



**HAL**  
open science

# Modélisation du métabolisme de la microalgue oléagineuse d'intérêt industriel *Microchloropsis gaditana*

Clémence Dupont Thibert

## ► To cite this version:

Clémence Dupont Thibert. Modélisation du métabolisme de la microalgue oléagineuse d'intérêt industriel *Microchloropsis gaditana*. Biotechnologies. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALV113 . tel-04617114

**HAL Id: tel-04617114**

**<https://theses.hal.science/tel-04617114>**

Submitted on 19 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : CSV - Chimie et Sciences du Vivant

Spécialité : Biotechnologie

Unité de recherche : LPCV - Laboratoire de Physiologie Cellulaire Végétale

**Modélisation du métabolisme de la microalgue  
oléagineuse d'intérêt industriel *Microchloropsis gaditana*.**

**Metabolic modeling of the industrial oleaginous microalga  
*Microchloropsis gaditana*.**

Présentée par :

**Clémence DUPONT THIBERT**

Thèse soutenue publiquement le **15 décembre 2023**, devant le jury composé de :

<b>M. Gilles CURIEN</b> CHARGE DE RECHERCHE HDR, CNRS délégation Alpes	Directeur de thèse
<b>Mme Yonghua LI-BEISSON</b> DIRECTRICE DE RECHERCHE, CEA Cadarache	Rapporteure
<b>Mme Sabine PERES</b> PROFESSEURE DES UNIVERSITES, Université Claude Bernard Lyon 1	Rapporteure
<b>M. Hans GEISELMANN</b> PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président du jury
<b>M. Guillaume COGNE</b> MAITRE DE CONFERENCES, Université de Nantes	Examineur

Invités :

<b>M. Florian DELRUE</b> INGENIEUR HDR, Plateforme Microalgues et Procédés	Co-encadrant de thèse
<b>M. Maxime DUROT</b> INGENIEUR DOCTEUR, TotalEnergies	Co-encadrant de thèse





# Remerciements

---

Maintenant que la soutenance est derrière moi, je peux me consacrer à la rédaction des derniers mots de ce manuscrit, que vous lirez probablement en premier... Tout cela aurait été impossible sans l'aide, le soutien, ou simplement la présence de nombreuses personnes. Dans cette partie, je tiens à exprimer ma gratitude envers tous ceux qui ont contribué de près ou de loin à la réalisation de cette thèse.

Je tiens d'abord à remercier chaleureusement mes directeurs de thèse **Maxime Durot**, **Gilles Curien** et **Florian Delrue**, pour m'avoir donné l'opportunité de réaliser ce doctorat. Merci à tous les trois pour la confiance que vous avez placée en moi, pour vos conseils avisés et pour m'avoir guidée tout au long de ces trois années de recherche doctorale.

J'aimerais exprimer mes remerciements à **Mme Yonghua Li-Beisson** et **Mme Sabine Pérès** pour avoir accepté d'être rapporteuses de mon manuscrit de thèse. Je vous remercie, ainsi que **M. Hans Geiselmann** et **M. Guillaume Cogne**, d'avoir accepté d'évaluer mon travail en participant à mon jury de thèse.

Mes remerciements vont également à **M. Jérémy Pruvost**, **Mme Caroline Baroukh** et **Mme Cécile Pereira**, pour leurs participations à mes comités de suivi de thèse. Chacun d'entre eux a apporté des retours constructifs qui ont grandement contribué à faire progresser ce projet de thèse.

Je tiens aussi à remercier **Éric Maréchal**, directeur du Laboratoire de Physiologie Cellulaire et Végétale du CEA de Grenoble ; **Jean-François Sassi**, directeur de la Plateforme Microalgues et Procédés du CEA de Cadarache et **Angélique Chanal**, directrice du programme R&D Biofuels de TotalEnergies, de m'avoir ouvert les portes de leurs laboratoires et programmes respectifs pour la réalisation de ce projet. Merci également à **Abril Lansoy** pour avoir permis à l'équipe Algue de m'accueillir pendant la majeure partie de ma thèse.

Un grand merci aussi à tous mes collègues du BioLab/programme Biofuels de TotalEnergies, des équipes Photosynthèse et Lipides du LPCV du CEA de Grenoble et de la plateforme MAP du CEA de Cadarache. Merci pour tous ces échanges professionnels et pour tous les bons moments partagés au CRES, au CEA et en dehors.

J'aimerais remercier en particulier **Juliette Jouhet, Elodie Billey, Séverine Collin et Sylvaine Roy.**

**Juliette**, merci de m'avoir fait découvrir le monde des lipides, d'avoir relu et corrigé l'intégralité de mes productions scientifiques (notamment ce manuscrit de thèse) et pour ton soutien indéfectible pendant la période de la rédaction ! **Elodie**, merci pour tes réponses à mes nombreuses questions, pour ta présence et ton soutien tout au long des trois dernières années et pour ta bonne humeur communicative ! **Séverine**, merci pour toutes tes questions et suggestions toujours pertinentes, et pour les conseils bienveillants que tu m'as donnés. **Sylvaine**, merci pour ton aide et tes conseils dans les tout derniers instants de ma thèse.

Je voudrais également remercier mes collègues (anciens et actuels) du CRES pour avoir partagé mon quotidien pendant ces dernières années. Merci pour votre accueil au sein du département, pour toutes les discussions, les pauses café, les afterworks, les week-ends et la bonne ambiance qui a régné au bâtiment 45 depuis mon arrivée en 2019 !

Tout d'abord merci à mes super co-bureaux : **Alice et Sylvain**. Merci pour votre accueil dans le carré VIP de l'open-space, pour toutes nos discussions et les conseils professionnels et personnels que vous m'avez donné. Merci aux membres de la Team Procédé : **Yann** (alias le sang) pour tous les week-ends randos/bières/saucissons que tu as organisé ! Merci de m'avoir accompagnée et attendue lors de quelques sorties ski de rando et alpinisme un peu périlleuses... **Alicia** pour nos rallyes autos quotidiens entre Lyon et Solaize puis sur les routes parfois chaotiques de Bulgarie (et bientôt en Namibie !!), nos soirées sport, bière(s) et jeux de sociétés pendant les différents confinements/couvres feux et tous les petits week-ends qui ont suivis... **Christine** pour ton optimisme à toute épreuve et nos discussions motivantes notamment sur les fins de randos compliquées. **Pierre** pour tes nombreux encouragements en particulier pendant la période de la rédaction puis de préparation de la soutenance. **Bernadette** pour ta gentillesse, ta bonne humeur et pour tous les petits plats Thermomix dont tu m'as fait profiter ces dernières années. Merci à la Team Biotech : **Patrick et Hervé**. Merci de m'avoir fait profiter de votre expertise et de m'avoir accueillie toujours avec le sourire dans vos laboratoires pour la réalisation de quelques manips. **Patrick** je tiens également à te remercier de m'avoir fait profiter de tes talents de photographe ! **Richard** pour être venu compléter de nombreuses fois notre team lors de pauses café et de weekends. **Mathilde, Nosra et Louis** pour votre gentillesse

et pour votre organisation et participation aux afterworks et aux soirées jeux de société des BioEvents qui ont été source de nombreux fous rires !

Je garde de supers souvenirs de toutes nos aventures ensemble : de la nuit sur de la glace à la survie à une invasion de zombies, en passant par la Fête de la Coquille St-Jacques de Villard de Lans... Et j'espère que nous aurons l'occasion d'en vivre d'autres !!

Merci également à tous les autres membres du département BioFuel et du CEA qui ont pu avoir un mot gentil au cours de ces dernières années, que je ne prendrai pas le risque de citer au risque d'en oublier mais qui se reconnaîtront !

Aussi, beaucoup de personnes ont contribué à la réussite de cette thèse en dehors des heures de travail. Leur présence et leur soutien sans faille ont été des plus précieux pour me permettre d'arriver au bout de cette aventure !

Merci à mes "co-doctorantes à distance" **Marjo** et **Rosa** ! Merci pour toutes nos petites discussions et tips échangés sur le PhD, et sur tout le reste ! Merci pour ces petits week-ends de pause (souvent sous le soleil de Montpellier). Et merci **Marjo** pour tous tes conseils et tes réponses à mes questions du tout début à la toute fin !

Merci aussi aux membres de la Greek Family : **Aurélie** et **Mégane**, à mes amies du Magistère (non je ne vous appellerai pas les Reines du Sud) : **Sofia**, **Justine**, **Charlotte**, **Alice** et **Tiphonie**, au Bruit de Fond alias **Carole**, **Pauline**, **Noémie**, **Cassandra** et **Hadjer**, mais également à **Yamna**, **Manon** et **Cécilia**. Merci d'avoir toujours répondu présentes et pour tout ce que nous avons partagé ces dernières années (il y a tellement de choses que je ne prendrai pas le risque d'essayer de les lister...).

Merci à mes Camarades des **Tripouettes**, **Marie** et **Carole**. Merci pour m'avoir divertie et bien souvent nourrie pendant les derniers mois de la thèse. Je garderai de très bons souvenirs de nos week-ends en montagne, nos aventures RP et nos concerts de rues !

**Carole**, tu mérites plus qu'amplement un petit paragraphe pour toi toute seule. Merci pour ton soutien quotidien ces derniers mois. Merci d'avoir été présente, d'avoir essayé de me remonter le moral et de me changer les idées en employant des techniques diverses et variées (que je ne détaillerai pas ici) sur la fin de la thèse. La période de la rédaction aura au moins permis une chose : trouver une

réponse partielle à ton éternelle question : mon premier est le melon ;). Hâte d'écrire le prochain chapitre de nos aventures ensemble autour du monde (180° pour être plus précise) !!!

Enfin, mes derniers remerciements vont évidemment à ma famille au sens large et plus particulièrement à **mes parents** et mes sœurs, **Azélie** et **Cassandra**, qui m'ont énormément soutenue et encouragée, et sans qui je ne serais pas arrivée jusque-là. Vous êtes une famille formidable !

# Table des matières

---

Remerciements.....	
Abréviations.....	
1 Chapitre 1.....	1
1.1 Contexte de la thèse.....	1
1.1.1 Les microalgues comme source de biocarburants.....	1
1.1.2 Diversité phylogénétique des microalgues.....	4
1.2 <i>M. gaditana</i> : la microalgue choisie pour l'étude.....	7
1.2.1 <i>M. gaditana</i> : une microalgue d'intérêt industriel.....	7
1.2.2 Importance et diversité des glycérolipides chez <i>M. gaditana</i> .....	8
1.2.3 Etat de l'art du métabolisme lipidique chez <i>M. gaditana</i> .....	16
1.3 Modèles métaboliques à l'échelle du génome.....	27
1.3.1 Modéliser le métabolisme.....	28
1.3.2 Les modèles métaboliques à l'échelle du génome.....	30
1.3.3 Applications des GEM à l'échelle du microorganisme.....	46
1.3.4 Modélisation du métabolisme des microalgues.....	53
1.3.5 Modéliser le métabolisme lipidique.....	55
1.4 Objectifs de la thèse.....	59
2 Chapitre 2.....	60
2.1 Préambule.....	60
2.2 Reconstruction et correction manuelle du modèle.....	61
2.2.1 Reconstruction de la première version du modèle par SilicoLife.....	61
2.2.2 Correction manuelle du modèle.....	61
2.2.3 Bilan des corrections réalisées au cours de la thèse.....	77
2.3 Résultats.....	80
2.4 Discussion.....	117
2.4.1 Choix de la stratégie de simulation de flux.....	117
2.4.2 Standards actuels de développement et de publication de GEM.....	118
2.4.3 Perspectives d'amélioration d'iMgadit.....	124

3	Chapitre 3.....	127
3.1	Introduction .....	127
3.1.1	Préambule.....	127
3.1.2	Cycle de développement de BioModTool.....	129
3.2	Spécifications des besoins .....	129
3.2.1	Identification des cas d'utilisation et des utilisateurs ciblés.....	130
3.2.2	Applicabilité à différents modèles métaboliques .....	130
3.2.3	Détermination des données de composition de biomasse exploitables .....	131
3.2.4	Structure et composition de la fonction de biomasse.....	132
3.2.5	Besoins spécifiques des utilisateurs du LPCV : exploitation des données lipidomiques de <i>M. gaditana</i> .....	133
3.3	Solutions techniques proposées .....	133
3.3.1	Implémentation de BioModTool.....	133
3.3.2	Format des données de composition de biomasse et structure de la fonction de biomasse .....	134
3.4	BioModTool – version tout public.....	137
3.4.1	Présentation de BioModTool .....	137
3.4.2	Implémentation et utilisation de BioModTool .....	145
3.4.3	Validation de BioModTool .....	158
3.5	Module spécifique aux données lipidomiques du LPCV .....	164
3.5.1	Présentation du module lipidomic_data_lpcv.py .....	164
3.5.2	Utilisation du module lipidomic_data_lpcv.py .....	166
3.5.3	Validation du module lipidomic_data_lpcv.py .....	168
3.6	Avantages et limites des solutions techniques implémentées dans BioModTool.....	168
3.7	Discussion et perspectives .....	173
3.7.1	BioModTool permet la valorisation de données variées .....	173
3.7.2	BioModTool est applicable à divers GEM d'organismes variés .....	175
3.7.3	Élargissement des applications des GEM disponibles pour les non-modeleurs.....	175
3.7.4	Détermination des identifiants des métabolites .....	176
3.7.5	Perspectives d'application à d'autres organismes du laboratoire.....	176
3.7.6	Perspectives d'automatisation de BioModTool.....	177
3.8	Conclusion.....	177

4	Chapitre 4.....	178
4.1	Conclusions relatives aux objectifs du projet de thèse .....	179
4.1.1	Correction du métabolisme lipidique de iMgadiT et intégration de données lipidomiques.....	179
4.1.2	Potentiel applicatif des méthodes développées pour les biologistes .....	180
4.1.3	Exploitation du modèle métabolique et de BioModTool pour l'interprétation de profils lipidiques détaillés .....	186
4.2	Autres perspectives d'applications pour iMgadiT .....	193
4.2.1	Perspectives pour l'ingénierie métabolique .....	193
4.2.2	Modélisation multi-échelle : passage à l'échelle du réacteur .....	196
4.3	Conclusion générale.....	201
	Annexes.....	202
A.	Article 1: Supplemental files .....	202
B.	Liste des contributions et communications.....	216
5	Bibliographie .....	218

# Abréviations

---

24E2COAR	2,4 di-enoyl-coa réductase
32ECOAI	delta3, delta2 – enoyl-coa isomerase
3OAR	3-oxoacyl réductase
3OAS	3-oxoacyl-ACP synthase
AA	Acide aminé
accoa	Acetyl-coa
ACCOAC	Acétyl-coa carboxylase
ACOADH	Acyl-coa déshydrogénase
ACOX	Acyl-coa oxidase
ACP	Acyl carrier protein (protéine porteuse d'acyle)
ACS	Acyl-coa synthétases
ADACT	S-adenosylmethionine:diacylglycérol 3-amino-3-carboxypropyl transférase
AND / DNA	Acide désoxyribonucléique / Deoxyribonucleic acid
AG / FA	Acide gras / Fatty acid
AG-ACP / FA-ACP	Acyl-ACP
AMP	Adénosine monophosphate
ARN / RNA	Acide ribonucléique / Ribonucleic acid
ATP	Adénosine triphosphate
ATPM	ATP maintenance
ATS1/ATS2	Glycérol-3-phosphate acyl-ACP acyltransférase
BiGG	Biochemical, Genetic and Genomic (base de données)
BLAST	Basic Local Alignment Search Tool (Outil de recherche d'alignement local)
BOF	Biomass objective function (Fonction objective de biomasse)
CAT	Catalase
CBM	Constraint-Based Method (Méthodes par contraintes)
CCUS	Carbon Capture, Utilization, and Storage (captage, stockage, utilisation du CO <sub>2</sub> )
CDP-DAG	Cytidine diphosphate diacylglycérol
CDP-DAGS	CDP-DAG synthase
CEPT	CDP-éthanolamine:DAG éthanolamine phosphotransférase
CHOPT	CDP-choline:DAG choline phosphotransférase
CoA	Coenzyme A
COBRA	Constraint-based Reconstruction and Analysis (Reconstruction et analyse basées sur les contraintes)
CSOM	Computational strain optimization method (méthode de conception de souches <i>in silico</i> )
CTP	Cytidine triphosphate
DAG	Diacylglycérol
DCW	Dry Cell Weight (Poids sec de cellules)
dFBA	Dynamic Flux Balance Analysis (Analyse de balance de flux dynamique)
dNTP	Désoxynucléotides triphosphates
DGAT	Diacylglycérol acyltransférase
DGDG	Digalactosyldiacylglycérol
DGDGS	DGDG synthase
DGTA	1,2-diacylglycéryl-3-O-2'-(hydroxyméthyl)-(N,N,N-triméthyl)-β-alanine
DGTS	1,2-diacylglycéryl-3-O-4'-(N,N,N-triméthyl)-homosérine

DGTSS	DGTS synthase
EAR	Énoyl-réductase
EC	Enzyme Commission (Numéro de commission d'une enzyme)
ECOAH	Enoyl-coa hydratase
EDO	Équations différentielles ordinaires
EFM	Elementary flux mode (Mode élémentaire de flux)
EIM / IMS	Espace intermembranaire mitochondrial / mitochondrial intermembrane space
ELO	Elongase
eMOMA	Version environnementale du MOMA
EPA	Acide eicosapentaénoïque
FAACPE	Acyl-ACP hydrolase
FAD	Fatty acid desaturase
FAIR	Findable, accessible, interoperable, reusable
FASI	Fatty acid synthase I
FASII	Fatty acid synthase II
FBA	Flux Balance Analysis (Analyse de balance de flux)
FC	Fold change
FVA	Flux Variability Analysis (analyse de variabilité de flux)
G3PAT	Glycérol-3-phosphate acyltransférase
GAM	Growth-Associated Maintenance (Coûts de maintenance associé à la croissance)
GC	Gas chromatography (chromatographie en phase gazeuse )
GEM	Genome-scale Model (Modèle métabolique à l'échelle du génome)
GLYC3P	Glycerol-3-phosphate
GO	Gene Ontology
GPR	Gène-Proteine-Réaction (association gène-réaction d'un GEM)
H <sup>+</sup> / h	Proton
HACD	Hydroxyacyl-coa-déshydrogénase NAD <sup>+</sup> dépendante
HPLC	High-performance Liquid Chromatography (Chromatographie en phase liquide à haute performance)
JSON	Java script object notation
KAS	Ketoacyl-ACP synthase
KAT	3-ketoacyl-coa thiolase / acétyl-coa acyltransférase
KEGG	Kyoto Encyclopedia of Genes and Genomes (base de données)
KO	Knock out
LC-MS	Liquid Chromatography Mass Spectrometry (Spectrométrie de masse couplé à la séparation par chromatographie en phase liquide)
LPAAT	Acyl-ACP/acide lysophosphatique acyltransférase
LPAT	Lysophosphatidate acyltransférase
lyso-PC	Lysophosphatidylcholine
MAG	Monoacylglycérol
MCOATA	Malonyl-coa ACP malonyltransférase
MCOATA	Malonyltransférase
MGDG	Monogalactosyldiacylglycérol
MGDGS	MGDG synthase
MOMA	Minimization Of Metabolic Adjustments (Minimization des ajustements métaboliques)
MRPA	Mean Rate of Photon Absorption (vitesse moyenne d'absorption des photons)
MS	Mass spectrometry (Spectrométrie de masse)
NGAM	Non-Growth Associated Maintenance (Coûts de maintenance non-associés à la croissance)

PA	Acide phosphatidique
PAD	Palmitic acid (16:0) desaturase
PAP	Phosphatidic acid phosphatase
PBR	Photobioréacteur
PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction (Reaction de polymerase en chaîne)
PDAT	Phospholipid:diacylglycérol acyltransférase
PDH	Pyruvate déshydrogénase
PE	Phosphatidyléthanolamine
pFBA	Parsimonious FBA (FBA parcimonieuse)
PG	Phosphatidylglycérol
PGPP	PG phosphate phosphatase
PGS	Phosphatidylglycérol synthase
PGSA	PG phosphate synthase
PI	Phosphatidylinositol
PIS	Phosphatidylinositol synthase
PP	Pentose phosphate
ppi	Pyrophosphate
PPP	Phosphoénoylpyruvate
PS	Phosphatidylsérine
PSII	Photosystème II
PSS	PS synthase
PUFA	Poly-unsaturated fatty acid (acide gras poly-insaturé)
RE / ER	Réticulum endoplasmique / Endoplasmic reticulum
ROOM	Regulatory on/off minimization
SAD	Stearic acid (18:0) désaturase
SBML	Systems biology markup language
SBO	Systems biology ontology
sn1-glyc3p	Acide lysophosphatidique
<i>sn-X</i>	Stereospecifically numbered (numérotation stéréochimique)
SQDG	Sulfoquinovosyldiacylglycérol
SQDGS	SQDG synthase
TAG	Triacylglycérol
UDP	Uridine diphosphate
UDPGAL	UDP-alpha-D-galactose
UDPSQ	UDP-6-sulfoquinovose
VLC-PUFA	Very-Long Chain polyunsaturated Fatty Acid (acides gras polyinsaturés à très longue chaîne)
WT	Wild-Type (souche sauvage)

# 1 Chapitre 1

---

## Introduction

Les travaux présentés dans ce manuscrit s'inscrivent dans le cadre d'une thèse CIFRE réalisée au sein de la branche Recherche et Développement de TotalEnergies (OneTech), du Laboratoire Physiologie Cellulaire et Végétale (LPCV) et l'équipe de Microalgues et Procédés (MAP) du Commissariat à l'énergie atomique et aux énergies alternatives (CEA). Ces travaux de recherche et développement s'intègrent dans des projets collaboratifs entre TotalEnergies et le CEA ayant pour objectif de caractériser le métabolisme lipidique des microalgues, et plus particulièrement de la microalgue oléagineuse *Microchloropsis gaditana*, pour des applications dans le domaine des biocarburants.

Ce chapitre introductif a pour objectif de situer la thèse dans son contexte scientifique tout en introduisant les concepts biologiques et bioinformatiques qui seront utilisés dans cette étude.

### 1.1 Contexte de la thèse

#### 1.1.1 Les microalgues comme source de biocarburants

La croissance démographique mondiale ainsi que le développement économique et social à travers le monde entraînent une augmentation des besoins en énergie. Cette tendance s'accompagne d'une préoccupation grandissante quant au changement climatique et à l'épuisement des énergies fossiles non renouvelables. Dans ce contexte, la recherche et le développement de ressources énergétiques alternatives et renouvelables, en particulier de biocarburants, jouent aujourd'hui un rôle essentiel pour répondre à ces défis environnementaux et énergétiques.

Plusieurs "générations" de biocarburants ont été définies, en fonction de l'origine de la biomasse utilisée et des procédés de transformation (Ministère de la Transition écologique et de la Cohésion des territoires, Ministère de la Transition énergétique 2023).

Les biocarburants de première génération (biocarburants dits conventionnels) proviennent de plantes agricoles. On distingue (1) les bioéthanol, obtenu par fermentation des glucides issus de plantes sucrières (comme la betterave ou la canne à sucre) et de plantes amylacées (comme le maïs ou le blé)

; et (2) les biodiesels, produits par transestérification d'acides gras issus de plantes oléagineuses telles que le colza, le tournesol ou le palmier à huile. Cette première génération de biocarburants est produite à l'échelle industrielle. Cependant, elle nécessite l'exploitation de terres arables et la consommation d'eau douce et entre donc directement en concurrence avec la production de nourriture humaine et animale.

La deuxième génération de biocarburants (biocarburants avancés) repose principalement sur la valorisation des parties lignocellulosiques des plantes qui sont inutilisables pour l'alimentation humaine. Ces déchets lignocellulosiques sont largement disponibles et proviennent de résidus agricoles (tels que la paille) et de rémanents forestiers (comme les chutes de bois). Des cultures à croissance rapide telles que le peuplier ou l'eucalyptus sont également utilisées. La matière lignocellulosique peut être transformée en biodiésel et biokérosène par voie thermochimique (gazéification) ou en bioéthanol par voie biochimique (hydrolyse enzymatique). Les biocarburants de deuxième génération restent en compétition avec l'agriculture en termes d'utilisation des terres arables et d'eau douce. Les premières unités industrielles de biocarburants de seconde génération ont vu le jour, et des projets de recherche et développement sont en cours afin de produire des biocarburants à des prix compétitifs tout en atteignant les meilleurs bilans environnementaux possibles.

La troisième génération de biocarburants repose sur l'exploitation de la biomasse des algues (macro- ou micro- algues). Ces biocarburants de troisième génération, sont principalement produits à partir de microalgues. Les microalgues sont des microorganismes unicellulaires photosynthétiques qui se développent dans des milieux aquatiques, qu'ils soient salins ou d'eau douce. A la différence des deux premières générations de biocarburants, la culture des microalgues n'entre pas en concurrence avec les surfaces agricoles. Certaines souches de microalgues, capables de produire et d'accumuler de grandes quantités de lipides notamment sous forme de triacylglycérols (TAG), constituent des sources prometteuses d'acides gras pour la production de biodiesel (Chisti 2007). D'autres souches, telles que *Chlamydomonas* ou *Chlorella*, dont la biomasse est riche en carbohydrates dans certaines conditions, constituent des sources potentielles pour la production de bioéthanol (Ho et al. 2013; Condor et al. 2022). Par ailleurs, les microalgues ont la capacité de produire d'autres molécules à haute valeur ajoutée, telles que des pigments, des antioxydants, des acides gras polyinsaturés à longue chaîne etc. De ce fait, elles présentent un intérêt considérable pour diverses applications biotechnologiques, autres que les biocarburants, notamment dans les domaines de la cosmétique, de l'agroalimentaire, de la pharmacie, et bien d'autres encore (Spolaore et al. 2006; Chu 2012; Rizwan et al. 2018). Dans cette thèse nous nous intéresserons à la production de lipides par les microalgues oléagineuses, notamment par *M. gaditana*, pour la production de biodiesels.

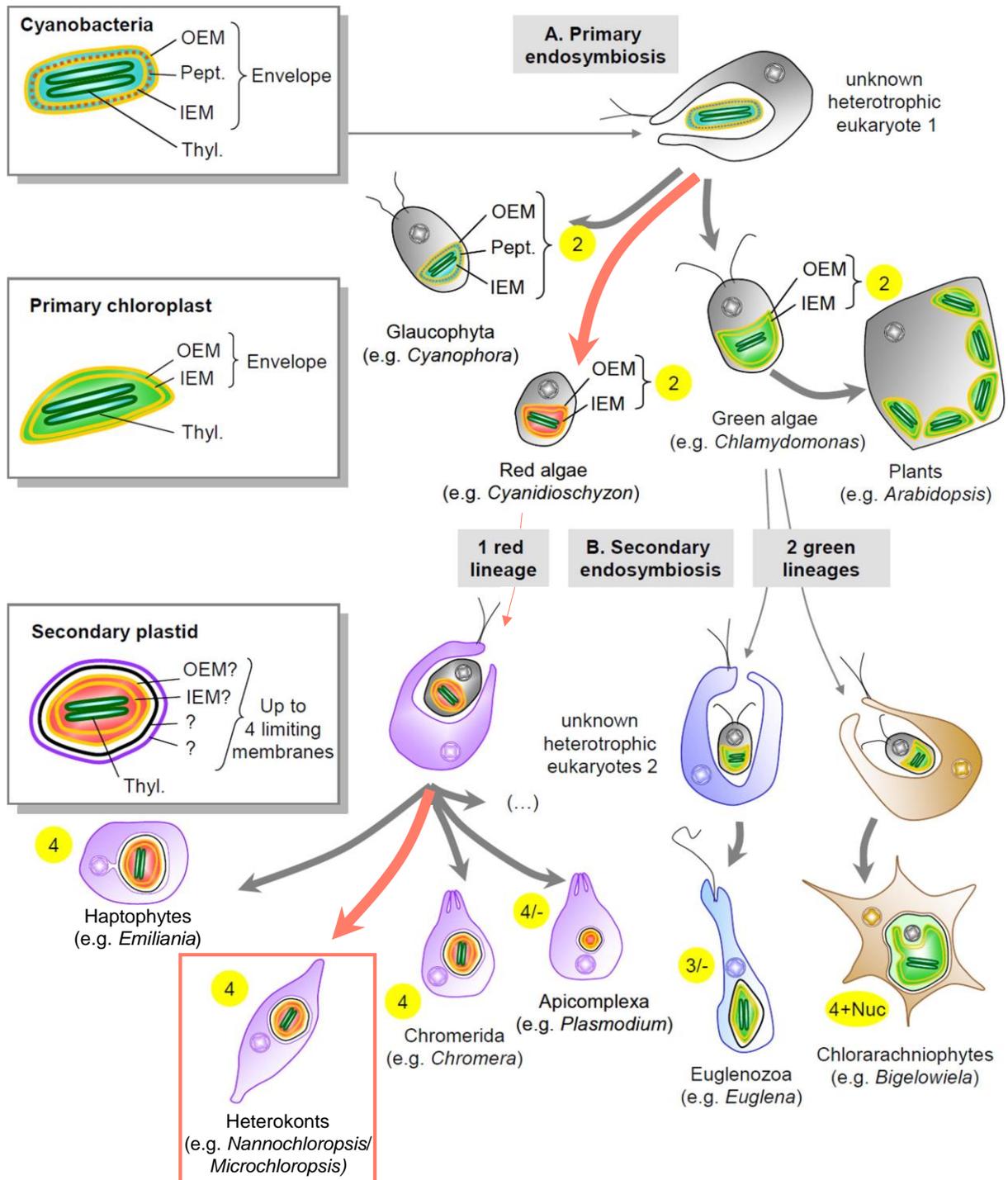
En plus de ne pas nécessiter de terres agricoles, les microalgues offrent de nombreux avantages (Sundaram et al. 2023). (1) Elles ont une croissance rapide avec un doublement de la biomasse de l'ordre de quelques heures (24 heures en moyenne). (2) Certaines souches peuvent se développer dans l'eau salée ou même dans des eaux usées (Henkanatte-Gedera et al. 2017) et n'entrent donc pas en concurrence avec l'agriculture pour l'eau douce. (3) Les microalgues fixent le CO<sub>2</sub> atmosphérique par la photosynthèse, intégrant ensuite le carbone dans des molécules organiques riches en énergie. Ces microorganismes peuvent ainsi être exploitées dans des technologies dites CCUS (captage, stockage, utilisation du CO<sub>2</sub> ou Carbon Capture, Utilization, and Storage) afin de valoriser le CO<sub>2</sub> issu de processus industriels (Marechal 2021). (4) Leur productivité surfacique est supérieure à celle des plantes terrestres. Par exemple, le colza produit 1190 L/ha d'huile, le palmier à huile en produit 5950 L/ha, alors que les microalgues sont capables d'avoir un rendement de production d'huile entre 58 000 et 136 000 L/ha (Chisti 2007). En effet, dans certaines conditions, certaines souches de microalgues peuvent accumuler jusqu'à de 80-90% de leur poids sec sous forme de lipides (Metting 1996; Spolaore et al. 2006; Mata, Martins et Caetano 2010). Cependant, lorsqu'elles sont cultivées dans des conditions normales, les microalgues ont une teneur en lipides relativement faible. Chez certaines espèces, la production de lipides peut être significativement augmentée par un stress, notamment par une carence nutritive en azote ou en phosphate (John J. Sheehan et al. 1998; Simionato et al. 2013; Jouhet et al. 2017; Janssen, Kastenhofer et al. 2018). Toutefois, ces conditions de carence entraînent l'arrêt de la division cellulaire. L'optimisation de la productivité en lipides nécessite ainsi un compromis entre croissance et production de lipides (avec un stress ralentissant la croissance). Des procédés en deux phases, commençant par une culture en milieu riche afin d'obtenir la biomasse désirée, suivi d'une carence en azote entraînant l'accumulation de lipides, sont des approches actuellement explorées. Néanmoins, ce processus, qui nécessite une modification du milieu de culture et un délai supplémentaire avant la récolte de la biomasse, est coûteux. Ainsi, la conciliation de la croissance et de l'accumulation de lipides fait partie des verrous critiques à lever pour arriver à une production d'huile algale économiquement viable. L'optimisation des conditions de culture, ainsi que le développement de souches génétiquement modifiées dans le but d'augmenter la production de lipides font actuellement l'objet de nombreuses recherches.

Par ailleurs, l'optimisation des procédés industriels, en particulier en ce qui concerne les systèmes de culture, de récolte, de prétraitement etc., est encore nécessaire afin de diminuer les coûts de production (Chisti 2007; Mata, Martins et Caetano 2010; Razzak et al. 2013; Aransiola et al. 2014; Park, Nguyen et Jin 2019; Kashyap et al. 2023; Sundaram et al. 2023).

## 1.1.2 Diversité phylogénétique des microalgues

Il existe une grande diversité phylogénétique au sein des microalgues (Metting 1996; Keeling 2010, 2013; Petroutsos et al. 2014). En effet, les microalgues comprennent à la fois des organismes procaryotes (cyanobactéries) et des organismes eucaryotes issus de plusieurs évènements d'endosymbioses. Les endosymbioses ont eu un impact significatif sur l'évolution et la diversification des eucaryotes (Figure 1.1). Selon la théorie de l'endosymbiose, les mitochondries et les plastes ont évolué à partir de procaryotes phagocytés par des hôtes eucaryotes et sont devenus des organites des cellules eucaryotes. Au cours de l'évolution, des transferts horizontaux de gènes entre l'organite et le génome de son hôte ont été observés (Zimorski et al. 2014). Un premier évènement d'endosymbiose primaire (internalisation d'un procaryote dans un eucaryote) a eu lieu entre une cyanobactérie ancestrale, capable de réaliser la photosynthèse grâce aux thylakoïdes présents à l'intérieur de ses deux membranes, et un eucaryote hétérotrophe inconnu. Une fois ingérée par l'hôte, la cyanobactérie a été conservée sous forme de chloroplaste primaire permettant ainsi à l'eucaryote phagocytaire de réaliser à son tour la photosynthèse. Cette endosymbiose a donné naissance à toutes les lignées contenant un chloroplaste. Trois lignées majeures ont émergé de cette endosymbiose primaire : (1) La lignée des algues vertes (telle que *Chlamydomonas*), c'est de cette lignée que les plantes supérieures ont évolué ; (2) la lignée des algues rouges et (3) la lignée des *Glaucophytes* (ou lignée bleue). De multiples évènements d'endosymbioses secondaires (un eucaryote possédant déjà un endosymbiote entre en endosymbiose dans un autre eucaryote) et tertiaires ont ensuite eu lieu entre les algues rouges ou vertes et des eucaryotes hétérotrophes différents, donnant ainsi naissance à de nombreuses lignées (Keeling 2010, 2013; Petroutsos et al. 2014). Le chloroplaste de ces lignées est caractérisé par la présence de quatre membranes : les deux plus internes provenant de la cyanobactérie initiale, et les deux plus externes provenant probablement des endosymbioses primaires et secondaires.

Au fil de l'évolution, les microalgues ont donc acquis des gènes à partir de divers génomes. Le génome des microalgues est ainsi devenu un "melting pot" de gènes, donnant naissance à un réseau métabolique ressemblant à une mosaïque (Bowler et al. 2008; Curtis et al. 2012). Du fait de leur histoire évolutive complexe, le métabolisme des microalgues est difficilement comparable et extrapolable entre les différentes souches et à fortiori avec celui des plantes supérieures.



**Figure 1.1. Représentation schématique de l'évolution des plastes (source: Petroustos et al. 2014).** L'histoire évolutive des genres *Nannochloropsis* et *Microchloropsis* est mise en avant (en rouge). (A) Endosymbiose primaire. Un évènement d'endosymbiose entre un eucaryote hétérotrophe inconnu et une cyanobactérie a conduit à la formation de trois lignées différenciées par leur composition en pigment et possédant un plaste primaire à deux membranes : la lignée bleue (Glaucocystophytes), la lignée rouge (algues rouges), et la lignée verte (algues vertes et plantes). (B) Endosymbioses secondaires. Au moins deux évènements d'endosymbiose secondaire ont eu lieu au cours de l'évolution des plastes. Une première endosymbiose entre une algue rouge et un hôte hétérotrophe inconnu aboutissant à de multiples lignées dont les Haptophytes, les Heterokonts (telles que les genres

*Nannochloropsis* et *Microchloropsis*), les *Chromerida* et les *Apicomplexa*. Au moins deux événements endosymbiotiques indépendants entre une algue verte et un eucaryote hétérotrophe inconnu ont généré les lignées des *Euglenozoa* et des *Chlorarachniophytes*. Les chiffres en jaune indiquent le nombre de membranes délimitant les chloroplastes (de 2 à 4). Les phyla incluant des espèces ayant perdu leurs plastides sont indiqués par 3/- et 4/-. Les phyla dans lesquels le noyau primaire a été conservé en tant que nucléomorphe sont indiqués par 4+Nuc.

Ainsi, les microalgues se révèlent être des ressources d'un grand intérêt pour diverses applications industrielles, notamment la production de biodiesels. Actuellement, l'une des raisons de la difficulté à rationaliser la recherche et le développement dans ce domaine réside dans le manque de connaissances détaillées d'espèces prometteuses pour l'industrie. Ces connaissances sont difficiles à acquérir car elles ne peuvent pas être extrapolées à partir d'espèces modèles étudiées en laboratoire depuis plusieurs décennies, telles que *Chlamydomonas* ou *Chlorella*. Ces souches se caractérisant par l'accumulation de carbohydrates lors de carences sont ainsi plus intéressantes pour la production de bioéthanol que de biodiesels. De nombreuses souches oléagineuses prometteuses, telles que *M. gaditana*, ont été identifiées chez les *Heterokontes*. Cependant leur métabolisme reste assez peu caractérisé. Approfondir ces connaissances est crucial afin d'appréhender de manière globale leur croissance et leurs réponses aux stress environnementaux dans l'objectif *in fine* d'améliorer les rendements de production pour les différentes applications biotechnologiques.

Ce travail de thèse s'inscrit dans le cadre de projets de recherche et développement visant à mieux comprendre le métabolisme lipidique des microalgues, en particulier de *M. gaditana* une microalgue oléagineuse prometteuse pour la production de lipides destinés à des applications industrielles, notamment la production de TAG pour les biodiesels. Plus précisément, l'objectif principal du projet de thèse était de réaliser une modélisation précise du métabolisme lipidique de cette microalgue au sein d'un modèle métabolique.

## 1.2 *M. gaditana* : la microalgue choisie pour l'étude

### 1.2.1 *M. gaditana* : une microalgue d'intérêt industriel

Le genre *Nannochloropsis* (aujourd'hui divisé en deux genres : *Nannochloropsis* et *Microchloropsis*) a été décrit pour la première fois par Hibberd en 1981 (Hibberd 1981). Il résulte d'une endosymbiose secondaire d'une algue rouge et d'un eucaryote hétérotrophe (Figure 1.1) (Petroutsos et al. 2014; Fawley, Jameson et Fawley 2015). Sa classification taxonomique est présentée dans la Table 1.1 (Hibberd 1981).

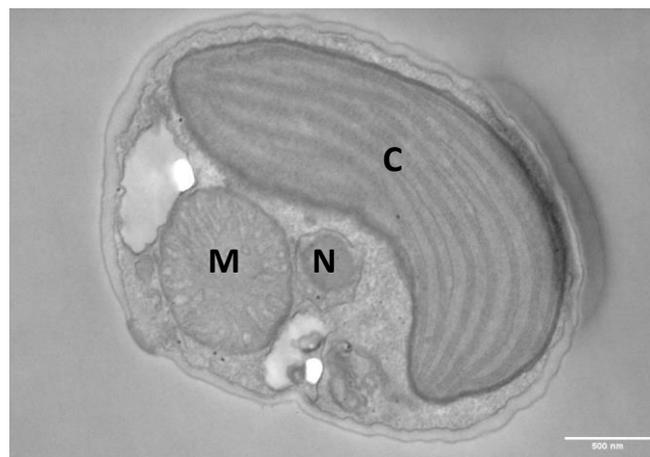
**Table 1.1. Classification taxonomique des *Nannochloropsis* et *Microchloropsis*.**

Classification	
Règne	<i>Chromista</i>
Infra-règne	<i>Heterokonta</i>
Phylum	<i>Ochrophyta</i>
Classe	<i>Eustigmatophyceae</i>
Ordre	<i>Eustigmatales</i>
Famille	<i>Monodopsidaceae</i>
Genre	<i>Nannochloropsis</i> / <i>Microchloropsis</i>

A ce jour, cinq espèces de *Nannochloropsis* : *N. granulata*, *N. limnetica*, *N. oceanica*, *N. oculata*, et *N. australis* et deux espèces de *Microchloropsis* : *M. gaditana* et *M. salina* ont été décrites (Fawley et al. 2015; Gong et al. 2020). Leur chloroplaste est caractérisé par la présence de quatre membranes (Petroutsos et al. 2014). Ces espèces de microalgues présentent les particularités de ne pas synthétiser de chlorophylles b et c mais seulement de la chlorophylle a et de ne pas accumuler d'amidon (Lubián et al. 2000; Archibald et Keeling 2002).

Les espèces de *Nannochloropsis* et *Microchloropsis* ont suscité un grand intérêt en raison de leur capacité à produire des niveaux élevés de lipides (Benvenuti et al., 2015; Radakovits et al., 2012). Elles accumulent de grandes quantités de lipides (jusqu'à 60% de leur poids sec) notamment lors de carence en azote (Simionato et al. 2013; Ma et al. 2014; Al-Hoqani, Young et Purton 2017; Janssen, Kastenhofer et al. 2018). Elles sont également riches en acides gras polyinsaturés oméga-3 ( $\omega$ 3) et oméga-6 ( $\omega$ 6). Une teneur en acide eicosapentaénoïque (EPA, 20:5) de 20 % molaire d'acide gras a par exemple été reportée chez *M. gaditana* (Jouhet et al., 2017). Ces acides gras  $\omega$ 3 et  $\omega$ 6 sont connus pour avoir des effets bénéfiques sur la santé humaine. Chez l'homme, la synthèse des acides gras  $\omega$ 3 étant insuffisante à ses besoins, son alimentation doit en contenir (Ma et al. 2016; Al-Hoqani, Young et Purton 2017).

La souche marine *M. gaditana* (Figure 1.2), en raison de sa capacité à produire de la biomasse et des lipides à des taux élevés, ainsi que sa tolérance à des conditions de culture variées en termes de pH, de température et de salinité, constitue une souche d'intérêt industriel pour la production de diesels. Cette souche est ainsi devenue un organisme modèle pour la production de biodiesel (Taleb et al. 2016; Janssen, Driessen et al. 2018a). En outre, plusieurs génomes de *M. gaditana* ont été séquencés et sont disponibles dans la littérature scientifique (Radakovits et al. 2012; Corteggiani Carpinelli et al. 2014). Diverses méthodes d'ingénierie génétique ont été développées pour cette souche, notamment les technologies TALEN et CRISPR-Cas9 (Poliner, Farré et Benning 2018; Kurita et al. 2023).



**Figure 1.2. Ultrastructure d'une cellule de *M. gaditana* (NG526) observée par microscope électronique en transmission à balayage (STEM). (Crédit photo : Grégory Si Larbi).**

Cellules cultivées en milieu liquide en présence d'azote et préparées par fixation chimique. C : chloroplaste, M : mitochondrie, N: noyau. Echelle : 500 nm. Image réalisée par le Laboratoire de Physiologie Cellulaire et Végétale du CEA de Grenoble sur un microscope MERLIN (Zeiss, Oberkochen, Germany ; 30 KV, 240pA).

### 1.2.2 Importance et diversité des glycérolipides chez *M. gaditana*

A ce jour, il n'existe pas de définition précise du terme «lipide». Contrairement aux glucides, aux protéines et aux acides nucléiques qui sont définis par leur structure chimique, les lipides qui comprennent notamment les huiles, les graisses et les cires, sont caractérisés par une propriété physico-chimique: leur hydrophobicité. Généralement les lipides sont non miscibles dans l'eau et sont solubles dans les solvants organiques tels que le chloroforme, l'éther, le benzène ou l'acétone. Les lipides présentent une grande diversité structurale : glycérolipides, stérols, terpénoïdes ; et fonctionnelle: stockage d'énergies (huiles), rôles structurels (membranes), signalisation (hormones).

En raison de l'abondance des glycérolipides (des esters d'acides gras et de glycérol) au sein de la biomasse des microalgues, cette thèse se concentrera sur les glycérolipides et les acides gras. Les glycérolipides sont les constituants majeurs des membranes biologiques et des gouttelettes lipidiques.

Ils jouent ainsi un rôle essentiel pour la structure cellulaire ainsi que pour l'adaptation des organismes à leur environnement. Ils sont notamment impliqués dans la réponse au stress et permettent le stockage d'énergie sous forme de réserves.

Les microalgues, telles que *M. gaditana*, présentent une très grande diversité moléculaire d'acides gras et de glycérolipides.

### 1.2.2.1 Diversité moléculaire des acides gras

Un acide gras est un acide carboxylique à chaîne aliphatique, possédant une chaîne carbonée linéaire de 8 à plus de 30 atomes de carbone, typiquement en nombre pair chez les algues. Les acides gras dits saturés sont dépourvus de doubles liaisons entre les atomes de carbone. Une double liaison (acides gras monoinsaturés) ou des liaisons multiples (acides gras polyinsaturés) peuvent être introduites à des positions spécifiques de la chaîne carbonée par des enzymes appelées désaturases (Dolch et Maréchal 2015). La position de ces doubles liaisons peut varier sur la chaîne carbonée. La nomenclature abrégée C:D est souvent utilisée pour désigner les acides gras : où C indique la longueur de la chaîne carbonée et D le nombre d'insaturations. Par exemple, l'acide gras 18:3 a une chaîne de 18 atomes de carbone et possède trois doubles liaisons. Dans cet exemple, le positionnement des doubles liaisons dans la chaîne carbonée et leur isométrie (*cis*, *trans*) ne sont pas précisés. D'autres nomenclatures sont également fréquemment utilisées pour désigner les acides gras. Ces différentes nomenclatures sont discutées plus en détail dans le [Chapitre 2](#).

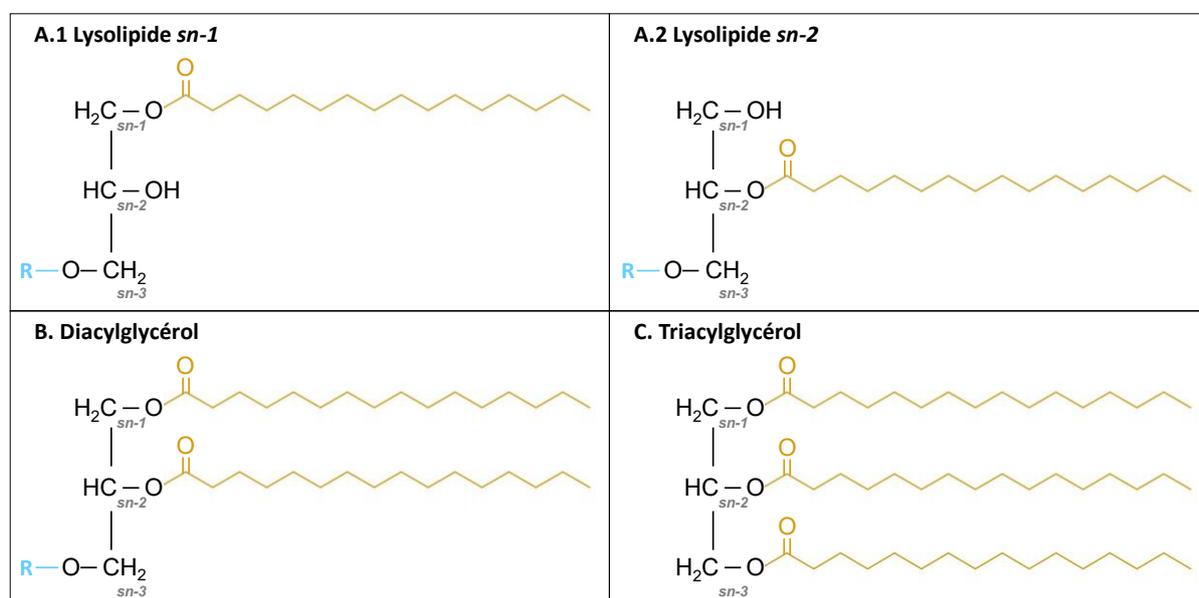
Les microalgues contiennent une grande diversité moléculaire en acides gras. Cette diversité repose sur la longueur de la chaîne carbonée, le nombre et de la position des insaturations. Chez *M. gaditana*, les acides gras polyinsaturés peuvent comporter jusqu'à cinq insaturations. La Table 1.2 présente la diversité et l'abondance relative des acides gras identifiés chez *M. gaditana*.

**Table 1.2. Exemple de profil en acides gras de *M. gaditana*. (D'après Billey et al. 2021).**  
Seuls les acides gras représentant au moins 1% des acides gras totaux sont listés.

Acides gras	% mol
14:0	5,1
16:0	22,1
16:1	28,3
16:2	1,0
16:3	1,0
18:0	1,9
18:1	1,6
18:2	1,2
18:3	1,7
20:4	3,3
20:5	30,5

### 1.2.2.2 Diversités moléculaire et fonctionnelle des glycérolipides

Les glycérolipides sont constitués d'une molécule de glycérol, un triol (propane 1,2,3-triol), à laquelle sont estérifiés un, deux ou trois acides gras. Ils peuvent posséder (ou non) une tête polaire de structure chimique variable. Les atomes de carbone du glycérol sont numérotés stéréo-spécifiquement suivant la nomenclature *sn-1*, *sn-2* et *sn-3* (*sn*: stereospecifically numbered). La Figure 1.3 présente la structure chimique de différents glycérolipides. Les glycérolipides membranaires possèdent un (lysolipides, Figure 1.3 A1 et A2) ou deux acides gras hydrophobes en position *sn-1* et/ou *sn-2* (Figure 1.3B), formant ainsi une "queue hydrophobe", et une tête polaire hydrophile en position *sn-3*. Les triacylglycérols (TAG) ne possèdent pas de tête polaire, mais un troisième acide gras hydrophobe (Figure 1.3C).

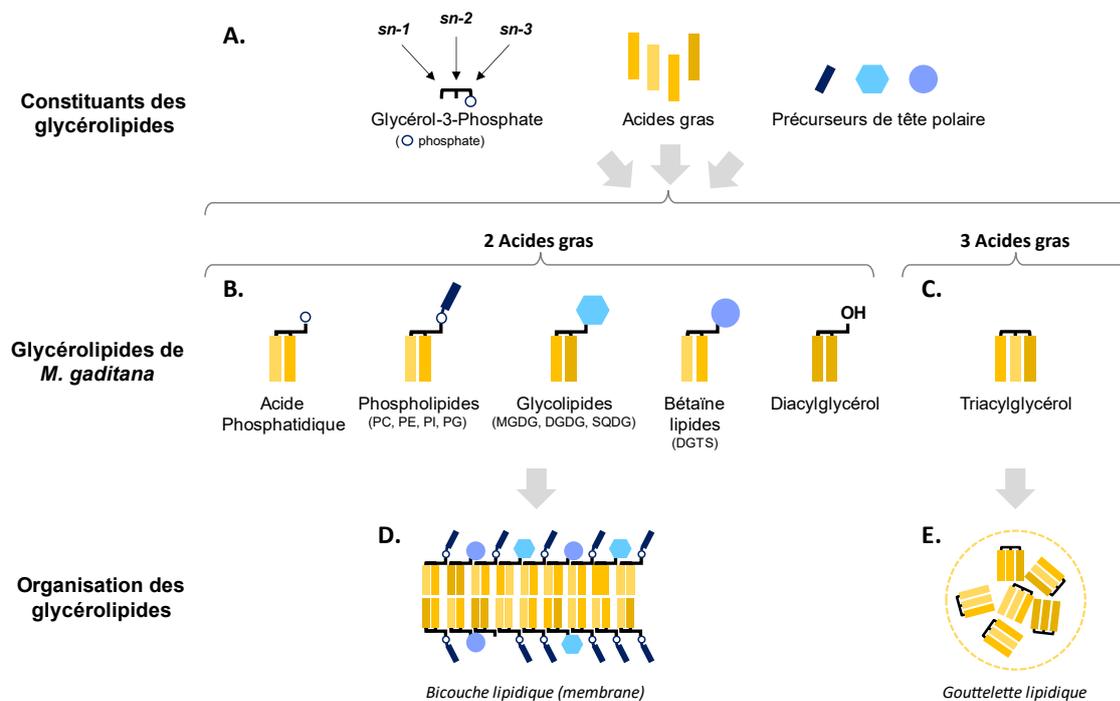


**Figure 1.3. Structure chimique des glycérolipides.**

Le squelette glycérol est représenté en noir. *sn-x* correspond à la numérotation stéréochimique des atomes de carbone du glycérol. En bleu, la tête polaire hydrophile, de structure chimique variable, en position *sn-3*. Les acides gras hydrophobes, dont les nombres de carbone et d'insaturation sont variables, sont représentés en orange. (A.1 et A.2) Les lysolipides possèdent un unique acide gras estérifié en position *sn-1* ou *sn-2*. (B) Les diacylglycérols naturels portent deux acides gras estérifiés en positions *sn-1* et *sn-2*. (C) Les triacylglycérols sont composés de trois acides gras.

Les glycérolipides sont généralement classés en différentes classes en fonction de leur structure et de leurs propriétés chimiques (Figure 1.4). Une classe de glycérolipides est définie par la nature de la tête polaire présente à la position *sn-3* du glycérol. Chez *M. gaditana*, il existe une dizaine de classes de glycérolipides regroupées en sous-groupes selon leur tête polaire : (1) les phospholipides : l'acide phosphatidique (PA), la phosphatidylcholine (PC), la phosphatidyléthanolamine (PE), le phosphatidylinositol (PI), le phosphatidylglycérol (PG) ; (2) les glycolipides : le

monogalactosyldiacylglycérol (MGDG), le digalactosyldiacylglycérol (DGDG), le sulfoquinovosyldiacylglycérol (SQDG) ; et (3) les bêtaines lipides : le 1,2-diacylglycéryl-3-O-4'-(N,N,N-triméthyl)-homosérine (DGTS). On retrouve également (4) du diacylglycérol (DAG) ayant une fonction alcool en position *sn*-3 et (5) du triacylglycérol (TAG). Un exemple de profil de classes glycérolipidiques de *M. gaditana* est détaillé dans la Table 1.3.



**Figure 1.4. Structure et caractéristiques des principales classes de glycérolipides de *M. gaditana* (D'après Lupette et Maréchal 2018).**

(A) Les briques de base des glycérolipides sont le glycérol-3-phosphate (donnant le squelette glycérol), les acides gras de longueur de chaîne et au nombre d'insaturations variés, et les têtes polaires. (B) Glycérolipides membranaires. (C) Triacylglycérol. (D) Les bicouches membranaires sont constituées de deux couches de glycérolipides. (E) Les triacylglycérols forment des gouttelettes lipidiques à l'intérieur des cellules.

**Table 1.3. Exemple de profil de classes glycérolipidiques de *M. gaditana* (D'après Billey et al. 2021).**

Classe	$\mu\text{g} \cdot 10^{-6}$ cellules
SQDG	$1,08 \cdot 10^{-1}$
MGDG	$1,07 \cdot 10^{-1}$
DGDG	$5,88 \cdot 10^{-2}$
PG	$5,57 \cdot 10^{-2}$
PI	$8,20 \cdot 10^{-2}$
PE	$5,02 \cdot 10^{-2}$
PC	$9,36 \cdot 10^{-2}$
DGTS	$5,69 \cdot 10^{-2}$
DAG	$6,08 \cdot 10^{-3}$
TAG	$4,71 \cdot 10^{-2}$

Les glycérolipides peuvent également différer par la nature de leurs chaînes carbonées, ce qui conduit à une très grande diversité de lipides spécifiques différents (Jouhet et al. 2017). Un lipide spécifique correspond à une molécule lipidique d'une classe donnée pour lequel les trois groupements fonctionnels présents sur le squelette glycérol sont identifiés et positionnés. Ces informations peuvent être précisées en indiquant le positionnement des acides gras dans l'ordre *sn-1/sn-2* (et si applicable *sn-3*). Par exemple le lipide spécifique PC(16:0/18:2) est formé d'un squelette glycérol comportant : (1) une choline en position *sn-3*, (2) un acide gras saturé à 16 carbones estérifié en position *sn-1* et (3) un acide gras à 18 carbones présentant deux insaturations en position *sn-2*. Lorsque la position des chaînes carbonées n'est pas connue, ces dernières sont séparées d'un symbole underscore "\_". Par exemple, la nomenclature PC(16:0\_18:2) indique que la PC est estérifiée par les acides gras 16:0 et 18:2 mais qu'on ne connaît pas leur position *sn-1* et *sn-2* respective (Liebisch et al. 2020). Par simplification ou manque d'information le nombre total d'atomes de carbone et d'insaturations peuvent être utilisés (PC(34:2) pour l'exemple précédent).

La composition et la position spécifique des différents acides gras sur le squelette glycérol peuvent être déterminées expérimentalement. Au LPCV, des analyses ont été réalisées afin de caractériser la composition et la position des acides gras des différents glycérolipides de *M. gaditana* (Jouhet et al. 2017). Chaque classe de glycérolipides présente un profil distinct d'acides gras (Jouhet et al. 2017; Billey et al. 2021). Un exemple de glycérolipidome détaillé est présenté dans la Table 1.4.

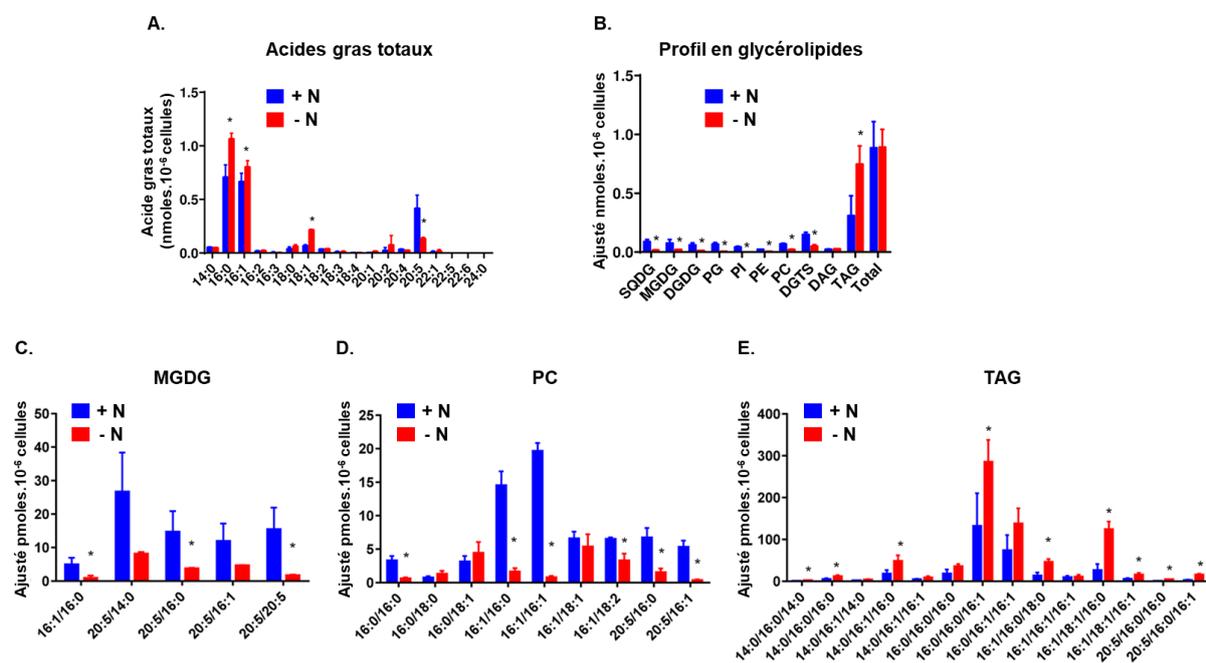
**Table 1.4. Exemple de profil glycérolipidique détaillé de *M. gaditana*.***Seuls les lipides spécifiques représentant plus de 1% de toutes les espèces présentes dans la classe sont indiqués.*

Classe	Glycérolipide (Identifiant du LPCV)	Positionnement des acides gras (Jouhet et al. 2017)	% mol/mol classe (Billey et al. 2021)
SQDG	SQDG 30:0	SQDG 14:0/16:0	12,06
	SQDG 30:1	SQDG 14:0/16:1	1,97
	SQDG 32:1	SQDG 16:1/16:0	83,47
MGDG	MGDG 32:1	MGDG 16:1/16:0	6,67
	MGDG 32:2	MGDG 16:1/16:1	5,39
	MGDG 34:4	MGDG 20:4/14:0	5,29
	MGDG 34:5	MGDG 20:5/14:0	43,71
	MGDG 36:4	MGDG 20:4/16:0	1,13
	MGDG 36:5	MGDG 20:5/16:0	8,91
	MGDG 36:6	MGDG 20:5/16:1	11,52
	MGDG 40:10	MGDG 20:5/20:5	14,92
DGDG	MGDG 40:9	MGDG 20:5/20:4	2,79
	DGDG 32:1	DGDG 16:1/16:0	21,54
	DGDG 32:2	DGDG 16:1/16:1	19,36
	DGDG 34:4	DGDG 20:4/14:0	2,38
	DGDG 34:5	DGDG 20:5/14:0	14,65
	DGDG 36:4	DGDG 20:4/16:0	2,35
	DGDG 36:5	DGDG 20:5/16:0	17,60
PG	DGDG 36:6	DGDG 20:5/16:1	21,83
	PG 32:0	PG 16:0/16:0	1,79
	PG 32:1	PG 16:1/16:0	14,88
	PG 32:2	PG 16:1/16:1	2,63
	PG 36:5	PG 20:5/16:0	73,58
PI	PG 36:6	PG 20:5/16:1	6,92
	PI 32:0	PI 16:0/16:0	9,34
	PI 32:1	PI 16:0/16:1	73,41
PE	PI 32:2	PI 16:1/16:1	18,33
	PE 36:4	PE 20:4/16:0	2,41
	PE 36:5	PE 20:5_16:0	9,39
	PE 36:6	PE 20:5/16:1	12,17
	PE 38:5	PE 20:4/18:1	1,50
	PE 38:6	PE 20:5/18:1	4,40
	PE 40:10	PE 20:5/20:5	9,92
	PE 40:6	PE 20:3/20:3	1,50
	PE 40:7	PE 20:3/20:4	8,90
	PE 40:8	PE 20:4/20:4	25,88
	PE 40:9	PE 20:4/20:5	24,15
PC	PC 32:0	PC 16:0/16:0	5,10
	PC 32:1	PC 16:0_16:1	20,77
	PC 32:2	PC 16:1/16:1	26,32
	PC 34:0	PC 16:0/18:0	1,27
	PC 34:1	PC 16:0/18:1	5,82

	PC 34:2	PC 16:1/18:1	12,95
	PC 34:3	PC 16:1/18:2	13,95
	PC 36:5	PC 20:5_16:0	8,57
	PC 36:6	PC 20:5_16:1	6,21
DGTS	DGTS 30:1	DGTS 14:0/16:1	9,05
	DGTS 30:2	DGTS 14:0/16:2	1,44
	DGTS 32:0	DGTS 16:0/16:0	5,51
	DGTS 32:1	DGTS 16:0_16:1	15,82
	DGTS 32:2	DGTS 16:1/16:1	12,62
	DGTS 34:3	DGTS 14:0/20:3	1,28
	DGTS 34:4	DGTS 14:0/20:4	1,41
	DGTS 34:5	DGTS 14:0/20:5	1,69
	DGTS 36:3	DGTS 20:3/16:0	1,82
	DGTS 36:4	DGTS 16:0/20:4	5,93
	DGTS 36:5	DGTS 16:0/20:5	7,02
	DGTS 36:6	DGTS 20:5/16:1	2,13
	DGTS 40:10	DGTS 20:5/20:5	14,42
	DGTS 40:8	DGTS 20:4/20:4	4,48
	DGTS 40:9	DGTS 20:4_20:5	14,76
DAG	DAG 14:0/16:1	DAG 14:0/16:1	33,98
	DAG 16:0/16:1	DAG 16:0/16:1	25,97
	DAG 16:1/16:1	DAG 16:1/16:1	10,61
	DAG 20:5/16:0	DAG 20:5/16:0	18,54
	DAG 20:5/20:5	DAG 20:5/20:5	8,85
TAG	TAG 14:0/16:0/16:0	TAG 14:0/16:0/16:0	1,94
	TAG 14:0/16:1/14:0	TAG 14:0/16:1/14:0	1,05
	TAG 14:0/16:1/16:0	TAG 14:0/16:1/16:0	5,77
	TAG 14:0/16:1/16:1	TAG 14:0/16:1/16:1	4,69
	TAG 16:0/16:0/16:0	TAG 16:0/16:0/16:0	3,80
	TAG 16:0/16:0/16:1	TAG 16:0/16:0/16:1	25,05
	TAG 16:1/16:1/16:0	TAG 16:1/16:1/16:0	32,76
	TAG 16:1/16:1/16:1	TAG 16:1/16:1/16:1	17,26
	TAG 16:1/18:1/16:1	TAG 16:1/18:1/16:1	3,43

### 1.2.2.3 Variabilité des profils lipidiques de *M. gaditana*

Dans la section précédente un profil glycérolipidique détaillé de *M. gaditana* a été présenté. Ce profil a été obtenu avec une souche sauvage de *M. gaditana* cultivée en condition de photoautotrophie à 24°C (Billey et al. 2021). Cependant, de nombreuses études montrent que le profil lipidique des *Nannochloropsis* et *Microchloropsis*, tout comme pour les microalgues en général est fortement affecté par les facteurs environnementaux tels que la lumière (Kim et al. 2014; Janssen, Driessen et al. 2018b), la température (Ferrer-Ledo et al. 2023), la limitation en éléments nutritifs (notamment l'azote et le phosphate) (Simionato et al. 2013; Jouhet et al. 2017; Janssen, Driessen et al. 2018a; Matsui et al. 2020; Savvidou et al. 2020), le mode de culture: en hétérotrophie (Marudhupandi, Sathishkumar et Kumar 2016) et en mixotrophie (Fang et al. 2004) etc. Ces remaniements lipidiques sont une réponse adaptative des microalgues aux conditions environnementales fluctuantes. Cela se traduit non seulement par des variations dans la quantité des différentes classes de glycérolipides, mais aussi par des modifications significatives des profils des chaînes carbonées au sein de ces différentes classes lipidiques (Jouhet et al. 2017). Un exemple de l'impact d'une carence en azote sur le profil lipidique de *M. gaditana* est présenté dans la Figure 1.5.



**Figure 1.5. Remodelage lipidique en carence d'azote chez *M. gaditana* (D'après Jouhet et al. 2017).** Impact de la carence en azote sur la quantité d'acide gras totaux (A), le profil glycérolipidique (B) et les espèces lipidiques constituant différentes classes telles que MGDG (C), PC (D) et TAG (E).

Le profil lipidique de *M. gaditana* étant très variable il est important de pouvoir le caractériser dans différentes conditions. Cette nécessité est d'autant plus marquée que de nombreuses tentatives d'ingénierie génétique visent à modifier le profil des acides gras, par exemple en augmentant la proportion d'acides gras ayant un intérêt industriel. Cela souligne l'importance de la compréhension approfondie du métabolisme lipidique de *M. gaditana* et de la capacité à contrôler sa composition lipidique pour des applications biotechnologiques spécifiques.

#### 1.2.2.4 Méthode d'analyse des profils lipidiques au LPCV

L'équipe "Biogénèse, Dynamique et Homéostasie des lipides membranaires dans la cellule végétale" du LPCV dispose d'une plateforme lipidomique (LIPANG) permettant de réaliser des glycérolipidomes (catalogue des glycérolipides présents dans une espèce d'algue). Ces analyses reposent sur la quantification des acides gras par chromatographie en phase gazeuse (GC), et quantification des glycérolipides par chromatographie liquide à haute pression et spectrométrie de masse en tandem (HPLC-MS/MS). La combinaison de ces méthodes permet de quantifier avec précision les différentes molécules de glycérolipides présentes dans une souche (Jouhet et al. 2017).

Des glycérolipidomes de différentes souches sauvages et mutantes de *M. gaditana* cultivées dans différentes conditions sont ainsi disponibles au LPCV. De nouvelles données sont également générées régulièrement. L'interprétation de ces données nécessite une considération globale du métabolisme lipidique.

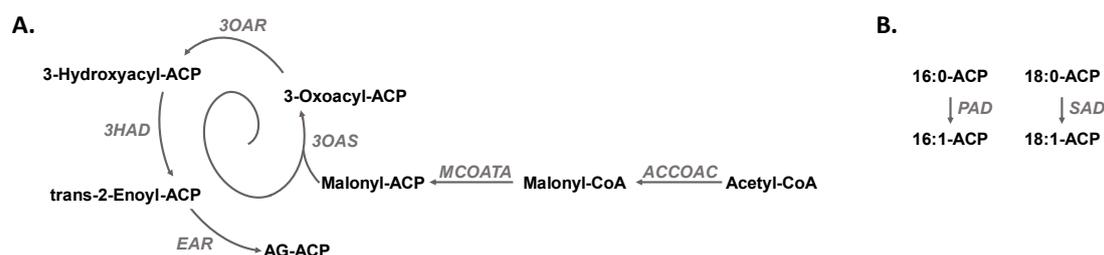
### 1.2.3 Etat de l'art du métabolisme lipidique chez *M. gaditana*

Malgré le potentiel industriel de *M. gaditana* pour diverses applications biotechnologiques, notamment la production de biodiesels, la connaissance de son métabolisme lipidique demeure relativement limitée. Les connaissances actuelles découlent majoritairement de l'analyse des génomes et transcriptomes de *M. gaditana* (Radakovits et al. 2012; Corteggiani Carpinelli et al. 2014) et d'analogies avec d'autres microalgues ainsi qu'avec des plantes supérieures. A partir de ces connaissances et analogies, les différentes voies de synthèse et de dégradation des acides gras et des glycérolipides ont été implantées dans le modèle métabolique au cours de la thèse. La section suivante présente le détail de ces réactions ainsi qu'un schéma récapitulatif à la fin. La stratégie de modélisation est présentée plus en détails dans le [Chapitre 2](#).

### 1.2.3.1 De la synthèse des acides gras à celle des glycérolipides

#### Synthèse chloroplastique des acyl-ACP

Le stroma du chloroplaste est le site principal de la synthèse des acides gras. Le pool d'acétyl-CoA (Coenzyme A) chloroplastique, produit par décarboxylation du pyruvate grâce à l'action du complexe pyruvate déshydrogénase (PDH), constitue le substrat primaire pour la synthèse d'acides gras (Figure 1.6.A). L'acétyl-CoA (deux atomes de carbone) est tout d'abord converti en malonyl-CoA (trois atomes de carbone) par une acétyl-CoA carboxylase (ACCOAC). La malonyl-CoA ACP malonyltransférase (MCOATA) catalyse ensuite la condensation du malonyl-CoA en malonyl-ACP (Acyl Carrier Protein). Les acyl-ACP (acides gras liés à l'ACP) sont synthétisés *de novo* par l'acide gras synthase II (FASII) via une succession de quatre réactions enzymatiques transférant deux carbones par cycle d'élongation. Le malonyl-ACP est condensé avec un acétyl-ACP par la 3-oxoacyl-ACP synthase (3OAS) formant un 3-oxoacyl-ACP. Le 3-oxoacyl-ACP est ensuite réduit par une 3-oxo-acyl réductase (3OAR) en 3-hydroxyacyl-ACP qui sert de substrat à une énoyl-déshydratase. Enfin, le 2,3-trans-énoyl-ACP produit est réduit par une énoyl-réductase (EAR) en un acyl-ACP contenant deux carbones supplémentaires par rapport au substrat initial. Chez *M. gaditana*, ce cycle produit essentiellement des acyl-ACP à 16 et 18 carbones (16:0-ACP et 18:0-ACP). Ces acyl-ACP peuvent subir une désaturation par une acyl-ACP désaturase stromale, conduisant à la production de 16:1-ACP et/ou de 18:1-ACP (Figure 1.6.B). Il existe ensuite deux voies majeures pour ces acyl-ACP. (1) Ils peuvent être incorporés dans la synthèse des glycérolipides chloroplastiques ou (2) être convertis en acide gras libres via une acyl-ACP hydrolase (FAACPE), puis exportés du chloroplaste vers le cytosol où ils sont allongés et désaturés (Li-Beisson et al. 2019).



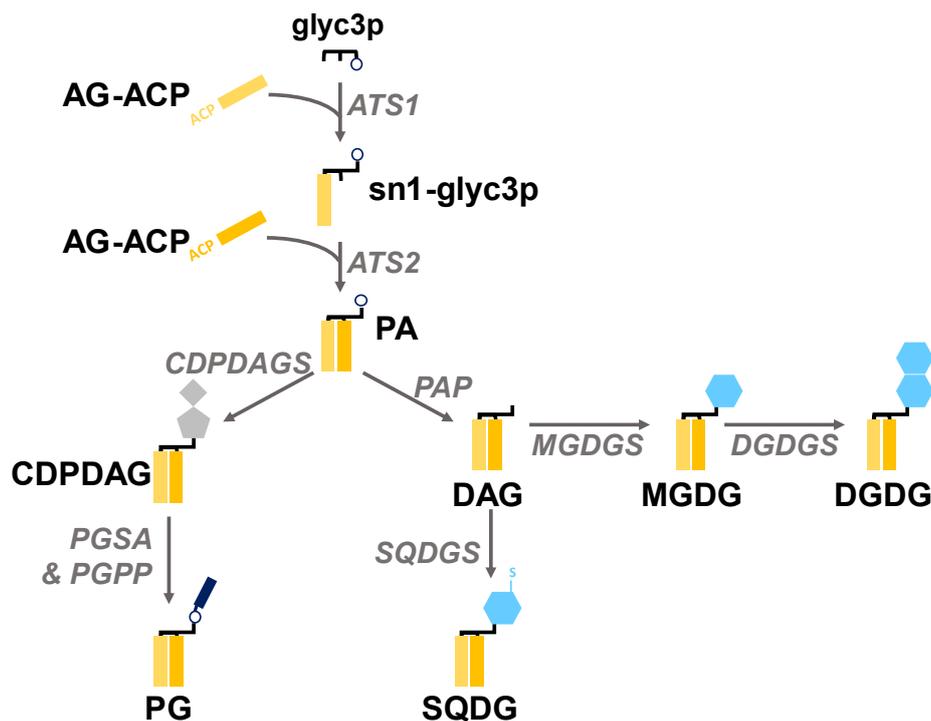
**Figure 1.6. Synthèse chloroplastique des acides gras.**

(A) Cycle de la synthèse des acides gras saturés. (B) Désaturation des acyl-ACP 16:0 et 18:0.

**Métabolites** : ACP : Acyl Carrier Protein, AG-ACP : acyl-ACP, CoA : Coenzyme A. **Réactions** : 3OAR : 3-oxo-acyl réductase, 3OAS : 3-oxoacyl-ACP synthase, ACCOAC : acétyl-CoA carboxylase, EAR : énoyl-réductase, MCOATA : malonyltransférase, PAD: Palmitic acid (16:0) désaturase, SAD: Stearic acid (18:0) désaturase.

## Synthèse des glycérolipides chloroplastiques

Quatre classes de glycérolipides sont synthétisées dans l'enveloppe du chloroplaste : le MGDG, le DGDG, le SQDG et le PG (Figure 1.7). La synthèse de ces glycérolipides est initiée par l'estérification successive de deux groupements acyle (provenant d'un d'acyl-ACP) en positions *sn-1*, puis *sn-2* d'un glycérol-3-phosphate (glyc3p) par des glycérol-3-phosphate acyl-ACP acyltransférases (ATS1 et ATS2) formant ainsi de l'acide phosphatidique (PA, son nom provenant du groupement phosphate en position *sn-3*). L'acide phosphatidique sert ensuite de précurseur pour les quatre classes de glycérolipides. La synthèse du PG est successivement catalysée par la CDP-DAG synthase (CDP-DAGS), la PG phosphate synthase (PGSA) et la PG phosphate phosphatase (PGPP). Le diacylglycérol (DAG) est synthétisé à partir du PA par une phosphatidate phosphatase plastidial (PAP), qui retire le groupement phosphate en position *sn-3*. Le DAG est ensuite le précurseur de la synthèse du SQDG et du MGDG, catalysée par la SQDG synthase (SQDGS) et la MGDG synthase (MGDGS). La formation du DGDG, à partir du MGDG, est catalysée par la DGDG synthase (DGDGS) (Li-Beisson et al. 2019).



**Figure 1.7. Représentation schématique de la synthèse plastidiale des glycérolipides (D'après Lupette 2018).**

**Métabolites** : ACP : Acyl Carrier Protein, AG-ACP : acyl-ACP, CDP-DAG : cytidine diphosphate diacylglycérol, DAG : diacylglycérol, DGDG : digalactosyldiacylglycérol, glyc3p : glycérol-3-phosphate, MGDG : monogalactosyldiacylglycérol, PG : phosphatidylglycérol, *sn1-glyc3p* : acide lysophosphatidique, SQDG : sulfoquinovosyldiacylglycérol. **Réactions**: ATS : acyl-ACP acyltransférases (1 et 2), CDP-DAGS : CDP-DAG synthase, DGDGS : DGDG synthase, MGDGS : MGDG synthase, PAP : phosphatidate phosphatase plastidial, PGSA : PG phosphate synthase, PGPP : PG phosphate phosphatase, SQDGS : SQDG synthase.

## Synthèse, élongation et désaturation cytosolique des acides gras

Les acyl-ACP chloroplastiques sont convertis en acides gras libres avant d'être exportés vers le cytosol (Li-Beisson et al. 2019). Par ailleurs, l'activation de la transcription d'une FAS cytosolique de type I a été signalée chez *M. gaditana* dans des conditions de stress telles qu'une forte exposition à la lumière, suggérant la possibilité d'une synthèse cytosolique des acides gras (Alboresi et al. 2016). La synthèse d'acide gras cytosolique, absente chez les plantes et les algues vertes, a donc été ajoutée au modèle. Les acides gras libres cytosoliques sont ensuite activés par thioestérification par les acyl-CoA synthétases (ACS) en présence d'ATP, formant un acyl-CoA, de l'AMP et du pyrophosphate (PPi) (Billey et al. 2021).

Le 16:0-CoA sert également de substrat initial pour une succession d'élongations et de désaturations formant d'abord le 18:0-CoA et aboutissant aux acides gras polyinsaturés à très longue chaîne (VLC-PUFA pour very-long chain polyunsaturated fatty acid) jusqu'à l'acide gras 20:5. A l'inverse de *Phaeodactylum tricornutum*, des plantes et des algues vertes, il n'y a pas de désaturation des C16 au-delà du 16:1 produit dans le stroma du chloroplaste. Au niveau du Réticulum Endoplasmique (RE), les élongations des acides gras à longue chaîne (entre 12 et 22 atomes de carbone) et très longues chaînes (plus de 22 atomes de carbone) sont réalisées au niveau des acyl-CoA (*i.e.* avant acylation au squelette glycérol) par les  $\Delta 0$ -ELO et  $\Delta 6$ -ELO élongases (Figure 1.8.A). Les désaturations, en revanche, ont lieu sur les chaînes carbonées des PC, PE et DGTS. Ces réactions sont catalysées par les  $\Delta 9$ ,  $\Delta 12$ ,  $\Delta 6$ ,  $\Delta 5$  et  $\Delta 15/\omega 3$  fatty acyl désaturase (FAD) (Figure 1.8.B) (Vieler et al. 2012; Dolch et al. 2017).



**Figure 1.8. Elongation et désaturation des acides gras.**

(A) Elongation des acyl-CoA 16:0 et 18:3. (B) Désaturation des chaînes carbonées estérifiées en position sn-2 (en vert) des PC, PE et DGTS. Elo: Elongase. FAD : fatty acid désaturase. Les positions des insaturations ne sont pas représentées.

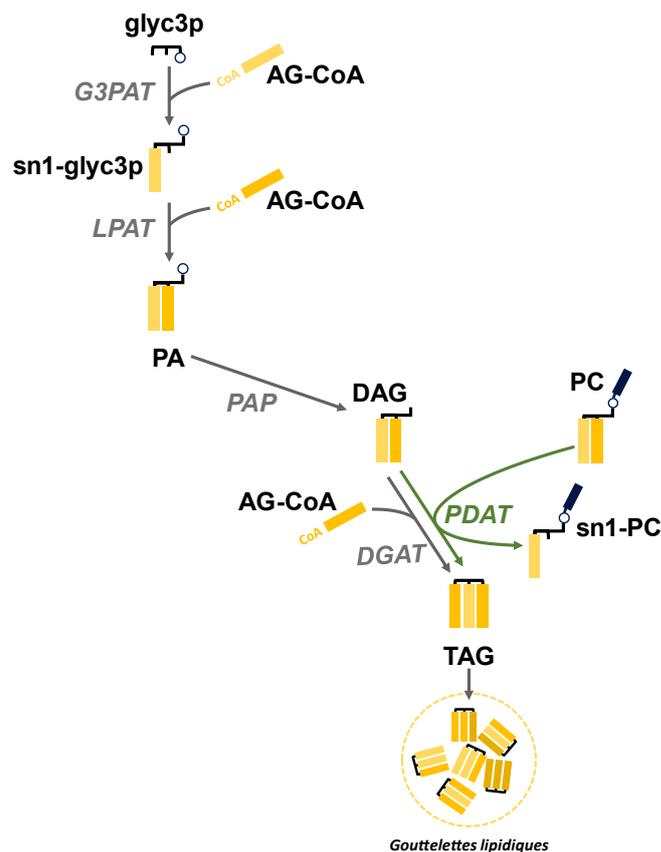
La réimportation des VLC-PUFA dans le chloroplaste est nécessaire à l'assemblage du MGDG, du DGDG, du SQDG et du PG. Ce processus, connu sous le nom de "voie oméga", a très peu été caractérisé (Petroutsos et al. 2014; Dolch et al. 2017). Dans le modèle ce processus a été modélisé par l'ajout de réactions de transports d'acides gras entre le cytosol et le chloroplaste.

## Synthèse des TAG

La synthèse des TAG peut s'effectuer par deux grandes voies distinctes au niveau du RE (Figure 1.9).

### (1) La voie de Kennedy :

Cette voie permet la synthèse *de novo* des TAG à partir d'acides gras issus du pool cytosolique d'acyl-CoA. Cette voie est initiée par l'estérification d'un groupement acyle (provenant d'un d'acyl-CoA) en position *sn-1* d'un glycérol-3-phosphate par la glycérol-3-phosphate acyltransférase (G3PAT). L'acide lysophosphatidique (*sn-1*-glyc3p) formé est ensuite de nouveau estérifié avec un acyl-CoA par une lysophosphatidate acyltransférase (LPAT) générant de l'acide phosphatidique (PA). Le DAG est ensuite formé par déphosphorylation du PA par une acide phosphatidique phosphatase (PAP). Enfin le TAG est produit par estérification d'un troisième groupement acyle par une diacylglycérol acyltransférase (DGAT) (Li-Beisson et al. 2019).



**Figure 1.9. Représentation schématique de la synthèse des triacylglycérols.**

La synthèse des TAG peut s'effectuer par deux grandes voies distinctes: la voie Kennedy (en gris) et la voie acyl-CoA indépendante (en vert). **Métabolites :** CoA : Coenzyme A, AG-CoA : acyl-CoA, DAG : diacylglycérol, glyc3p : glycérol-3-phosphate, PA : acide phosphatidique, PC : phosphatidylcholine, *sn-1*-glyc3p : acide lysophosphatidique, *sn-1*-PC : *sn-1* lyso-phosphatidylcholine, TAG : triacylglycérol. **Réactions:** DGAT : diacylglycérol acyltransférase, G3PAT : glycérol 3-phosphate acyltransférase, LPAT : lysophosphatidate acyltransférase, PAP : acide phosphatidique phosphatase, PDAT : phospholipid:diacylglycérol acyltransférase.

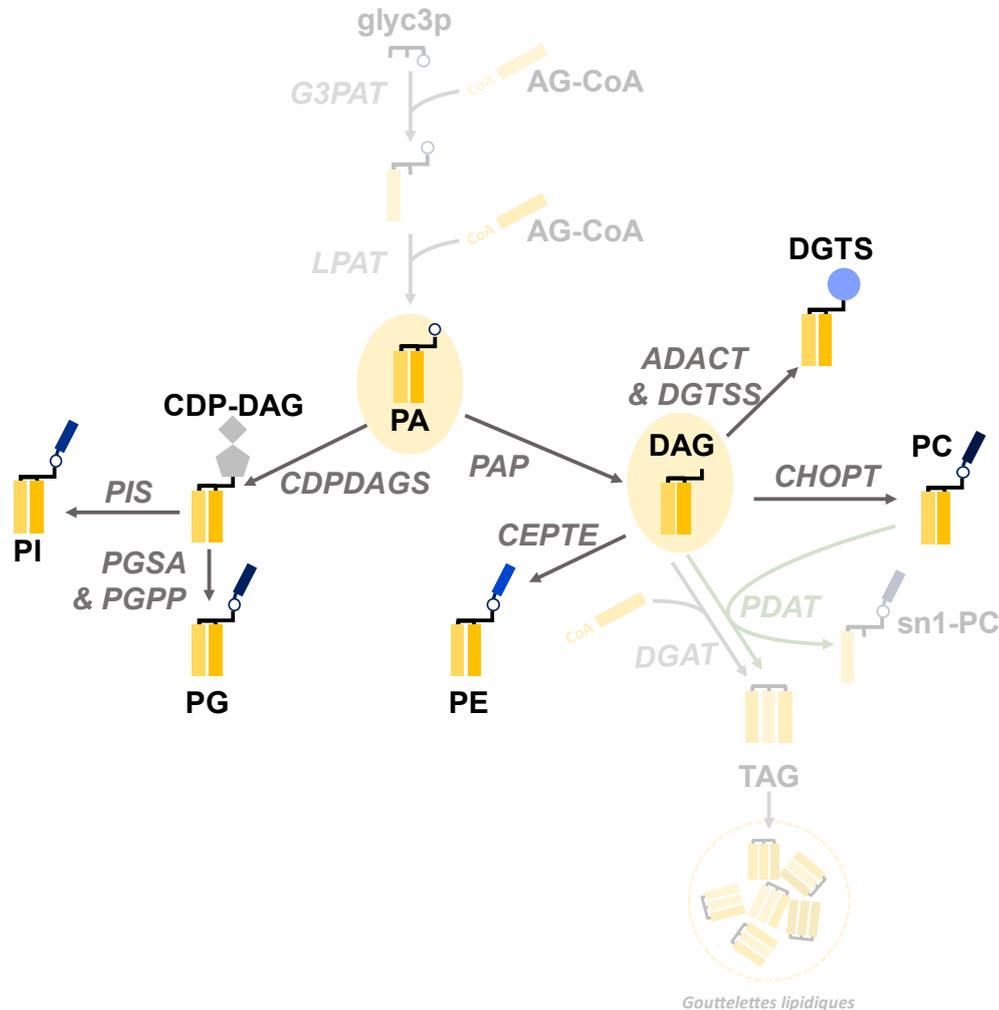
**(2) La voie acyl-CoA indépendante, utilisant la PC comme donneur d'acyl :**

La phospholipid:diacylglycérol acyltransférase (PDAT) transfère un acide gras en position *sn*-2 de la PC vers la position *sn*-3 du DAG formant d'une part une molécule de TAG et d'autre part une lysophosphatidylcholine (lyso-PC) en coproduit (Dahlqvist et al. 2000; Zienkiewicz et al. 2016). Cette voie alternative à la voie Kennedy reste assez peu caractérisée, notamment chez les *Heterokontes*.

**Synthèse des autres glycérolipides au niveau du Réticulum Endoplasmique**

Le PA formé via la voie Kennedy sert également de substrat pour la synthèse du cytidine diphosphate diacylglycérol (CDP-DAG) par une CDP-diacylglycérol synthase (CDPDAGS) et du DAG par une acide phosphatidique phosphatase (PAP) (Figure 1.10). Le CDP-DAG est ensuite utilisé pour la synthèse du PI par la phosphatidylinositol synthase (PIS) et celle du PG par la PG phosphate synthase (PGSA) et la PG phosphate phosphatase (PGPP). Le DAG sert de substrat à la synthèse de la PC et de la PE, respectivement par les enzymes CDP-choline:DAG choline phosphotransférase (CHOPT) et CDP-éthanolamine:DAG éthanolamine phosphotransférase (CEPT) (Li-Beisson et al. 2019).

Le DAG est également le précurseur de la synthèse du DGTS (Murakami et al. 2018). Cette synthèse nécessite deux étapes. Dans un premier temps, un groupement 3-amino-3-carboxypropyl est transféré de la S-adenosylméthionine au DAG formant du diacylglycérolhomosérine. Trois groupements méthyles sont ensuite successivement transférés de la S-adenosylméthionine à la diacylglycérolhomosérine formant ainsi le DGTS. Les enzymes S-adenosylméthionine:diacylglycerol 3-amino-3-carboxypropyl transférase (ADACT) et DGTS synthase (DGTSS) catalysent ces deux réactions.



**Figure 1.10. Représentation schématique de la synthèse des glycérolipides au niveau du RE.**

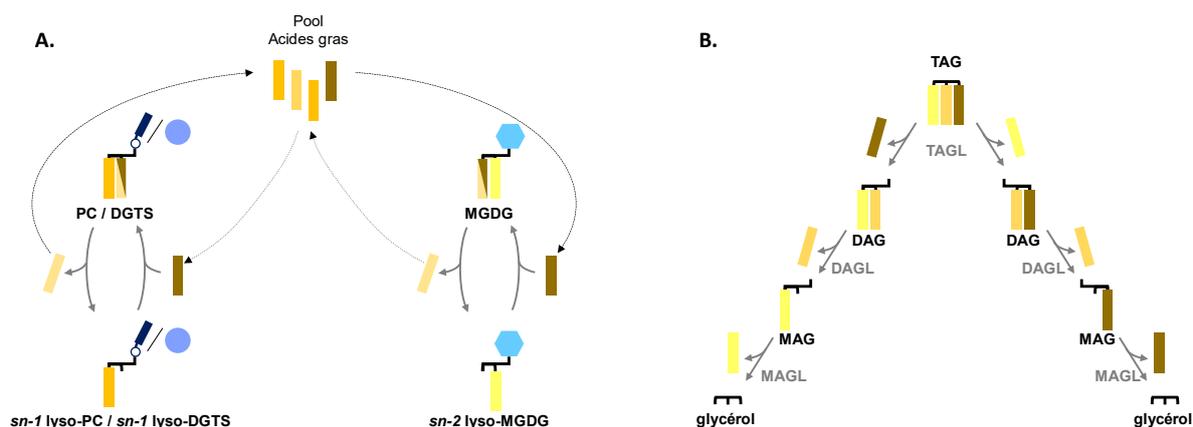
PA et DAG servent de précurseurs pour la synthèse des glycérolipides au niveau du RE (mis en évidence en beige). Les réactions présentées dans la Figure 1.9 sont mises en arrière plan. **Métabolites :** CoA : Coenzyme A, AG-CoA : acyl-CoA, CDP-DAG : cytidine diphosphate diacylglycérol, DAG : diacylglycérol, glyc3p : glycérol-3-phosphate, DGTS : 1,2-diacylglycéryl-3-O-4'-(N,N,N-triméthyl)-homosérine, PA : acide phosphatidique, PC : phosphatidylcholine, PE : phosphatidyléthanolamine, PI : phosphatidylinositol, PG : phosphatidylglycérol, sn1-glyc3p : acide lysophosphatidique, sn1-PC : sn-1 lyso-phosphatidylcholine, TAG : triacylglycérol. **Réactions:** ADACT : S-adenosylmethionine:diacylglycerol 3-amino-3-carboxypropyl transférase, CDPDAGS : CDP-diacylglycérol synthase, CEPTTE : CDP-éthanolamine:DAG éthanolamine phosphotransférase, CHOPT : CDP-choline:DAG choline phosphotransférase, DGAT : diacylglycérol acyltransférase, DGTSS : DGTS synthase, G3PAT : glycérol 3-phosphate acyltransférase, LPAT : lysophosphatidate acyltransférase, PAP : acide phosphatidique phosphatase, PDAT : phospholipid:diacylglycérol acyltransférase, PGPP : PG phosphate phosphatase, PGSA : PG phosphate synthase, PIS : phosphatidylinositol synthase.

### 1.2.3.2 Le catabolisme lipidique

#### Lipolyse et édition des acides gras

La croissance et la survie des microalgues dans des conditions environnementales fluctuantes nécessitent un remodelage ou un renouvellement permanent des lipides membranaires ainsi qu'une mobilisation rapide des lipides de stockage. Ainsi, la dégradation des lipides et l'édition des acides gras sont des processus importants chez ces organismes (Kong et al. 2018; Hoffmann et Shachar-Hill 2023). Le processus d'édition se réfère à la dé-acylation d'un groupement acyl d'un lipide donné et à sa ré-acylation sur un autre lipide (Figure 1.11A).

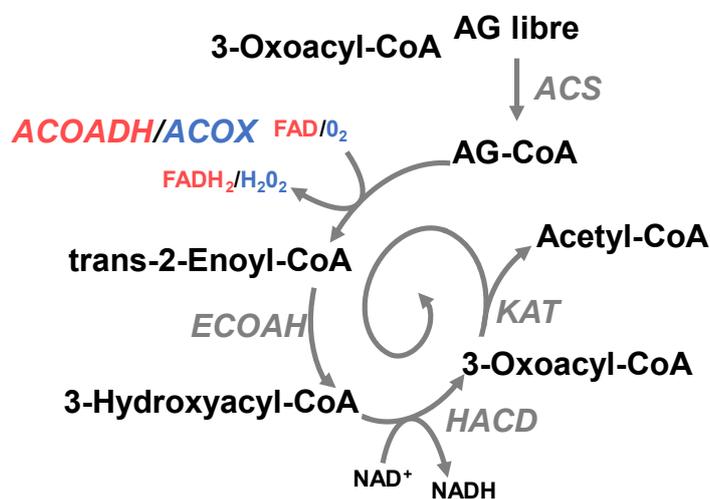
Les TAG, lipides de stockages, peuvent être catabolisés via trois hydrolyses successives aboutissant à la formation d'une molécule de glycérol et de trois acides gras (Figure 1.11B). La plupart des connaissances actuelles sur la dégradation des glycérolipides et l'édition des acyles ont été obtenues chez les plantes. Ces processus sont beaucoup moins caractérisés chez les microalgues (Kong et al. 2018; Hoffmann et Shachar-Hill 2023).



**Figure 1.11. Représentation schématique de l'édition des acides gras et lipolyse des triacylglycérols.** (A) Edition de l'acide gras estérifié en sn-2 des PC et DGTS. Edition de l'acide gras estérifié en sn-1 des MGDG. (B) Lipolyse des TAG. **Métabolites** : DAG : diacylglycérol, DGTS : 1,2-diacylglycéryl-3-O-4'-(N,N,N-triméthyl)-homosérine, MAG : monoacylglycérol, MGDG : monogalactosyldiacylglycérol, PC : phosphatidylcholine, sn-1 lyso-DGTS : sn-1 lyso-diacylglycérol-N,N,N-triméthylhomosérine, sn-1 lyso-PC : sn-1 lyso-phosphatidylcholine, sn-2 lyso MGDG : sn-2 lyso-monogalactosyldiacylglycérol, TAG : triacylglycérol. **Réactions** : DAGL : DAG lipase, MAGL : MAG lipase, TAGL : TAG lipase.

## $\beta$ -oxydation

La  $\beta$ -oxydation est une voie métabolique de dégradation des acides gras produisant de l'acétyl-CoA ainsi que du pouvoir réducteur. Chez les espèces *Nannochloropsis* et *Microchloropsis*, la  $\beta$ -oxydation s'effectue à la fois dans le peroxysome et la mitochondrie (Poliner et al. 2015; Li-Beisson et al. 2019). Les acides gras libres du cytosol sont donc importés dans ces compartiments où ils sont dans un premier temps activés par thioestérification par les acyl-CoA synthétases (ACS) (Figure 1.12). La dégradation s'effectue en quatre étapes. Elle commence par la déshydrogénation de l'acyl-CoA en trans-2-énoyl-CoA, soit par la réduction du dioxygène ( $O_2$ ) pour générer du peroxyde (ACOX dans le peroxysome), soit par la réduction du FAD en  $FADH_2$  (ACOADH dans la mitochondrie). Dans la mitochondrie, le  $FADH_2$  produit par cette réaction est oxydé par la chaîne respiratoire. Dans le peroxysome, cette réaction produit du peroxyde d'hydrogène ( $H_2O_2$ ) qui est dismuté en eau et oxygène par la catalase (CAT). Dans la mitochondrie et le peroxysome, le trans-2-énoyl-CoA est ensuite hydraté par une énoyl-CoA hydratase (ECOAH) formant le 3-hydroxyacyl-CoA. Le 3-hydroxyacyl-CoA est ensuite oxydé par une hydroxyacyl-CoA-déshydrogénase  $NAD^+$  dépendante (HACD), formant du 3-oxoacyl-CoA. Enfin, une acétyl-CoA acyltransférase ou 3-ketoacyl-CoA thiolase (KAT) clive le 3-oxoacyl-CoA en un acétyl-CoA et un acyl-CoA, deux carbones plus courts que le substrat initial. Ce composé peut de nouveau entrer dans le cycle de dégradation.

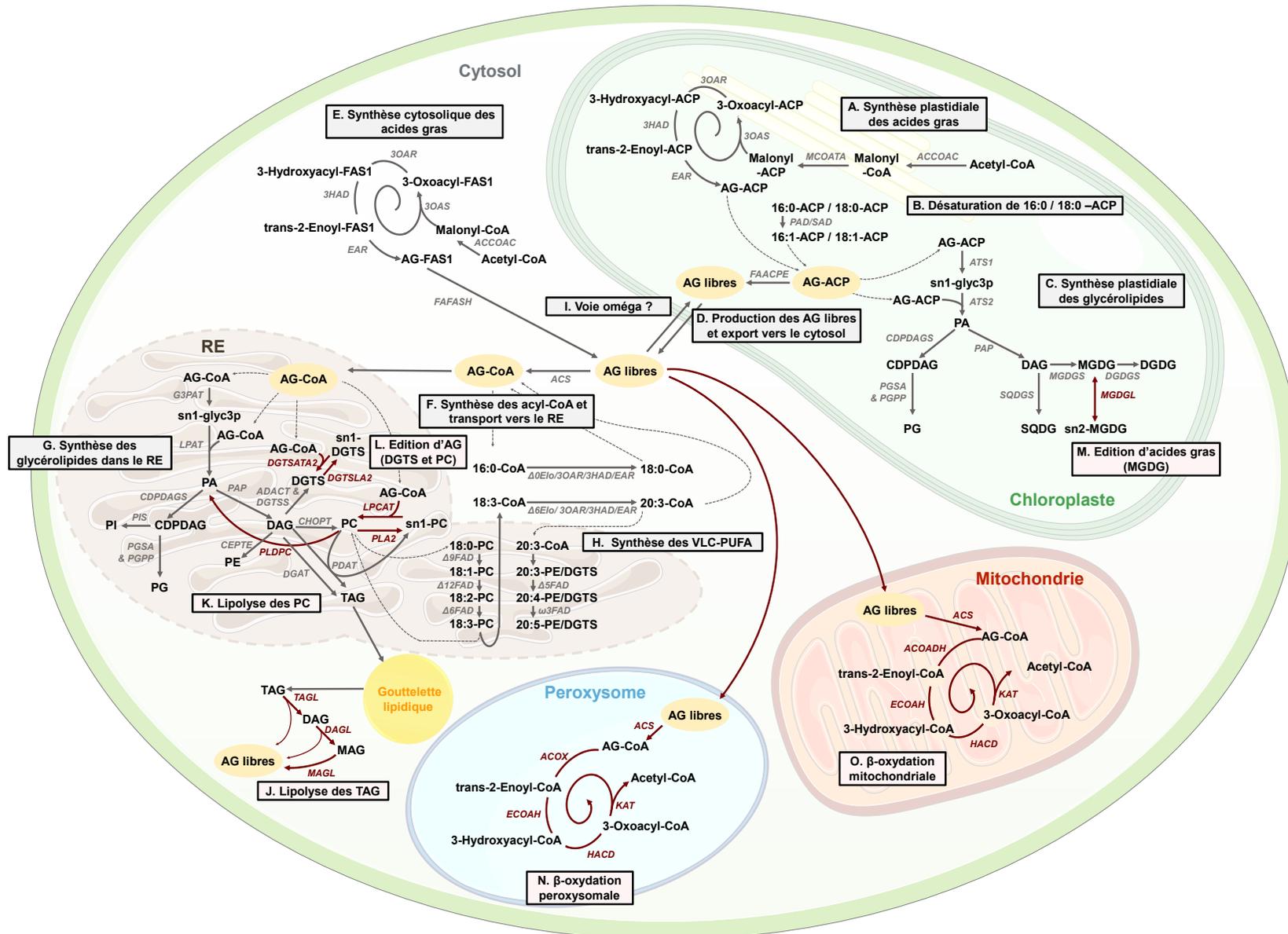


**Figure 1.12.  $\beta$ -oxydation des acides gras.**

La première étape du cycle de dégradation diffère entre la mitochondrie: ACOADH (rouge) et le peroxysome: ACOX (bleu). **Métabolites** : AG : acides gras, AG-CoA : acyl-CoA, CoA : Coenzyme A. **Réactions** : ACS: acyl-CoA synthétase, ACOADH : Acyl-CoA déshydrogénase, ACOX : Acyl-CoA oxydase, ECOAH : énoyl-CoA hydratase, HACD : hydroxyacyl-CoA-déshydrogénase, KAT : 3-ketoacyl-CoA thiolase.

Pour les acides gras insaturés, deux enzymes supplémentaires sont nécessaires, la cis- $\Delta^3$ -enoyl CoA isomérase qui convertit une liaison cis- $\Delta^3$  en une liaison trans- $\Delta^2$  (32ECOAI) et la 2,4-dienoyl CoA réductase qui réduit le 2,4-dienoyl-CoA en trans- $\Delta^3$ -enoyl-CoA (24E2COAR).

La Figure 1.13 résume de façon schématique le métabolisme lipidique supposé de *M. gaditana*.



**Figure 1.13. Bilan du métabolisme lipidique chez *M. gaditana*.**

Les flèches grises et rouges représentent respectivement la synthèse des lipides (A à I), la dégradation et l'édition des acides gras (L à O). (AG : Acides gras). Dans un souci de clarification, les réactions ayant lieu dans les membranes sont représentées dans les compartiments correspondants. Les acronymes sont définis dans les figures précédentes.

Le métabolisme lipidique décrit dans cette introduction repose principalement sur des analogies avec d'autres microalgues et des plantes supérieures ainsi que sur la caractérisation détaillée du glycérolipidome de *M. gaditana*. Plusieurs gènes codant pour des enzymes impliquées dans ces voies lipidiques ont été identifiés dans le génome de *M. gaditana* (Radakovits et al. 2012; Corteggiani Carpinelli et al. 2014). Cependant, il convient de noter que peu d'analyses de caractérisation fonctionnelle de ces enzymes ont été entreprises jusqu'à présent. Les caractérisations fonctionnelles reposent généralement sur l'analyse de souches mutantes. Malgré la disponibilité de plusieurs technologies d'ingénierie génétique chez *M. gaditana* (Poliner, Farré et Benning 2018; Kurita et al. 2023), la génération de souches mutantes reste une tâche complexe. Cette complexité est due en partie à l'annotation peu précise de certaines zones du génome de *M. gaditana* et au temps expérimental requis. En effet, au LPCV il faut en moyenne six à neuf mois pour générer et caractériser une souche mutante de *M. gaditana*.

Les modèles métaboliques à l'échelle du génome (GEM) se sont révélés être des outils bioinformatiques intéressants pour l'élucidation du métabolisme de nombreux organismes. Un modèle métabolique de *M. gaditana*, avec un métabolisme lipidique détaillé, constituerait un atout majeur pour approfondir la compréhension du métabolisme de cette souche.

### **1.3 Modèles métaboliques à l'échelle du génome**

Les modèles métaboliques à base de contraintes permettent une compréhension approfondie du métabolisme des microorganismes en simulant les flux intracellulaires au sein d'un réseau métabolique. Les divergences entre les prédictions de ces modèles et les données expérimentales sont utilisées pour concevoir des expériences ciblées visant à corriger ces imprécisions dans les connaissances métaboliques (Bordbar et al. 2014). Ces modèles présentent également des applications potentielles aussi bien pour la recherche fondamentale que pour des projets industriels (O'Brien, Monk et Palsson 2015; Ye et al. 2022; Cunha et al. 2023). Ils permettent notamment d'optimiser les souches et les conditions de culture.

Cependant le développement de tels modèles requiert souvent un long processus de corrections manuelles basé sur des connaissances et des données détaillées de l'organisme modélisé (Thiele et Palsson 2010).

## 1.3.1 Modéliser le métabolisme

### 1.3.1.1 Les réactions métaboliques : définitions

Les **réactions métaboliques** sont des transformations chimiques : des métabolites substrats réagissent entre eux pour former des produits. On représente généralement une réaction par son équation bilan en prenant en compte la stœchiométrie de la réaction c'est-à-dire les proportions dans lesquelles les métabolites sont consommés ou produits. Les réactions métaboliques sont soumises à la loi de conservation des masses. Il y a donc un équilibre entre la quantité de produit formé et de réactif consommé. La majorité des réactions métaboliques sont catalysées par des enzymes. Ces dernières sont des catalyseurs qui permettent d'accélérer, voire de rendre possibles les réactions du métabolisme. Certaines réactions sont réversibles grâce à l'action de la même enzyme ou grâce à une enzyme différente catalysant la réaction inverse. De plus, plusieurs enzymes peuvent catalyser la même réaction (isoenzymes). Inversement, une même enzyme peut avoir différents substrats et donc catalyser plusieurs réactions distinctes, entraînant ainsi de la redondance et de la complexité au sein du métabolisme.

Le métabolisme d'un organisme se compose de plusieurs centaines, voire de milliers de réactions métaboliques interconnectées. L'ensemble de ces réactions peut être représenté sous la forme d'un réseau métabolique. Pour chaque réaction, le flux métabolique associé correspond à la vitesse de synthèse de son ou ses produits. La fluxomique est une mesure de la dynamique du métabolisme et représente le résultat final de toutes les interactions métaboliques et régulatrices qui déterminent le comportement du métabolisme à l'échelle du microorganisme (ou d'une cellule, d'un tissu, d'un organisme pluricellulaire etc.). Il n'existe pas de méthode de mesure directe des flux métaboliques. Les méthodes d'analyses de fluxomique les plus abouties sont basées sur la combinaison d'expériences de marquage isotopique (le plus fréquemment au  $^{13}\text{C}$ ) et d'outils mathématiques permettant de calculer les vitesses des réactions à partir des métabolites marqués (Wiechert 2001; Antoniewicz 2015; Hendry et al. 2020). Cependant ces méthodes sont coûteuses, lourdes expérimentalement et ne sont applicables qu'à des réseaux de quelques dizaines (jusqu'à une centaine) de réactions. C'est précisément dans ce contexte que les modèles métaboliques jouent un rôle crucial, en permettant d'estimer ces valeurs de flux qui sont difficilement mesurables. Les formalismes les plus fréquemment utilisés pour modéliser le métabolisme reposent sur une description quantitative des réactions, soit par des équations différentielles ordinaires (EDO) soit par des méthodes basées sur des contraintes (Resat, Petzold et Pettigrew 2009; Orth, Thiele et Palsson 2010; Clark et al. 2020).

### 1.3.1.2 Modélisation du métabolisme par équations différentielles

Les EDO permettent de décrire l'évolution d'un système au cours du temps, en décrivant la variation des concentrations des différents substrats, produits, enzymes (Resat, Petzold et Pettigrew 2009). Ce type de formalisme se révèle très efficace pour la modélisation d'un espace métabolique restreint, tel que l'étude d'un sous-système métabolique ou d'un modèle simplifié. Cependant, son application à l'échelle d'un organisme entier reste encore limitée. Cette limitation découle du fait que la modélisation d'un système métabolique à l'aide des EDO nécessite la caractérisation d'un grand nombre de paramètres et de constantes enzymatiques, qui varient en fonction des conditions étudiées (paramètres dynamiques). Dans le cadre des approches de modélisation à grande échelle, le nombre de paramètres dynamiques à prendre en compte devient souvent trop important pour pouvoir être évalué. Or il est souvent difficile d'obtenir les données expérimentales nécessaires pour estimer les paramètres cinétiques de l'ensemble des enzymes d'un organisme (Mannan et al. 2015). Enfin, les systèmes d'équations différentielles deviennent difficiles à résoudre numériquement lorsque le réseau métabolique devient trop conséquent.

### 1.3.1.3 Modélisation à base de contraintes

La modélisation à base de contraintes prédit le phénotype sous la forme de flux métaboliques. Chaque réaction du réseau métabolique possède une valeur numérique qui représente la quantité de matière convertie par unité de temps et de matière (e.g. matière sèche). Ce formalisme repose sur la stœchiométrie des réactions et sur les taux d'absorption/sécrétion des métabolites extracellulaires, en supposant que le système se trouve dans un état stationnaire : les concentrations de chaque métabolite à l'intérieur du système sont considérées comme constantes. Ainsi, contrairement aux modèles basés sur les EDO, la modélisation à base de contraintes ne nécessite pas la caractérisation biologique de nombreux paramètres cinétiques. Ces approches de modélisation par contraintes ont donc été largement utilisées pour modéliser le métabolisme à grande échelle, notamment pour modéliser le métabolisme des microorganismes à l'échelle du génome (O'Brien, Monk et Palsson 2015; Gu et al. 2019).

Les méthodes à base de contraintes ont également été utilisées pour modéliser des organismes pluricellulaires tels que les plantes ou l'homme, ainsi que des microbiomes (une communauté de microorganismes) (Gu et al. 2019). Dans cette thèse nous nous intéresserons plus particulièrement aux modèles de microorganismes unicellulaires. Le processus de reconstruction et d'exploitation de ces modèles ainsi que certaines de leurs applications sont détaillées dans la section suivante.

**N.B.** : Dans la suite de ce manuscrit le terme "modèle métabolique" fera référence aux modèles à base de contraintes.

## 1.3.2 Les modèles métaboliques à l'échelle du génome

### 1.3.2.1 Les réseaux métaboliques à l'échelle du génome

#### **Contenu et organisation du réseau métabolique**

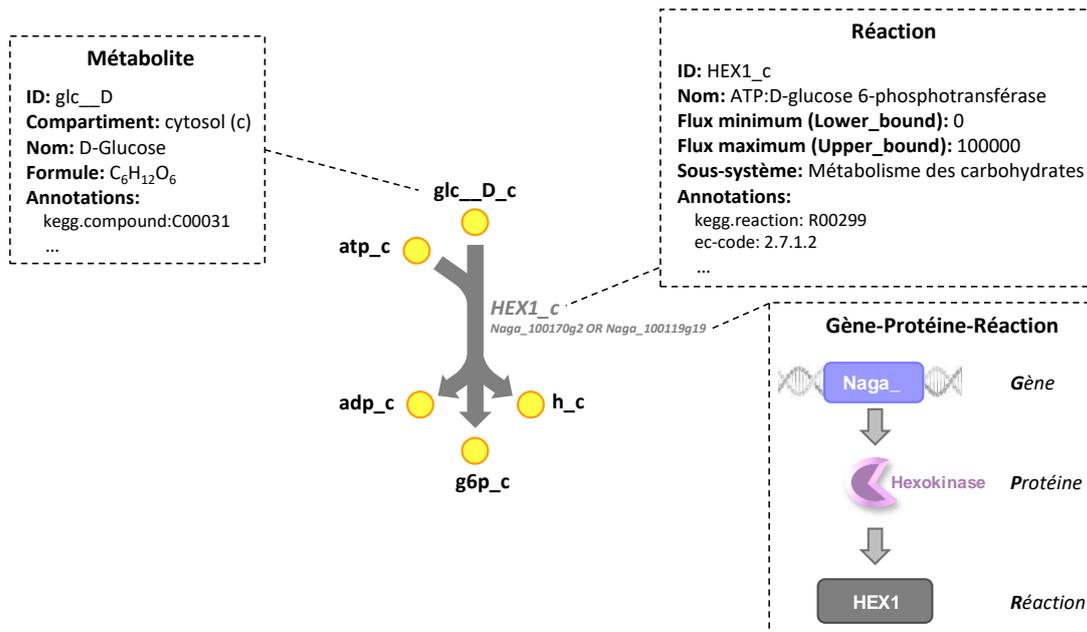
Un réseau métabolique à l'échelle du génome décrit l'intégralité des fonctions métaboliques codées par le génome d'un organisme d'intérêt. Il englobe ainsi l'ensemble des réactions métaboliques associées à cet organisme en mettant en relation les gènes codant pour les enzymes avec les réactions métaboliques qu'elles catalysent. Chaque réaction peut être associée à un ou plusieurs gènes via une relation Gène-Protéine-Réaction (GPR). Des gènes codant des isoenzymes sont reliés par l'opérateur logique "OR", tandis que les gènes codant les protéines d'un complexe enzymatique sont séparés par l'opérateur logique "AND".

Un réseau rassemble et organise ainsi dans un format structuré un ensemble de connaissances synthétiques sur les capacités métaboliques d'un organisme. Pour chaque réaction, le réseau comporte les informations permettant de répondre à différentes questions :

- (1) Quels sont les substrats et les produits de la réaction ?
- (2) Quels sont les coefficients stœchiométriques pour chaque métabolite ?
- (3) La réaction est-elle réversible ?
- (4) Dans quel compartiment cellulaire la réaction se produit-elle ?
- (5) Quel(s) gène(s) code(nt) pour la protéine (ou le complexe protéique) associé à la réaction ?

Toutes ces informations sont rassemblées à partir d'un éventail de sources, y compris des bases de données spécifiques aux organismes, des données à haut débit et de la littérature (Durot, Bourguignon et Schachter 2009; Thiele et Palsson 2010). Le processus de reconstruction des réseaux métaboliques est présenté dans la section "1.3.2.2 Reconstruction d'un réseau métabolique à l'échelle du génome".

La Figure 1.14 illustre l'ensemble des informations associées à un exemple de réaction métabolique : la première étape de la glycolyse qui convertit le glucose en glucose-6-phosphate, catalysée par des hexokinases.



**Figure 1.14. Représentation d'une réaction métabolique.**

Le réseau rassemble des informations relatives aux métabolites (ronds) et aux réactions (flèches). Les règles Gène-Protéine-Réaction associent des gènes aux réactions du réseau. La représentation est inspirée de celle utilisée par Escher, une application qui sera utilisée ultérieurement dans ce manuscrit (King et al. 2015).

### Compartmentation du réseau

Afin de mieux représenter ce qui se passe au sein des cellules, les réseaux métaboliques sont compartimentés. Pour les organismes eucaryotes, les modèles comportent généralement un milieu extracellulaire, le cytosol ainsi que des compartiments représentant les différents organites : mitochondrie, peroxysome, noyau, réticulum endoplasmique (RE), appareil de Golgi, vacuole, chloroplaste etc. (Duarte, Herrgård et Palsson 2004; Tibocha-Bonilla et al. 2018). La compartimentation d'un réseau nécessite l'affectation des métabolites et des réactions dans les différents compartiments et l'ajout de réactions de transport permettant le passage des métabolites d'un compartiment à l'autre (Klitgord et Segrè 2010; Seif et Palsson 2021).

### Définition du système modélisé

Le modélisateur est libre de définir les limites du système étudié. Ces limites définissent ce qui est inclus dans le modèle et ce qui en est exclu. Pour les réseaux métaboliques de microorganismes le système considéré est généralement "un microorganisme dans son milieu de culture" (Figure 1.15A). Les métabolites peuvent entrer et sortir du système par des pseudos-réactions ne respectant pas la loi de conservation des masses. Ces réactions sont conceptuelles, créées uniquement à des fins de modélisation, et n'ont pas de correspondance biologique (Thiele et Palsson 2010). On distingue notamment les réactions d'échange et les réactions de demande.

- Réactions d'échange :

Ces réactions définissent l'environnement nutritionnel extracellulaire en permettant un afflux ou un efflux de métabolites dans le système, par exemple pour réapprovisionner un réseau métabolique en matières premières (telles que les sources de carbone ou d'énergie) (Figure 1.15B). Par convention, un flux négatif à travers une réaction d'échange représente l'absorption du métabolite correspondant, et un flux positif représente sa sécrétion. Les réactions d'échange peuvent fonctionner dans le sens négatif (absorption) ou positif (sécrétion). Ces réactions d'échange peuvent être laissées "ouvertes" (pour permettre à un métabolite d'entrer dans le système à un taux spécifié) ou "fermées" (le métabolite ne peut que quitter le système) à l'aide de contraintes.

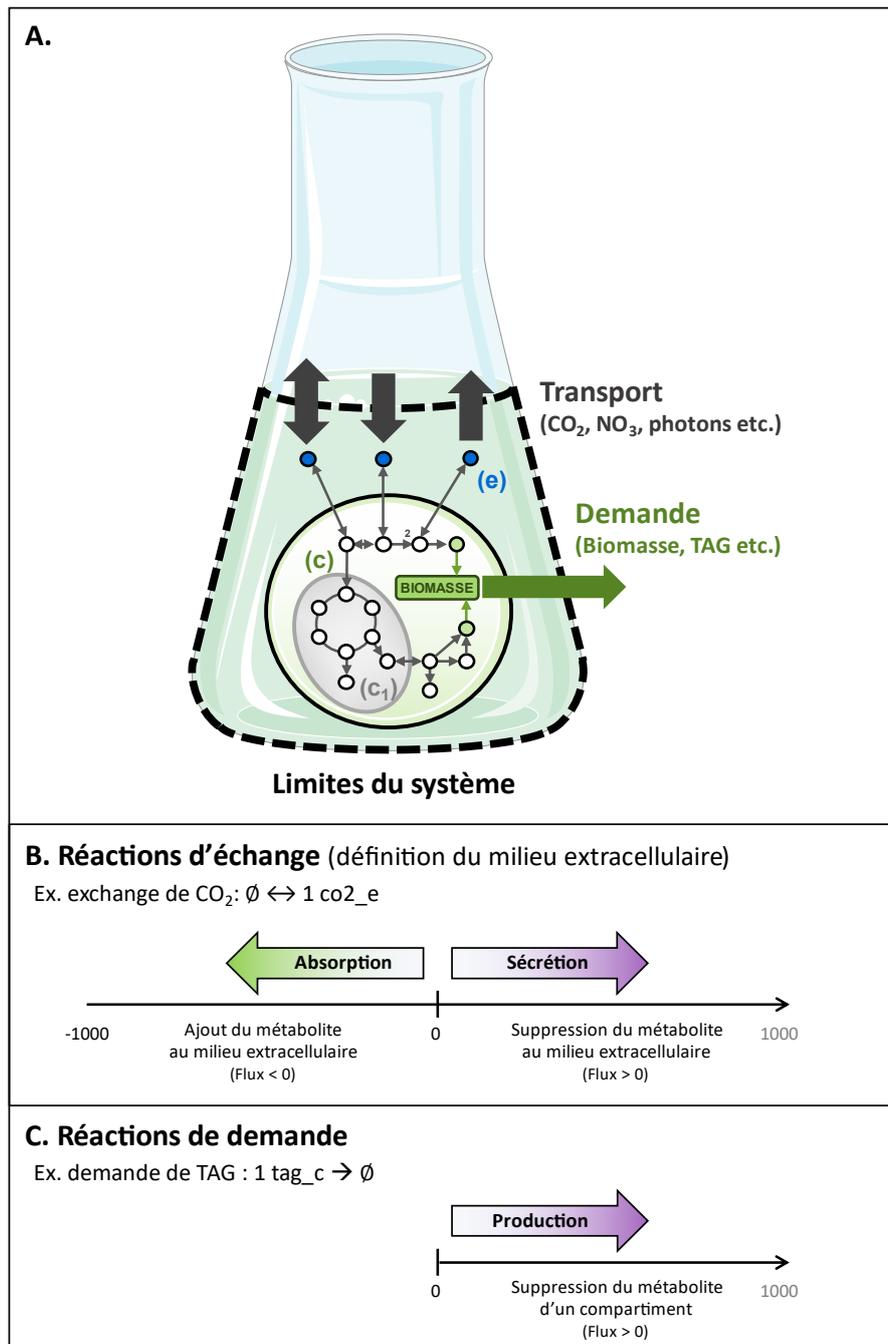
- Réactions de demande :

Le processus de modélisation est analogue à celui d'une réaction d'échange, mais opérant sur des métabolites "internes" c'est-à-dire localisés dans tout autre compartiment que le compartiment extracellulaire (Figure 1.15C). Les métabolites consommés par ces réactions sont supposés être utilisés dans des processus intracellulaires qui ne sont pas représentés explicitement dans le modèle (par exemple, la réaction de demande pour le pseudo-métabolite représentant la biomasse, ou un pseudo-métabolite représentant une molécule de stockage etc). Ces réactions diffèrent des réactions d'échange par le fait qu'elles sont irréversibles et que les métabolites ne sont pas prélevés dans l'environnement extracellulaire, mais dans l'un des compartiments de l'organisme.

Les réactions d'échange et de demande diffèrent des réactions de transport qui permettent le passage d'un métabolite d'un compartiment à un autre et qui respectent donc la loi de conservation des charges et des masses.

- Définition du système modélisé :

Prenons l'exemple de la modélisation d'un microorganisme tel qu'une microalgue. Le système peut être défini comme une microalgue évoluant dans son milieu de culture (Figure 1.15B). Ce système est divisé en plusieurs compartiments, notamment le compartiment extracellulaire, le cytosol de la microalgue, et éventuellement d'autres compartiments intracellulaires (mitochondrie, chloroplaste, peroxyosome etc.). La composition du milieu de culture est déterminée par les réactions d'échange, qui spécifient la disponibilité des différents substrats.



**Figure 1.15. Représentation du système modélisé (Adaptée de Thiele et Palsson 2010).**

(A) Définition des compartiments et des limites du système. Le réseau schématisé comporte trois compartiments : le milieu extracellulaire (e) qui correspond au milieu de culture, le cytosol (c) et un organite ( $c_1$ ) (qui correspond par exemple à la mitochondrie ou au chloroplaste etc.). Les limites du système sont représentées par les lignes noires pointillées. Pour chaque métabolite pouvant être consommé ou sécrété par le microorganisme (par exemple l' $\text{O}_2$ , le  $\text{CO}_2$ , le glucose ou autre source de carbone, les photons etc), une réaction d'échange doit être définie (flèche grise foncée). Ces réactions d'échange permettent de définir la composition du milieu de culture. Des réactions de demande peuvent également être définies pour les métabolites produits par l'organisme que l'on souhaite faire sortir du système (par exemple le métabolite représentant la biomasse, ou des métabolites de stockage tels que les TAG) (flèche verte). Les conventions de définition des réactions d'échange (B) et de demande (C) sont illustrées.

La modélisation d'un organisme photoautotrophe, qui effectue la photosynthèse, nécessite la prise en compte de l'énergie lumineuse reçue par l'organisme. Un photon est une particule élémentaire de lumière, dépourvue de masse, transportant une quantité discrète d'énergie. Dans les GEM les photons sont modélisés dans les réactions où ils interviennent de la même manière que les métabolites classiques, afin de prendre en compte leur apport énergétique. Par exemple, la réaction d'oxydation de la plastocyanine et de réduction de la ferrédoxine catalysée par le photosystème I (PSI) peut être modélisé par la réaction suivante :



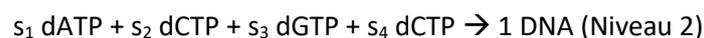
### La fonction de biomasse

Les réseaux métaboliques comportent généralement une (ou plusieurs) réactions de biomasse, également appelées fonctions objectives de biomasse. Une réaction de biomasse simule la consommation des métabolites permettant la croissance d'un organisme donné, pondérés pour représenter la composition d'un gramme de matière sèche (Figure 1.16). Cette réaction est utilisée pour représenter la croissance du microorganisme d'intérêt.

Il est possible de modéliser une fonction de biomasse en utilisant une pseudo-réaction unique consommant l'ensemble des métabolites dans les proportions appropriées. Cependant, il est courant de représenter la fonction de biomasse par plusieurs pseudo-réactions organisées en deux ou trois niveaux. Le premier niveau consomme les macromolécules principales de la cellule, telles que l'ADN, l'ARN, les protéines, les carbohydrates, les lipides, etc. :



où  $s_i$  (souvent exprimé en  $\text{mmol.gDW}^{-1}$ ) correspond à l'abondance de la macromolécule (ou du pool de molécules)  $i$  dans la biomasse cellulaire. ADN, ARN, PROTÉINES, etc. sont ici des pseudo-métabolites. Le niveau 2 correspond à la consommation des monomères (les désoxynucléotides triphosphates (dNTPs), les acides aminés etc.), symbolisant la polymérisation des macromolécules. À titre d'exemple, l'ADN est un pseudo-métabolite produit par une pseudo-réaction qui consomme les quatre monomères de dNTPs :



où  $s_i$  désigne la fraction molaire de chaque métabolite au sein de la macromolécule (ici en  $\text{mol.mol}_{\text{ADN}}^{-1}$ ).

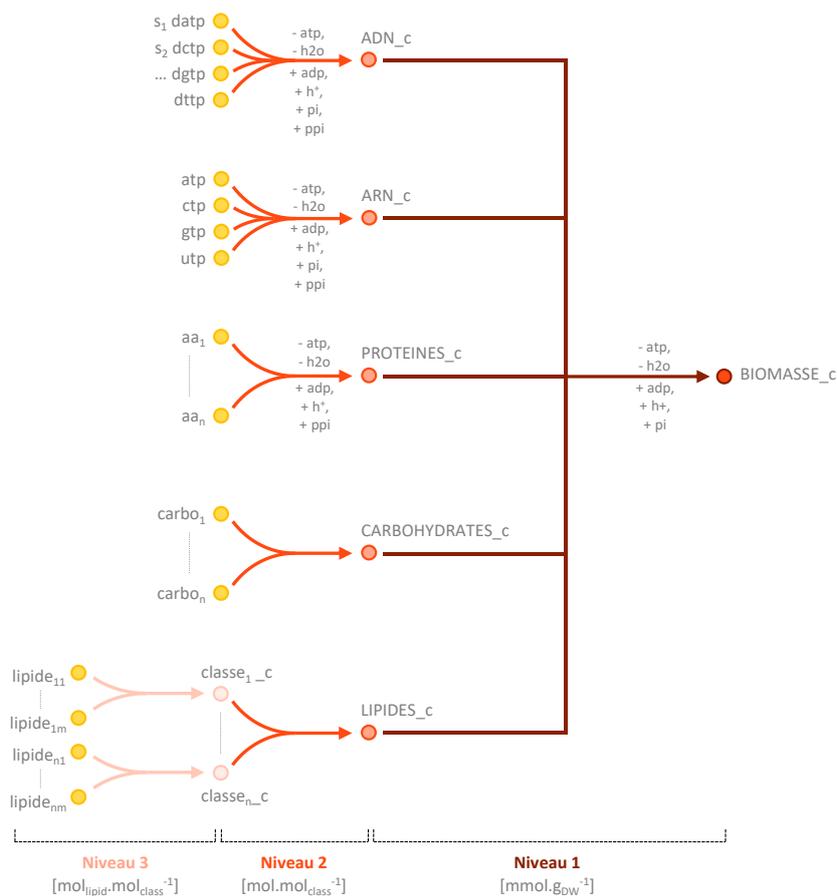
De par la multitude des entités lipidiques existantes, la formulation des pseudo-réactions lipidiques est généralement complexe. Une solution fréquemment utilisée consiste à introduire des pseudo-réactions supplémentaires afin de regrouper les entités lipidiques par classe :



où  $s_i$  représente la fraction molaire de la classe de lipides dans l'ensemble des lipides ( $\text{mol}_{\text{classe}} \cdot \text{mol}_{\text{LIPIDES}}^{-1}$ ). PC (phosphatidylcholine), PE (phosphatidyléthanolamine), etc. sont des pseudo-métabolites représentant chaque classe de lipides qui doivent à leur tour être produits par des pseudo-réactions supplémentaires. Par exemple, pour PC :



où  $s_i$  représente la fraction molaire de chaque lipide spécifique dans sa classe (ici en  $\text{mol} \cdot \text{mol}_{\text{PC}}^{-1}$ ).



**Figure 1.16. Exemple de fonction objective de biomasse.**

La fonction objective de biomasse agit comme un "drain" sur les composants cellulaires. Elle est composée de tous les métabolites nécessaires à la création d'un nouveau microorganisme, y compris l'ADN, les acides aminés, les carbohydrates, lipides etc., pondérés en fonction de données expérimentales et pouvant être organisés en plusieurs niveaux. Dans les simulations FBA, la fonction biomasse est utilisée pour simuler la croissance.

La fonction de biomasse consomme ainsi tous les composés nécessaires à la création d'une nouvelle cellule, y compris l'ADN, les acides aminés, les lipides et les polysaccharides. Les fonctions de biomasse peuvent également incorporer des détails sur l'énergie (ATP), les vitamines et les cofacteurs nécessaires à la croissance (Feist et Palsson 2010).

### 1.3.2.2 Reconstruction d'un réseau métabolique à l'échelle du génome

Afin d'homogénéiser les reconstructions, Thiele et Palsson ont publié un protocole complet de reconstruction de réseaux métaboliques (Thiele et Palsson 2010). Le processus est divisé en quatre étapes: (1) la construction d'une ébauche métabolique, (2) l'amélioration de cette ébauche, (3) la conversion du réseau en modèle mathématique et l'ajout de contraintes, et (4) l'évaluation de la qualité du modèle final (Figure 1.17). Afin d'étudier la fonctionnalité d'un réseau ce dernier doit être converti en modèle (étapes 3 et 4), une première itération complète du cycle est ainsi nécessaire. Les étapes 2 à 4 seront itérées jusqu'à ce que les prédictions réalisées par le réseau métabolique soient en accord avec les observations biologiques.

Ces différentes étapes sont présentées plus en détail dans la suite du manuscrit.

#### **Etape 1 : Construction d'une ébauche métabolique**

La première étape de la reconstruction d'un réseau métabolique consiste à élaborer une ébauche (ou "draft" en anglais) du réseau (Figure 1.17A). Un réseau de réactions métaboliques brut est généré à partir des informations extraites de l'annotation du génome de l'espèce étudiée et/ou issues de l'étude de réseaux métaboliques d'espèces phylogénétiquement proches.

Certaines réactions existantes (et fonctionnelles) chez l'organisme d'intérêt pourront être absentes ou erronées dans cette ébauche en raison d'annotations de gènes manquantes, erronées ou incomplètes. La qualité de l'annotation du génome a donc un fort impact sur la qualité du réseau métabolique reconstruit. En effet, l'objectif de l'annotation est de proposer une fonction potentielle pour chaque gène d'intérêt présent dans le génome en utilisant par exemple des mots clés ou des catégories de Gene Ontology (GO). A partir de l'annotation, on peut ensuite identifier les réactions métaboliques que l'organisme est capable de réaliser en associant des réactions métaboliques précises à chaque gène annoté, par exemple via l'association d'un nombre Enzyme Commission Number (EC Number) (Egelhofer, Schomburg et Schomburg 2010). Cette étape repose sur l'utilisation de bases de données de réactions métaboliques telles que KEGG REACTION (Kanehisa et Goto 2000b), MetaCyc (Caspi et al. 2018; Caspi et al. 2020) et/ou de modèles métaboliques BiGG Models etc. (Kanehisa et Goto 2000b; King et al. 2016; Norsigian et al. 2020). Elle peut être réalisée automatiquement et/ou manuellement. Différents outils et logiciels ont été développés pour faciliter

et accélérer le processus de reconstruction métabolique tels que AuReMe (Aite et al. 2018), CarveMe (Machado et al. 2018), FAME (Boele, Olivier et Teusink 2012), Merlin (Dias et al. 2015), ModelSEED (Seaver et al. 2021), Pathway Tools (Karp et al. 2021), RASP (Metcalf, Nagygyor et Boyle 2020), RAVEN (Wang et al. 2018) etc. Cette étape de reconstruction du métabolisme appliquée à la microalgue *M. gaditana* a été réalisée en amont de la thèse par l'entreprise SilicoLife (<https://www.silicolife.com/>) via une méthode interne. Les différents outils de reconstructions de réseaux métaboliques ne seront donc pas décrits dans ce manuscrit. Des revues détaillant et comparant ces différents outils sont disponibles dans la littérature (Mendoza et al. 2019; Passi et al. 2021; Cunha et al. 2023).

## **Etape 2 : Amélioration du réseau**

L'ébauche initiale d'un réseau métabolique, souvent générée à partir de données d'annotations automatiques, contient généralement des réactions incorrectes ou manquantes en raison d'erreurs d'annotations ou d'annotations incomplètes. Par conséquent, une étape de correction manuelle est essentielle pour améliorer la qualité du réseau métabolique. Cette étape est souvent basée sur la connaissance d'experts du métabolisme de l'organisme d'intérêt, des recherches bibliographiques ainsi que sur des données expérimentales. Les bases de données telles que BRENDA (Schomburg et al. 2000; Chang et al. 2021), BiGG Models (King et al. 2016; Norsigian et al. 2020), Chebi (Degtyarenko et al. 2008; Hastings et al. 2016), Ensembl (Martin et al. 2023), KEGG (Kanehisa et Goto 2000b), LIPID MAPS (Fahy et al. 2007), MetaCyc (Caspi et al. 2018; Caspi et al. 2020), LIPID MetaNetX (Bernard et al. 2014; Moretti et al. 2021), SwissLipids (Aimo et al. 2015), Rhea (Bansal et al. 2022), UniProt (Consortium 2022) etc. constituent également des sources d'informations importantes dans cette étape de correction pour obtenir des informations sur les gènes, les enzymes, les réactions, les métabolites et les voies métaboliques. Par ailleurs, des bases de données spécifiques à un organisme, telles que PlantCyc pour les algues et les plantes (Hawkins et al. 2021), ou encore ChloroKB pour *Arabidopsis thaliana* (Gloaguen et al. 2017; Gloaguen et al. 2021) peuvent aussi être utilisées. Ces bases de données, qui comportent des informations souvent élaborées et axées sur l'organisme cible, présentant ses spécificités métaboliques et physiologiques connues, sont ainsi des sources d'informations importantes.

Cette étape d'amélioration repose sur plusieurs éléments clés (Figure 1.17B).

### Vérification de l'équilibre des charges et des masses

Une formule et une charge doivent être ajoutées à chaque métabolite du modèle. Le bilan massique de chaque réaction doit être effectué afin de respecter la loi de conservation des masses : le nombre d'atomes dans les réactifs doit être égal au nombre d'atomes dans les produits. De même pour la charge des réactions : les équations équilibrées doivent avoir une charge neutre.

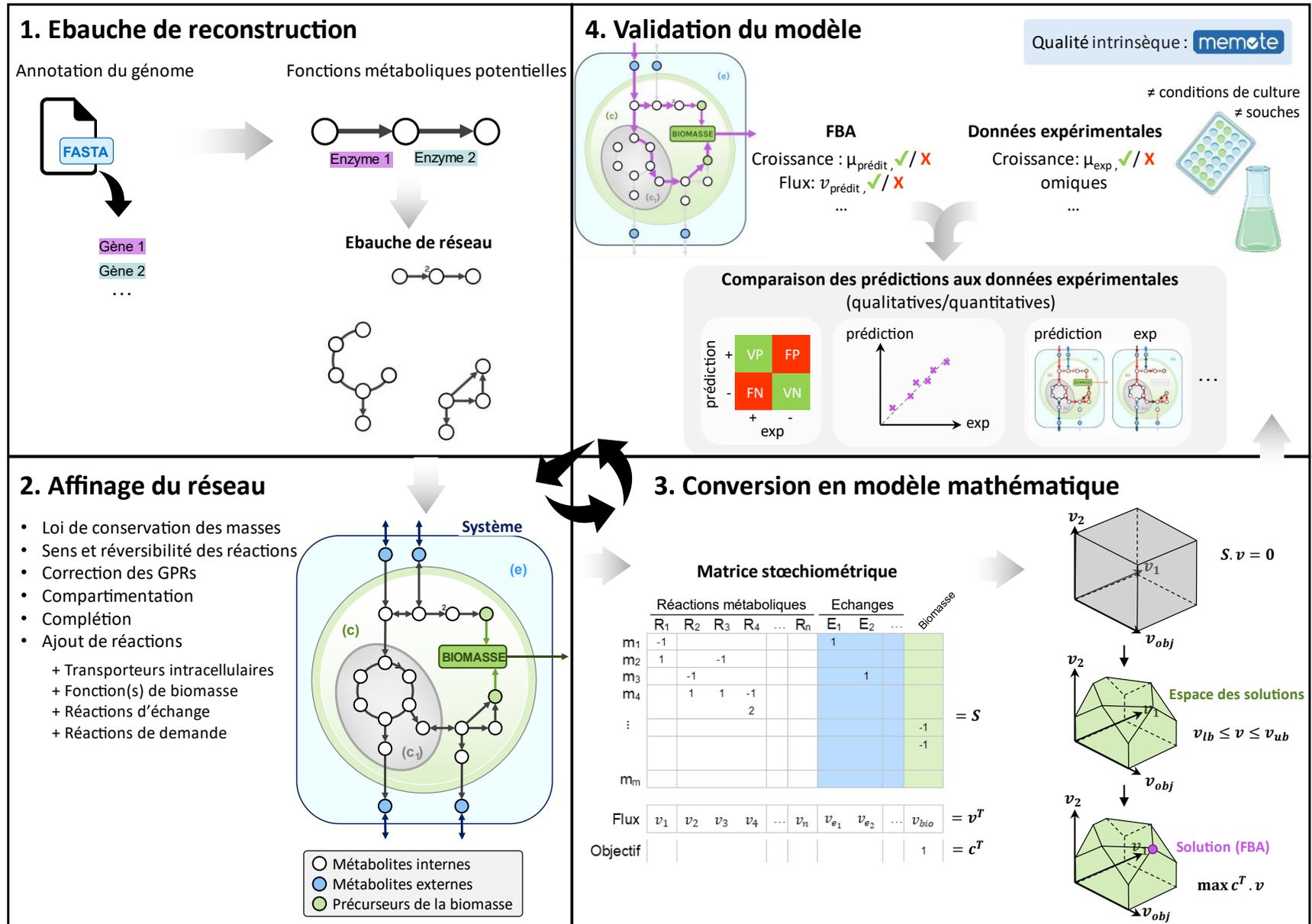


Figure 1.17. Les quatre principales étapes de la reconstruction d'un modèle métabolique. (Texte de la légende : page suivante)

(A) Une ébauche de réseau est générée à partir de l'annotation du génome. (B) L'ébauche de réseau est affinée. (C) Le réseau est converti en modèle mathématique permettant la prédiction de flux par FBA. (D) Les prédictions du modèle sont comparées aux données expérimentales qualitatives et quantitatives. Les écarts entre les prédictions et les observations biologiques sont utilisés pour améliorer le modèle dans un processus itératif (étapes 2-4).

### Réversibilité des réactions

Cette étape consiste à déterminer si les réactions sont réversibles ou non. La réversibilité des réactions se base sur les propriétés thermodynamiques des réactions (telles que leur énergie libre de Gibbs). Ce type d'information est généralement disponible dans des bases de données telles que BRENDA (Schomburg et al. 2000; Chang et al. 2021), BiGG Models (King et al. 2016; Norsigian et al. 2020) ou MetaCyc (Caspi et al. 2018; Caspi et al. 2020). Des outils comme eQuilibrator (Noor et al. 2013) peuvent également être utiles pour déterminer ces paramètres thermodynamiques. Ces contraintes doivent être soigneusement analysées car elles exercent une forte influence sur les prédictions du modèle. Un trop grand nombre de réactions réversibles peut créer des cycles générateurs d'énergie. Inversement, restreindre excessivement la réversibilité des réactions peut sur-ajuster le modèle et le rendre incapable de prédire le comportement métabolique dans différentes conditions environnementales ou génétiques.

### Définition du système et compartimentation des réactions et des métabolites

La compartimentation d'un réseau nécessite l'affectation des métabolites et des réactions dans différents compartiments et l'ajout de réactions de transport permettant le passage des métabolites d'un compartiment à l'autre. Ces transports peuvent être associés à des transporteurs (et donc avoir une GPR) ou à la diffusion des métabolites à travers les membranes. Les réactions de transports sont définies entre les compartiments intracellulaires (entre le cytosol et la mitochondrie par exemple) mais aussi entre le cytosol et le compartiment extracellulaire. Suite à la compartimentation, moins de réactions contribuent aux pools de métabolites dans chaque compartiment, et certains métabolites deviennent spécifiques à un compartiment. Ces modifications structurelles du réseau ajoutent indirectement des contraintes car l'échange de métabolites entre les compartiments n'est possible qu'à travers les réactions de transport ajoutées (Klitgord et Segrè 2010; Seif et Palsson 2021).

### Définition des associations Gène-Protéine-Réaction

Les associations entre gènes et réactions (GPR) générées automatiquement par les outils d'annotation sont confrontées à la connaissance des experts du métabolisme et à la bibliographie. Des corrections à ces associations GPR sont également souvent apportées lorsque des différences sont constatées entre les essentialités de gènes prédites par le modèle et celles observées expérimentalement.

### Ajout de la fonction de biomasse

Une ou plusieurs réactions de biomasse sont ajoutées au réseau. La composition de cette fonction de biomasse est déterminée à partir de données expérimentales, ou par similitude avec des organismes phylogénétiquement proches (Feist et Palsson 2010; Thiele et Palsson 2010). Plusieurs outils permettent la génération de la fonction de biomasse. Elle peut être ajoutée lors de la reconstruction automatisée de l'ébauche de réseau par des outils automatiques, tels que ModelSEED (Devoid et al. 2013) qui permet à l'utilisateur de choisir la fonction objective parmi une liste de réactions de biomasse définies selon les groupes phylogénétiques, ou CarveMe qui propose une biomasse pour les bactéries à Gram positif et une pour celles à Gram négatif lors de la reconstruction de modèles d'organismes procaryotes (Machado et al. 2018). Les fonctions de biomasse générées automatiquement ne sont généralement qu'une représentation qualitative de la biomasse, décrivant la présence ou l'absence de composants de biomasse essentiels. Leur composition est souvent très éloignée de la composition réelle de la biomasse des organismes modélisés. Pour remédier aux limitations de ces méthodes automatiques, différentes approches telles que ObjFind (Burgard et Maranas 2003), BOSS (Gianchandani et al. 2008), et invFBA (Zhao et al. 2016) ont été développées dans le but de mieux définir les compositions quantitatives des fonctions de biomasse. Ces approches utilisent des données fluxomiques pour prédire les objectifs métaboliques finaux de la cellule. Récemment, BOFdat, un outil permettant de générer des fonctions de biomasse à partir de multiples données omiques (génomique, transcriptomique, protéomique, lipidomique) a également été développé (Lachance et al. 2019).

La croissance d'un organisme, mais également son organisation interne, les échanges entre les différents compartiments cytosoliques et la maintenance d'un métabolisme intracellulaire fonctionnel nécessitent de l'énergie. Cette dépense énergétique est modélisée par la consommation d'ATP. On sépare classiquement : (1) l'énergie (sous forme d'ATP) nécessaire à la réplication d'une cellule, notamment pour la synthèse des macromolécules (protéines, ADN et ARN). Cette consommation d'ATP (GAM, growth associated ATP maintenance), étant proportionnelle au taux de croissance, est directement prise en compte dans la fonction de biomasse. Et (2) la consommation d'énergie indépendante de la croissance (NGAM, non-growth associated ATP maintenance) liée à des processus de maintenance cellulaire tels que les transports, le repliement des protéines etc. La NGAM est prise en compte dans le modèle via l'ajout d'une réaction d'hydrolyse de l'ATP simulant une dissipation d'énergie, dont la vitesse minimale est contrainte à partir de données expérimentales (Varma et Palsson 1994b; Thiele et Palsson 2010).

### Complétion du réseau

En raison de l'incomplétude des annotations et/ou des limites des outils d'annotation, des réactions peuvent être manquantes dans l'ébauche de réseau. Inversement, l'ébauche peut contenir des réactions inutiles (faux-positifs). Ces réactions peuvent être isolées dans le réseau, laissant à tort penser qu'il manque d'autres réactions pour les relier. Une étape de complétion (ou gap-filling) doit donc être réalisée afin d'identifier des réactions permettant d'établir la fonctionnalité du système. Un réseau métabolique doit être fonctionnel pour un objectif métabolique spécifique, généralement la production de biomasse. En d'autres termes, il doit être possible de définir un flux entre les substrats et la fonction objectif. Afin d'étudier la fonctionnalité d'un réseau, il est donc essentiel de définir préalablement la composition du milieu dans lequel l'organisme d'intérêt est cultivé. Pour tester si un flux est possible, il faut utiliser le réseau en tant que modèle (cf. étapes 3 et 4). Une première itération complète du cycle est ainsi nécessaire pour identifier les voies métaboliques non fonctionnelles qui doivent être complétées.

La complétion est ainsi une étape essentielle d'amélioration du réseau. Elle peut être réalisée manuellement quand des experts de l'organisme connaissent parfaitement le métabolisme de celui-ci. Des nombreuses méthodes bioinformatiques de complétion ont également été développées (Pan et Reed 2018). La majorité de ces méthodes reposent sur le principe de parcimonie, visant ainsi à minimiser le nombre de modifications à effectuer, et donc minimiser le nombre de réactions à ajouter. Les réactions spontanées (non catalysées par des enzymes et donc non associées à un gène) sont également ajoutées au réseau au cours de cette étape.

Les étapes 1 et 2 déterminent la qualité du réseau métabolique reconstruit et donc la qualité des résultats des analyses se basant sur ces réseaux.

### **Etape 3 : Transformation du réseau en modèle computationnel**

L'une des utilisations les plus courantes et les plus puissantes des réseaux métaboliques à l'échelle du génome est la prédiction des flux métaboliques. Nous utiliserons dans cette thèse le formalisme de la modélisation métabolique à base de contraintes, qui permet de modéliser les distributions de flux métaboliques dans les réseaux à l'échelle du génome. Ce formalisme a été très largement utilisé pour modéliser les réseaux métaboliques à cette échelle, et une très grande diversité de méthodes et d'applications ont été définies sur cette base (Lewis, Nagarajan et Palsson 2012; O'Brien, Monk et Palsson 2015). La plupart de ces méthodes reposent sur deux éléments clés : les contraintes appliquées au système biologique et la méthode d'analyse pour prédire les flux. C'est notamment le

cas de l'analyse en balance de flux (FBA : Flux Balance Analysis), qui en dépit d'être la méthode la plus ancienne, reste l'une des méthodes les plus largement utilisées (Orth, Thiele et Palsson 2010; Varma et Palsson 1994a). Cette approche nécessite la transformation du réseau en modèle mathématique et repose sur deux hypothèses : (1) que le système est à l'état stationnaire et (2) que les organismes ont un objectif général pendant la croissance.

**N. B. :** Le modèle mathématique à base de contrainte d'un réseau métabolique à l'échelle du génome est couramment désigné comme modèle métabolique à l'échelle du génome (GEM, "Genome Scale Metabolic model").

### La matrice stœchiométrique

D'un point de vue mathématique, un réseau de  $m$  métabolites et  $r$  réactions peut être représenté par une matrice stœchiométrique  $S$  formée de  $m$  lignes et  $n$  colonnes (Figure 1.17C). Chaque ligne représente un métabolite et chaque colonne une réaction.  $S_{ij}$  correspond au coefficient stœchiométrique du métabolite  $i$  dans la réaction  $j$ .  $S_{ij}$  est ainsi :

- positif si le métabolite est un produit de la réaction,
- négatif si c'est un substrat,
- nul s'il n'y participe pas.

La dynamique du réseau métabolique peut alors être exprimée par l'équation suivante :

$$\frac{dM(t)}{dt} = S \cdot v \quad (\text{Eq. 1.1})$$

Où  $M(t)$  correspond au vecteur des concentrations des métabolites du réseau à l'instant  $t$ , et  $v$  au vecteur des vitesses des réactions du réseau. La vitesse d'une réaction représentant son flux, le vecteur  $v$  est également appelé distribution de flux. Par convention les flux sont généralement exprimés en millimole par gramme de matière sèche par heure ( $\text{mmol.gDW}^{-1}.\text{h}^{-1}$ ).

### Contraintes de flux et espace des solutions

L'une des hypothèses sur lesquelles repose la FBA est que le système est à l'état stationnaire. Cette hypothèse de stationnarité signifie que les concentrations de chaque métabolite à l'intérieur du système sont considérées comme constantes. En effet, les transitions métaboliques sont généralement rapides par rapport aux taux de croissance cellulaire et aux changements

environnementaux (Varma et Palsson 1994a). La conséquence de cette hypothèse est que tous les flux métaboliques conduisant à la formation et à la dégradation d'un métabolite doivent s'équilibrer.

Mathématiquement, cette hypothèse implique que le produit de la matrice  $S$  et du vecteur  $v$  est égal à zéro (Eq. 1.2). Cette supposition permet alors de définir un système d'équations linéaires et de calculer l'ensemble des vecteurs de flux  $v$  possibles.

$$\frac{dM(t)}{dt} = S \cdot v = 0 \quad (\text{Eq. 1.2})$$

La matrice stœchiométrique impose des contraintes d'équilibre des flux sur le réseau, garantissant que la quantité totale de tout composé produit doit être égale à la quantité totale consommée à l'état stationnaire. Ces contraintes définissent principalement des relations entre les flux. Typiquement, deux réactions successives dans une voie métabolique sans branchement sont contraintes à avoir des flux égaux pour conserver la quantité du métabolite intermédiaire.

D'autres contraintes peuvent également être appliquées au GEM. Chaque réaction peut se voir assigner des bornes qui définissent le flux minimal (limite inférieur ou lower bound) et maximal (limite supérieure ou upper bound) autorisés à travers la réaction. Typiquement, une réaction irréversible a toujours une borne inférieure positive ou nulle et une borne supérieure positive. Inversement, une réaction réversible a une borne inférieure négative et une borne supérieure positive. Ces limites de flux sont notamment appliquées pour contraindre la disponibilité et le taux d'absorption des substrats. Comme indiqué précédemment, les réactions d'échange permettent de définir le milieu extracellulaire et peuvent être laissées "ouvertes" (pour permettre à un substrat d'entrer dans le système à un taux spécifié) ou "fermées" (le substrat ne peut que quitter le système). Des mesures de taux d'échange avec l'environnement peuvent être effectuées, et utilisées pour définir les limites de ces réactions. Les bornes des réactions du réseau sont représentées dans les vecteurs  $b_l$  (lower bound) et  $b_u$  (upper bound) et sont utilisées pour contraindre le vecteur de flux :

$$b_l \leq v \leq b_u \quad (\text{Eq. 1.3})$$

Diverses contraintes d'ordre biologique, telles que des contraintes liées aux ressources (Goelzer, Fromion et Scorletti 2011; Goelzer et al. 2015; Bodeit et al. 2023), aux enzymes (Beg et al. 2007; Adadi et al. 2012; Sánchez et al. 2017), au protéome (Elseman et al. 2022) ou à la thermodynamique (Beard, Liang et Qian 2002; Niebel, Leupold et Heinemann 2019) peuvent également être ajoutées afin de réduire l'espace des solutions.

Plusieurs vecteurs  $v$  satisfont les contraintes imposées par les équations (Eq. 1.2) et (Eq. 1.3). L'espace mathématique des solutions respectant ces contraintes, appelé espace des solutions, définit l'ensemble des distributions de flux possibles (Figure 1.17C).

#### Objectif de simulation

La FBA repose ensuite sur la définition d'une "fonction objective" à maximiser ou minimiser. Cette fonction représente l'objectif biologique du système. Cet objectif de simulation se définit mathématiquement par l'équation suivante :

$$Z = c^T \cdot v \quad (\text{Eq. 1.4})$$

Où  $c$  est un vecteur de poids qui représente la proportion dans laquelle chaque réaction de  $v$  participe à l'objectif. L'objectif de simulation le plus largement utilisé en FBA est la maximisation de la réaction de biomasse, ce qui permet de simuler la croissance de l'organisme d'intérêt. Biologiquement cet objectif repose sur l'hypothèse que l'objectif de l'organisme est de proliférer. Cela se traduit mathématiquement par la maximisation du flux de la fonction de biomasse (réaction qui consomme l'ensemble des précurseurs de la biomasse dans des proportions spécifiques). Le vecteur  $c$  est alors entièrement constitué de 0 et d'un unique 1 à la position de la fonction de biomasse. La fonction objective devient ainsi  $Z = v_{\text{biomasse}}$ .

La FBA peut donc être décrite mathématiquement de la manière suivante :

$$\begin{aligned} \text{maximiser } Z &= c^T \cdot v \\ \text{tel que } S \cdot v &= 0 \\ \mathbf{b}_l &\leq v \leq \mathbf{b}_u \end{aligned} \quad (\text{Eq. 1.5})$$

Ce problème mathématique peut être résolu en programmation linéaire et permet ainsi de trouver une solution optimale d'un réseau métabolique à l'état stationnaire (Figure 1.17C). Plusieurs solutions sont souvent possibles. Des méthodes ont été développées pour explorer la diversité des solutions telles que l'analyse de variabilité de flux (FVA, Flux Variability Analysis) (Gudmundsson et Thiele 2010).

#### **Etape 4 : Validation du GEM**

Cette étape consiste à vérifier que le réseau métabolique obtenu est fonctionnel et présente une qualité suffisante. Deux aspects du modèle sont vérifiés : (1) ses caractéristiques intrinsèques de qualité et (2) ses capacités prédictives (Figure 1.17D) (Thiele et Palsson 2010).

### Caractéristiques intrinsèques de qualité

La suite de test MEMOTE a récemment été développée afin de standardiser les tests qualités des GEM et faciliter le suivi de la qualité d'un modèle ou de comparer la qualité de différents modèles (Lieven et al. 2020). MEMOTE automatise la réalisation de différents types de tests.

(1) Test de la cohérence du modèle : vérification du respect de la loi de conservation des masses pour l'ensemble des réactions du modèle, vérification de l'équilibre des charges, identification des réactions bloquées (qui ne peuvent pas avoir de flux) et des réactions pouvant avoir un flux illimité. La présence d'une grande proportion de réactions pouvant avoir un flux illimité au sein d'un GEM révèle des problèmes de réversibilité de réactions, de cofacteurs manquants ou des erreurs de définition des réactions de transports. MEMOTE identifie également les cycles produisant de l'énergie (EGC erroneous energy-generating cycles). Les EGC sont des cycles qui produisent des métabolites énergétiques, tels que l'ATP ou le NAD(P)H, sans source d'énergie. La présence d'EGC fausse l'équilibre énergétique modélisé et peut entraîner une augmentation des taux de croissance allant jusqu'à 25%, rendant peu fiable la prédiction de ces derniers (Fritzemeier et al. 2017; Lieven et al. 2020). Enfin, les métabolites orphelins et "cul-de-sac" (dead-end) sont également identifiés. La présence d'une fonction de biomasse est vérifiée.

(2) Vérification de la présence de GPR et d'identifiants des gènes, des réactions et métabolites de diverses bases de données couramment utilisées pour les GEM: BioCyc (Karp et al. 2019), BiGG (King et al. 2016; Norsigian et al. 2020), BRENDA (Schomburg et al. 2000; Chang et al. 2021), Chebi (Degtyarenko et al. 2008; Hastings et al. 2016), EC Number (Egelhofer, Schomburg et Schomburg 2010), InChI (InChI Trust), KEGG (Kanehisa et Goto 2000b), Rhea (Bansal et al. 2022), UniProt (Consortium 2022) etc.

Ces premières vérifications qualité alimentent en général les premières itérations de correction du modèle (Etape 2).

### Validation des phénotypes prédits avec des données expérimentales

L'évaluation du modèle au niveau fonctionnel est réalisée par comparaison entre les phénotypes prédits et des données expérimentales. Divers phénotypes, tant qualitatifs que quantitatifs, peuvent être testés pour valider le modèle, notamment la croissance dans différentes conditions (croissance/absence de croissance, ou taux de croissance), type trophique, produits sécrétés, essentialité des gènes. Ces tests *in silico* sont principalement basés sur des simulations par FBA (ou méthodes dérivées). Les écarts entre les prédictions et les observations sont utilisés pour améliorer le modèle dans un processus itératif (Thiele et Palsson 2010). Des prédictions inexactes peuvent

découler d'une connaissance incomplète ou erronée du métabolisme de l'organisme modélisé. Les GEM peuvent révéler ces prédictions incorrectes, ce qui permet d'identifier les lacunes ou les inexactitudes dans les connaissances métaboliques et aider à concevoir des expériences ciblées visant à corriger ces dernières (Bordbar et al. 2014).

### 1.3.3 Applications des GEM à l'échelle du microorganisme

Les GEM sont particulièrement bien adaptés aux expériences de type "what if" impliquant des perturbations génétiques ou environnementales, telles que : comment la cellule se comporterait-elle dans tel environnement ? Comment telle mutation affecterait-elle ses capacités métaboliques ? Quelles mutations maximiseraient la croissance et/ou la production du métabolite d'intérêt ? Les GEM sont également adaptés à l'intégration et à l'analyse de diverses données expérimentales, notamment les données omiques. Ainsi, de nombreuses méthodes par contraintes (ou Constraint-Based Method, CBM) ont été développées pour exploiter les modèles GEM. En 2012, Lewis, Nagarajan et Palsson en avaient déjà recensé plus d'une centaine (Lewis, Nagarajan et Palsson 2012) et de nombreux autres algorithmes ont été développés au cours des dix dernières années. Une description exhaustive de ces méthodes dépasse le cadre de ce manuscrit, mais la diversité de ces approches a fait l'objet de plusieurs revues (Durot, Bourguignon et Schachter 2009; Lewis, Nagarajan et Palsson 2012; Bordbar et al. 2014; Ng et al. 2022; Ye et al. 2022). Dans cette section, nous aborderons certaines applications des modèles métaboliques pertinentes pour les biotechnologies industrielles, notamment la prédiction des phénotypes et l'aide à l'ingénierie de souches. Nous détaillerons également quelques méthodes par contraintes pouvant être utilisées pour aborder ces problématiques.

#### 1.3.3.1 Prédire le phénotype

##### **Prédiction d'un état "optimal"**

Le phénotype d'un microorganisme peut être simulé sous forme de distributions de flux représentant un état "optimal". Diverses CBM ont été développées pour prédire un état optimal du réseau métabolique. Ces méthodes reposent sur l'application de contraintes et sur l'optimisation d'une fonction objective dans le but d'identifier une distribution de flux pertinente sur le plan biologique. La FBA est l'approche la plus ancienne et la plus simple de ces méthodes. Comme présenté dans la section précédente, la FBA repose sur les contraintes induites par l'hypothèse de l'état stationnaire et par les limites (lower et upper bounds) appliquées aux réactions du réseau. Elle permet d'identifier une solution maximisant ou minimisant un objectif de simulation (souvent la maximisation de la fonction objective de biomasse qui permet de simuler la croissance d'un microorganisme). Diverses stratégies

ont été proposées pour élargir les applications de la FBA, telles que (1) la définition d'autres objectifs de simulations, (2) l'introduction de contraintes supplémentaires fondées sur des hypothèses biologiques ou sur l'utilisation de données supplémentaires et (3) la remise en question de l'hypothèse de l'état stationnaire.

### **Objectifs de simulation alternatifs à la maximisation de la fonction de biomasse**

Utiliser la maximisation de la fonction de biomasse comme objectif de simulation par FBA repose sur l'hypothèse que les organismes ayant un métabolisme maximisant la prolifération ont un avantage sélectif. Ainsi, les microorganismes se développant rapidement peuvent surpasser d'autres espèces ou souches de microorganismes. Une question fondamentale qui a entouré l'approche FBA depuis ses débuts est de savoir si, et dans quelles conditions, l'hypothèse selon laquelle les systèmes biologiques fonctionnent au voisinage d'un optimum prévisible est valide, et le cas échéant, quelle fonction objective représente le mieux les objectifs métaboliques d'un organisme (Schuetz, Kuepfer et Sauer 2007; Schuster, Pfeiffer et Fell 2008; Bernstein et al. 2021). C'est pourquoi des objectifs de simulation alternatifs ont été proposés (Schuetz, Kuepfer et Sauer 2007; Yasemi et Jolicoeur 2021) tels que (1) la maximisation de la production d'ATP, ce qui postule que l'organisme d'intérêt utilise ses ressources pour maximiser la production d'énergie sous forme d'ATP (Fell et Small 1986; Savinell et Palsson 1992), (2) la minimisation de la consommation de la source de carbone, en supposant que l'évolution favorise l'utilisation la plus efficace du substrat (Kavšček et al. 2015), (3) la minimisation de l'utilisation du potentiel redox, ce qui suppose que l'organisme d'intérêt réduit le nombre de réactions d'oxydation du NAD(P)H et conserve ainsi son énergie ou l'utilise de la manière la plus efficace possible (Knorr, Jain et Srivastava 2007). Plusieurs études ont également proposé de déterminer plusieurs objectifs successifs (Patané et al. 2019; Schnitzer, Österberg et Cvijovic 2022). Par exemple, lorsque les microorganismes sont exposés à des conditions non optimales, telles que des températures extrêmes, un pH extrême ou des carences en nutriments, il est souvent insuffisant d'utiliser une seule fonction objective pour prédire de manière précise les données expérimentales (Serrano-Bermúdez et al. 2017). Une optimisation en deux étapes peut alors être réalisée, en maximisant la production de biomasse dans un premier temps, puis dans un second temps en minimisant une fonction auxiliaire telle que la production d'ATP, le taux d'absorption de substrat ou le taux de production de potentiel d'oxydoréduction (Hosseini 2020). Ainsi, la diversité des réactions pouvant être utilisées comme objectif de simulation permet d'élargir les contextes d'application de la FBA.

Un autre axe d'expansion des applications des GEM implique la modification des contraintes définissant l'espace des solutions possibles, pour obtenir des solutions se rapprochant davantage de la physiologie de l'organisme d'étude.

### **Modifier les contraintes du réseau**

Diverses contraintes biologiques, telles que l'économie de l'utilisation des enzymes, la dilution métabolique ou la variation du niveau de transcrits ont ainsi été proposées pour contraindre les réseaux métaboliques. Une large variété de CBM a ainsi été développée dans cette optique. Par exemple, la FBA parcimonieuse (pFBA ou parcimonious FBA) suppose qu'en cas de croissance exponentielle, la sélection favorise les souches ayant la croissance la plus rapide et nécessitant le plus faible flux métabolique global (une approximation de la minimisation de la masse totale d'enzymes nécessaire pour mettre en œuvre la solution optimale) (Lewis et al. 2010). Les algorithmes de pFBA permettent ainsi de prédire une distribution de flux permettant de maximiser l'objectif de simulation tout en minimisant la somme totale des flux à travers le réseau (Ebrahim et al. 2013). Il a été démontré que la solution de flux générée à l'aide de la pFBA est cohérente avec l'expression différentielle des gènes dans différentes conditions, ce qui confirme la pertinence biologique de la pFBA (Machado et Herrgård 2014).

D'autre part, la minimisation de l'ajustement métabolique (MOMA, Minimization Of Metabolic Adjustment) prédit la distribution des flux après une perturbation, de manière à ce que cette distribution de flux soit la plus proche possible de la distribution de flux de référence (par exemple, la solution FBA avant la perturbation) (Segrè, Vitkup et Church George M. 2002). Initialement, la MOMA avait été développée pour simuler l'impact d'une perturbation génétique (délétion d'un gène), mais récemment cette méthode a été appliquée à la simulation de modifications des conditions environnementales (eMOMA, environmental version of MOMA) (Kim et al. 2019).

Par ailleurs avec l'essor des analyses à haut débit, des méthodes permettant d'intégrer les données omiques (génomique, transcriptomique, protéomique, lipidomique et métabolomique) dans les modèles GEM ont été développées. Cette intégration peut être basée sur une approche "interrupteur" où une réaction est maintenue activée ou désactivée, ou sur une approche "valve" qui détermine les contraintes de flux des réactions sur la base des données quantitatives (Haas et al. 2017; Cunha et al. 2023).

### **S'affranchir de l'hypothèse de l'état stationnaire.**

Bien que la FBA suppose qu'un système ne change pas à l'état stable, ces pseudo-états stables peuvent être couplés à un système dynamique avec des variables environnementales changeantes à l'aide de la FBA dynamique (dFBA, dynamic FBA). Ainsi, les méthodes d'analyse dynamique du bilan de flux (dFBA) (Varma et Palsson 1994b; Mahadevan, Edwards et Doyle 2002) permettent de prédire la consommation et la production de métabolites au cours du temps. Deux approches principales ont été proposées : (1) l'approche d'optimisation dynamique (DOA), où l'optimisation est réalisée sur

l'ensemble de la période d'intérêt afin d'obtenir des profils temporels des flux et des niveaux de métabolites ; et (2) l'approche d'optimisation statique (SOA) qui consiste à diviser la période d'étude en intervalles de temps restreints, à chaque pas de temps une simulation de flux par FBA est réalisée, les résultats obtenus sont propagés au pas de temps suivant jusqu'à résolution du temps total.

Ces dernières années, les approches dFBA ont été appliquées aux modèles GEM de microalgues pour analyser la production de lipides et de pigments entre les phases d'obscurité et de lumière (Baroukh et al. 2016; Flassig et al. 2016). Loira et al. 2017 ont appliqué la dFBA pour étudier la production de TAG au fil du temps, en concevant des stratégies d'optimisation de souche (par délétion de gènes) visant à maximiser les TAG chez *M. salina*.

D'autre part, appliquer des contraintes sur la composition de la biomasse a récemment été proposée comme une méthode alternative pour simuler le comportement métabolique dynamique (Zuñiga et al. 2018). Les changements dans la composition de la biomasse influencent l'état du réseau métabolique, affectant ainsi directement les prédictions. Cette approche a permis d'améliorer la précision de la prédiction des flux en tenant compte des mesures effectuées au cours de la croissance de la microalgue *Chlorella vulgaris* dans des conditions de culture photoautotrophes et hétérotrophes. Les distributions des séries temporelles de flux simulent avec précision 75 % des données d'expression et de protéomique recueillies au cours de la croissance, y compris les réactions allostériques et les enzymes à plusieurs sous-unités. Cette approche a également permis de déterminer le contenu net des pools d'azote dans chaque condition.

### 1.3.3.2 Caractérisation globale de l'espace des solutions

La FBA permet d'identifier une solution qui maximise un objectif. L'une des limites de la FBA est que souvent la solution proposée n'est pas unique. Il peut exister des optimaux alternatifs, c'est-à-dire des distributions de flux alternatives permettant d'atteindre la même valeur optimale de l'objectif. L'analyse de variabilité de flux (FVA, Flux Variability Analysis) permet d'étudier ces optimaux alternatifs (Gudmundsson et Thiele 2010). Cette analyse permet de déterminer les flux minimaux et maximaux possibles pour chaque réaction tout en maintenant la fonctionnalité du réseau. Cette fonctionnalité globale peut être complète (on garde la même valeur optimale), ou suboptimale si on ne conserve qu'une fraction de l'optimum. L'analyse de la variabilité de flux (FVA, Flux Variability Analysis) est aussi employée pour tester les flux en regardant quelles sont les réactions essentielles et les réactions bloquées. Ainsi, la FVA est un moyen de tester la robustesse du modèle.

D'autres méthodes telles que le sampling ou l'analyse des modes de flux élémentaires (EFM) permettent également de caractériser l'espace des solutions (Lewis et al. 2010).

### 1.3.3.3 Applications pour l'ingénierie métabolique

L'ingénierie métabolique est une stratégie d'optimisation de souches consistant à manipuler les voies métaboliques d'un organisme en vue d'améliorer ses performances. Etant donné que les modèles GEM représentent l'activité de centaines d'enzymes, ils ont le potentiel d'évaluer les phénotypes des mutants *in silico*. En conséquence, diverses CBM ont été développées pour prédire l'essentialité d'un gène ou d'une réaction, évaluer l'impact d'une modification génétique sur la production de métabolites d'intérêt ou encore identifier les goulots d'étranglement de voies métaboliques. De nombreuses méthodes d'aide au design de souches ont également été développées (Lewis, Nagarajan et Palsson 2012; Machado et Herrgård 2015; Maia, Rocha et Rocha 2016; Vieira et al. 2019).

#### **Prédire l'essentialité des gènes**

Les GEM peuvent être utilisés pour formuler des prédictions quant à l'essentialité des gènes ou des réactions. Pour réaliser ce type d'analyse, chaque gène (ou chaque réaction) est inactivé individuellement au sein du modèle. Le modèle est alors optimisé pour la production de biomasse. Si le taux de croissance prédit est supérieur à un certain seuil, le gène (ou la réaction) est considéré comme non essentiel. À l'inverse, si le taux de croissance est nul ou inférieur au seuil, le gène (ou la réaction) est considéré comme essentiel (Joyce et Palsson 2008; Suthers, Zomorodi et Maranas 2009). Pour rappel, dans un modèle métabolique, la suppression d'un gène entraîne l'inactivation des réactions associées (les contraintes supérieures et inférieures des flux de ces réactions sont fixées à zéro, sauf si d'autres gènes sont associés à une même réaction avec le connecteur logique "OR"). Lorsqu'une réaction est inactivée son flux est fixé à zéro (Heirendt et al. 2019). Cette approche revient à déterminer quel(les) gènes/réactions sont essentiel(les) pour assurer le flux à travers la réaction de biomasse du modèle. Les GEM peuvent être très efficaces pour prédire l'essentialité des gènes. Des modèles très élaborés comme celui d'*E. coli* ont permis de prédire l'essentialité dans différentes conditions de croissance avec une précision allant jusqu'à 93,4% (Monk et al. 2017).

#### **Prédire les distributions de flux de souches mutantes**

Les algorithmes de prédiction de flux simulant des délétions de gènes se basent sur l'hypothèse que la variation des flux dans une souche mutante doit être minimale comparée aux flux prédits pour la souche sauvage (ou condition de référence). Deux CBM sont fréquemment utilisées : la MOMA (Minimization Of Metabolic Adjustment, (Segrè, Vitkup et Church George M. 2002)) et la ROOM (Regulatory On/Off Minimization (Shlomi, Berkman et Ruppin 2005)). Comme précisé précédemment, la MOMA consiste à minimiser l'ajustement de chacun des flux par rapport à une distribution de flux

de référence. La seconde approche, la ROOM, consiste à trouver la solution optimale tout en minimisant le nombre de flux qui vont être modifiés par rapport à la condition de référence. A la différence de la MOMA, la ROOM permet une reconfiguration plus large des flux individuels dans le réseau, permettant ainsi une meilleure identification des voies alternatives mise en place en réponse à la délétion d'un gène. Ces méthodes ont ouvert la voie à plusieurs méthodes d'aide au design de souche.

### **Identification de stratégies d'ingénierie métabolique**

De nombreuses méthodes de conception de souches *in silico* ont été développées (CSOM, computational strain optimization method). Ces méthodes ont fait l'objet de diverses revues (Lewis, Nagarajan et Palsson 2012; Maia, Rocha et Rocha 2016; Machado et Herrgård 2015; Vieira et al. 2019). Les CSOM peuvent notamment être utilisées pour identifier des stratégies d'édition génomique telles que (1) la délétion de gènes : OptKnock (Burgard, Pharkya et Maranas 2003), ReactKnock (Xu et al. 2013), OptCouple (Jensen et al. 2019), OptGene (Patil et al. 2005) ; (2) l'addition de gènes : OptSwap, (King et Feist 2013) ; (3) l'up-/down-régulation de l'expression génique : OptForce (Ranganathan, Suthers et Maranas 2010), IdealKnock (Gu et al. 2016) ; et (4) l'addition de voies métaboliques exogènes : OptStrain (Pharkya, Burgard et Maranas 2004) ou SimOptStrain (Kim, Reed et Maravelias 2011).

Plusieurs de ces méthodes ont été appliquées à des modèles métaboliques de microalgues. Par exemple, la méthode OptKnock a été utilisée pour évaluer *in silico* l'effet des délétions de gènes/réactions sur la surproduction de lipides chez *Rhodospiridium torulides*, aboutissant à l'identification de quatre souches mutantes potentielles (Castañeda et al. 2019).

Ces différentes méthodes reposent sur des formulations mathématiques et des hypothèses différentes (Lewis, Nagarajan et Palsson 2012). Plusieurs pipelines, tels que OptPipe (Hartmann et al. 2017), ou le "minimal cut sets framework" proposé par (Schneider, Kamp et Klamt 2020), ont été développés afin d'appliquer simultanément différentes méthodes dans le but d'identifier un maximum de stratégies d'ingénierie métabolique et de comparer la robustesse de ces stratégies. Ces pipelines comportent plusieurs étapes. (1) L'application de différents algorithmes de conception de souche, (2) le filtrage des cibles d'ingénierie identifiées selon différents critères qui varient en fonction du pipeline (les taux de croissance minimum et maximum supérieurs à un seuil etc.), (3) le classement des différentes cibles via l'attribution d'un score, ici encore basé sur des critères variables en fonctions des méthodes (par exemple en fonction du nombre d'interventions à réaliser, de la distance entre les flux prédits pour le mutant et la souche sauvage etc.) permettant ainsi la sélection des stratégies les plus pertinentes.

#### 1.3.3.4 Impact de la composition de la biomasse sur les prédictions du modèle

Plusieurs études ont démontré l'impact significatif de la composition de la biomasse sur les prédictions et les performances du modèle, notamment la prédiction de l'essentialité des gènes, des taux de croissance et des distributions de flux (Pramanik et Keasling 1998; Schulz et al. 2021; Lavoie et al. 2020; Nookaew et al. 2008).

La composition de cette fonction de biomasse doit donc être définie minutieusement afin d'augmenter le potentiel prédictif d'un GEM. Comme la composition de la biomasse des microorganismes, en particulier des microalgues, est fortement influencée par les conditions de culture, de plus en plus de modèles proposent plusieurs fonctions de biomasse et constituent ainsi des modèles spécifiques à différents contextes. Par exemple, plusieurs modèles de microalgues ont été publiés récemment, comportant trois fonctions de biomasse représentant la composition de la biomasse de la microalgue cultivée dans des conditions de photoautotrophie, mixotrophie et hétérotrophie (Zuñiga et al. 2016; Shah et al. 2017).

Des outils de génération de biomasse tels que BOFdat visent également au développement de ces modèles spécifiques à différents contextes (Lachance et al. 2019).

#### 1.3.3.5 Outils d'analyse des GEM

Bon nombre de ces méthodes ont été initialement développées en utilisant le logiciel MATLAB et ont été intégrées dans la COBRA Toolbox (constraint-based reconstruction and analysis) (Heirendt et al. 2019). Par la suite, la plupart d'entre elles ont été implémentées en Python (bibliothèque COBRAPy) (Ebrahim et al. 2013) dans le but d'accroître leur accessibilité et leur utilisation au sein de la communauté scientifique, Python ne nécessitant en effet pas de licence payante, contrairement à MATLAB (Carey et al. 2020; Ng et al. 2022; Cunha et al. 2023).

Diverses applications permettent de visualiser les modèles métaboliques. L'application web Escher, par exemple, est largement utilisée pour la visualisation de leur contenu sous forme de cartes métaboliques (King et al. 2015). Escher permet de créer des cartes métaboliques à partir des informations contenues dans un modèle. Le format d'entrée d'Escher est le format JSON (Java Script Object Notation), qui est l'un des formats standards de sauvegarde des GEM, proposé notamment par les bibliothèques COBRA (Ebrahim et al. 2013; Heirendt et al. 2019; Norsigian et al. 2020). Escher a également été intégré à des packages Python COBRA tels que Cameo, afin de visualiser des distributions de flux au sein des scripts de simulation (Cardoso et al. 2018). Par ailleurs, Escher-FBA permet la réalisation "dynamique" de simulation de flux sans écriture de lignes de code ni téléchargement de logiciel (Rowe, Palsson et King 2018).

### 1.3.4 Modélisation du métabolisme des microalgues

Le développement des GEM a débuté peu de temps après le séquençage du premier génome complet d'*Haemophilus influenzae* en 1995 (Fleischmann et al. 1995), avec la publication de son premier modèle en 1999 (Edwards et Palsson 1999). Au fil du séquençage de nouveaux organismes, d'autres GEM ont été reconstruits, couvrant ainsi un éventail phylogénétique de plus en plus large, incluant les microalgues (Gu et al. 2019).

#### 1.3.4.1 Modéliser le métabolisme des microalgues

Le premier modèle métabolique de microalgues eucaryote, un modèle de *Chlamydomonas reinhardtii*, a été publié en 2009 (Boyle et Morgan 2009). Du fait de leur potentiel pour diverses applications industrielles, des modèles de diverses souches ont été publiés au cours des dernières années. La modélisation du métabolisme des microalgues a fait l'objet de plusieurs revues (Tibocha-Bonilla et al. 2018; Correa et al. 2020; Cunha et al. 2023). Des modèles GEM notamment de *Chlamydomonas reinhardtii* (May et al. 2008; Boyle et Morgan 2009; Christian et al. 2009; Chang et al. 2011; Cogne et al. 2011; Dal'Molin, Cristiana Gomes de Oliveira et al. 2011; Chaiboonchoe et al. 2014; Imam et al. 2015b; Mora Salguero et al. 2018; Shene, Asenjo et Chisti 2018), *Chlorella vulgaris* (Zuñiga et al. 2016; Zuñiga et al. 2018), *Chlorella variabilis* (Juneja, Chaplen et Murthy 2016), *Fragilariopsis cylindrus* (Lavoie et al. 2020), *Microchloropsis gaditana* (Shah et al. 2017), *Microchloropsis salina* (Loira et al. 2017), *Ostreococcus lucimarinus* et *Ostreococcus tauri* (Krumholz et al. 2012), *Phaeodactylum tricornutum* (Levering et al. 2016), *Scenedesmus obliquus* (Ray, Kundu et Ghosh 2023), *Synechococcus elongatus* (Reimers et al. 2017; Broddrick et al. 2016), *Synechococcus sp.* (Hamilton et Reed 2012; Hendry et al. 2016; Qian et al. 2017; Vu et al. 2013) et *Synechocystis sp.* (Montagud et al. 2010; Yoshikawa et al. 2011; Nogales et al. 2012) sont disponibles. Par ailleurs, un modèle du métabolisme du chloroplaste de *Chlamydomonas reinhardtii*, *Microchloropsis gaditana* et *Phaeodactylum tricornutum* a également été publié (Røkke, Hohmann-Marriott et Almaas 2020).

Ces modèles métaboliques de microalgues ont été exploités pour diverses applications. Des simulations de croissance dans différentes conditions (photoautotrophie, hétérotrophie, mixotrophie, utilisation de différentes sources de carbone et d'azote) ont été réalisées chez *M. gaditana* (Shah et al. 2017), *C. reinhardtii* (Mora Salguero et al. 2018) et *C. vulgaris* (Zuñiga et al. 2016). Un modèle métabolique de *C. reinhardtii* a servi à évaluer différentes sources de lumières (Chang et al. 2011). L'augmentation de la production de lipides par les microalgues a également fait l'objet de plusieurs études, notamment la prédiction de l'impact d'une carence en azote (Xiong et al. 2010; Yoshikawa et

al. 2011; Bogen et al. 2013; Imam et al. 2015a; Levering et al. 2016; Loira et al. 2017; Shah et al. 2017; Zuñiga et al. 2018). L'influence d'une faible concentration en CO<sub>2</sub> et d'une faible luminosité sur l'augmentation globale de la teneur en lipides a également été étudiée (Levering et al. 2016). L'ensemble de ces analyses montrent le fort potentiel des modèles métaboliques pour l'étude du métabolisme lipidique chez les microalgues. Cependant, ces applications doivent reposer sur une modélisation détaillée du métabolisme lipidique.

#### 1.3.4.2 Etat de l'art des modèles des genres *Nannochloropsis* et *Microchloropsis*

Compte tenu de leur potentiel pour diverses applications biotechnologiques, en particulier pour la production de lipides, des efforts ont été effectués ces dernières années pour reconstruire les GEM des genres *Nannochloropsis* et *Microchloropsis* (Table 1.5). Ainsi, la première ébauche de modèle métabolique de *Nannochloropsis* sp. a été publiée en 2016 (Pham 2016). L'année suivante, deux modèles de *M. gaditana* (iRJ1321) et *M. salina* (iNS937) ont été publiés (Loira et al. 2017; Shah et al. 2017). En 2019, un nouveau modèle métabolique de *M. gaditana* (iMgadit\_SilicoLife) a été reconstruit par SilicoLife, TotalEnergies et le CEA (modèle non publié). Enfin, très récemment, un modèle de *M. gaditana* reconstruit automatiquement à l'aide de AuCoMe (Automated Comparison of Metabolism) a été publié (Belcour et al. 2023).

Parmi ces cinq GEM, seuls iRJ1321, iNS937 et iMgadit\_SilicoLife permettent l'optimisation de la production de biomasse par FBA. Dans le modèle de Pham, tous les précurseurs de cette réaction de biomasse ne peuvent pas être synthétisés par le modèle, rendant la simulation de croissance impossible. De plus, la réaction de biomasse a été ajoutée sur la base des données de plantes vertes (*Viridiplantae*) issues de PlantSEED (Seaver et al. 2018). Elle n'est donc pas spécifique de *Nannochloropsis*, ce qui limite ainsi les applications de ce modèle. Le modèle de Belcour et al. 2023, a été reconstruit dans le but d'effectuer des comparaisons phylogénétiques entre espèces plutôt que pour la prédiction de distributions de flux. Ainsi, ce modèle est dépourvu de réaction de biomasse.

L'un des axes majeurs de la recherche et du développement pour l'exploitation industrielle des microalgues réside dans l'optimisation des conditions de culture et la création de souches optimisées. Ces axes d'optimisation reposent sur un approfondissement de la connaissance du métabolisme de la souche étudiée. Comme présenté précédemment, les GEM sont des outils pertinents pour répondre à ces problématiques. Dans l'objectif d'approfondir les connaissances du métabolisme lipidique, il est nécessaire d'avoir un modèle métabolique de *M. gaditana* détaillé pour le métabolisme lipidique.

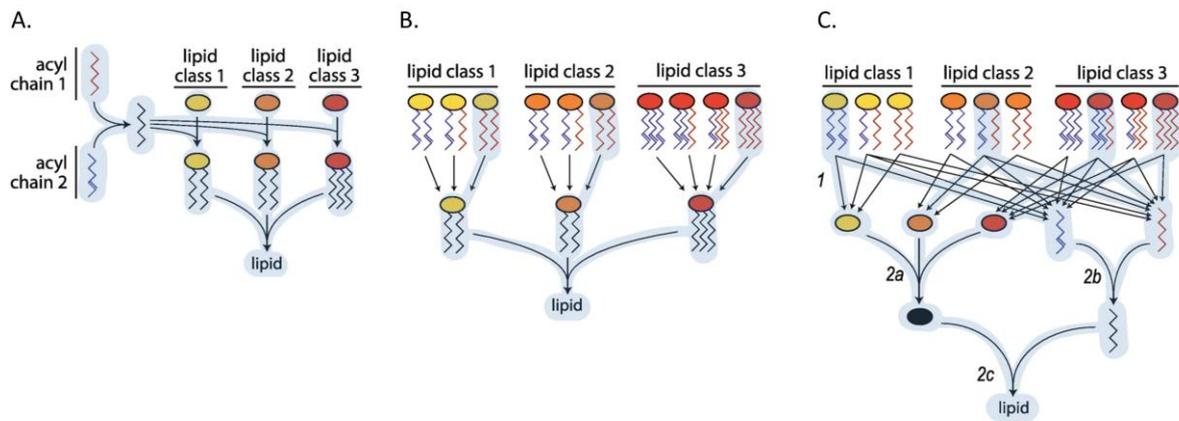
**Table 1.5. Modèles métaboliques à l'échelle du génome disponibles pour les microalgues des genres *Nannochloropsis* et *Microchloropsis*.**

Modèle	Modèle de Pham	iNS937	iRJ1321	iMgadit_SilicoLife	<i>Nannochloropsis</i> _gaditana
Année de reconstruction/ publication	2016	2017	2017	2019	2023 (Non disponible au début de la thèse)
Souche	<i>N. sp</i>	<i>M. salina</i>	<i>M. gaditana</i>	<i>M. gaditana</i>	<i>M. gaditana</i>
Réactions	987	2345	1918	1890	2729
Métabolites	1024	1985	1862	1899	3076
Gènes	383	934	1321	864	3474
<b>Compartiments</b>	10	10	4	8	2
Milieu extracellulaire	✓	✓	✓	✓	✓
Cytosol	✓	✓	✓	✓	✓
Mitochondrie	✓	✓	✓	✓	
Chloroplaste	✓	2 compartiments: stroma & lumen	✓	2 compartiments: matrice & EIM	
Peroxisome	✓	✓		✓	
RE	✓	✓		✓	
Appareil de Golgi	✓	✓			
Lysosome		✓			
Noyau	✓	✓			
Vacuole	✓				
Paroi cellulaire	✓				
<b>Possibilité de FBA</b> (Objectif = maximisation de la fonction de biomasse)	X	✓	✓	✓	X
<b>Références</b>	(Pham 2016)	(Loira et al. 2017)	(Shah et al. 2017)	Non publié	(Belcour et al. 2023)

### 1.3.5 Modéliser le métabolisme lipidique

Du fait de la très grande diversité moléculaire des acides gras et des glycérolipides, la modélisation du métabolisme lipidique présente un défi supplémentaire lors de la modélisation du métabolisme des microalgues (Correa et al. 2020; Cunha et al. 2023). En effet, il est nécessaire de décrire la synthèse des différents acides gras, ainsi que l'ensemble des réactions de synthèse des différentes classes de glycérolipides qui incorporent ces divers acides gras. Cette modélisation nécessite l'ajout de plusieurs dizaines voire centaines de réactions (Correa et al. 2020). Par ailleurs, un problème fréquemment rencontré lors de la modélisation des lipides est l'incorporation précise de ces lipides dans la fonction de biomasse (Sánchez et al. 2019). En raison de la multitude des combinaisons possibles d'espèces

lipidiques et du manque de données détaillées, le calcul des coefficients stœchiométriques peut s'avérer extrêmement complexe. De ce fait, divers formalismes, plus ou moins simplifiés, sont couramment utilisés pour décrire la composition lipidique de la biomasse. Le choix du formalisme est souvent déterminé par la disponibilité des données pour l'organisme d'intérêt. La modélisation la plus précise consiste à prendre en compte le profil en glycérolipides (cf section "La fonction de biomasse" Niveau 2 – lipides) ainsi que le profil en acide gras de chaque classe de glycérolipides (cf section "La fonction de biomasse" Niveau 3). Les combinaisons et positions des acides gras au sein des glycérolipides (positions *sn-1*, *sn-2* (et *sn-3* pour les TAG)) sont ainsi prises en compte dans la composition de la biomasse. Cependant, ce niveau de détail nécessite de disposer de glycérolipidomes complets et détaillés. Ces mesures sont rarement disponibles du fait de leur complexité d'acquisition (Jouhet et al. 2017; Sánchez et al. 2019). Il est donc plus courant de mesurer séparément (1) un profil des différentes classes de glycérolipides, et (2) une distribution des différentes chaînes d'acides gras. Les structures des fonctions de biomasse des GEM ont été adaptées pour tenir compte de ces données. Plusieurs approches de simplification sont fréquemment utilisées. La première approche consiste à supposer que toutes les classes glycérolipidiques ont le même profil en acide gras (Figure 1.18A). Une deuxième approche permet à tout lipide spécifique de former une classe générique correspondante (Figure 1.18B). Par exemple, chaque phosphatidylcholine (PC) décrite dans le modèle (PC 16:0/16:0, PC 16:0/16:1 etc.) peut être convertie en un pseudo-métabolite "PC". Seul le pseudo-métabolite de chaque classe est alors consommé dans la fonction de biomasse dans des proportions déterminées à partir de profils glycérolipidiques expérimentaux. Les profils des acides gras ne sont ainsi pas pris en compte dans cette approche. De plus, les simulations privilégient la consommation des entités de lipides les moins coûteuses en énergie (celles ayant les chaînes carbonées les plus courtes et les plus saturées), ce qui ne reflète pas la situation *in vivo*. Enfin, une troisième approche a récemment été proposée : le formalisme SLIME (Split Lipids Into Measurable Entities) (Figure 1.18C). Cette approche "coupe" les glycérolipides en deux (ou trois) acides gras et un squelette glycérol (et son éventuelle tête polaire). Ces différentes "entités" sont alors consommées dans différentes proportions dans la biomasse, permettant ainsi de prendre en compte à la fois les profils des classes de glycérolipides et les profils des acides gras (Sánchez et al. 2019; Alves et al. 2021; Correa et al. 2020).



**Figure 1.18. Différents formalismes de simplification de la composition lipidique de la biomasse (d'après Sánchez et al. 2019).**

(A) Tous les glycérolipides ont le même profil en acide gras. (B) Seuls les profils glycérolipidiques sont pris en compte. (C) Les profils d'acides gras (2a) et des glycérolipides (2b) sont pris en compte par le formalisme SLIME qui laisse flexible les entités de lipides spécifiques à produire.

Grâce aux avancées des méthodes de reconstruction et d'analyses de modèles métaboliques à l'échelle du génome, ces derniers sont de plus en plus utilisés pour étudier le métabolisme des lipides, notamment chez les levures, les microalgues, les plantes ou encore chez l'homme (Nielsen 2009; Sánchez et al. 2019; Correa et al. 2020; Alves et al. 2021; Córdoba et al. 2023; Cunha et al. 2023). Ils peuvent ainsi contribuer à une meilleure compréhension du métabolisme lipidique, à la manière dont ce dernier s'intègre et interagit avec les autres voies métaboliques, à la conception de souches et à l'optimisation des conditions de cultures en vue d'obtenir des phénotypes lipidiques d'intérêt.

### 1.3.5.1 Limites de la modélisation des lipides au sein des GEM de *M. gaditana*

La modélisation du métabolisme lipidique dans les deux modèles de *M. gaditana* disponibles au début de la thèse présente des limitations. En effet, le premier modèle disponible dans la littérature: iRJ1321 (Shah et al. 2017), se caractérise par une simplification notable de la modélisation du métabolisme lipidique. En effet, pour chaque glycérolipide les acides gras aux positions *sn-1*, *sn-2* (et *sn-3* pour les TAG) sont identiques. Par exemple le tag140\_c correspond au TAG 14:0/14:0/14:0, tag160\_c au TAG 16:0/16:0/16:0 etc. Ce modèle ne permet donc pas de représenter la variété des lipides observée chez *M. gaditana* (Jouhet et al. 2017). De plus, ce modèle ne comporte que quatre compartiments (le milieu extra-cellulaire, le cytoplasme, la mitochondrie et le chloroplaste). Le RE où une part significative de la synthèse des glycérolipides est réalisée, ainsi que le peroxyosome où les acides gras sont dégradés, ne sont pas présents dans le modèle. Enfin, ce modèle se caractérise par l'utilisation d'identifiants génériques pour l'ensemble des réactions du modèle. Les réactions sont identifiées "NanoG0001",

"NanoG0002" etc., et ne comportent pas d'identifiants de bases de données de réactions métaboliques classiques (KEGG ou BiGG etc.). Ces limitations rendent particulièrement complexe l'exploration, la vérification du contenu ainsi que l'utilisation du modèle iRJ1321.

Le second modèle, iMgadiit\_SilicoLife, quant à lui comporte une description partielle du métabolisme lipidique. Les voies de synthèse des acides gras et des glycérolipides sont présentes dans le modèle mais comportent diverses incohérences vis-à-vis de l'état des connaissances actuelles du métabolisme lipidique de *M. gaditana* : (1) la présence d'acides gras 16:2 et 16:3, (2) l'absence des voies de synthèse, élongation et désaturation cytosolique des acides gras, (3) la dégradation des acides gras n'est présente que dans le peroxyosome et seulement pour certains acides gras. Cette voie de dégradation, n'ayant pas subi de vérification manuelle (notamment de complétion), n'est donc pas fonctionnelle : les réactions la composant sont bloquées. (4) La dégradation des glycérolipides n'est pas décrite dans ce modèle. (5) Les couples *sn-1*, *sn-2* (et *sn-3* pour les TAG) décrits dans le modèle ne correspondent pas aux glycérolipides identifiés chez *M. gaditana*.

Cependant, malgré ces limitations, iMgadiit\_SilicoLife offre une base solide pour une description détaillée du métabolisme lipidique, notamment de par la présence des compartiments les plus importants pour le métabolisme lipidique (et le métabolisme central) : le chloroplaste (séparé en stroma et lumen), le RE, la mitochondrie (séparé en une matrice et l'espace intermembranaire), le peroxyosome, le cytosol ainsi qu'un milieu extracellulaire. De plus, sa reconstruction a été basée sur une ré-annotation automatique du génome de *M. gaditana* par les pipelines développés par SilicoLife et manuellement validée. La nomenclature BiGG a été utilisée pour les identifiants des métabolites et des réactions qui comportent également des annotations de bases de données classiquement utilisées notamment KEGG, EC Number, et MetaNetX. Les caractéristiques intrinsèques de qualité de ce modèle ont été validées avec MEMOTE et sont cohérentes avec la qualité actuelle des modèles métaboliques (Lieven et al. 2020). Une première étape de validation préliminaire à partir de taux de croissance de la littérature a également été effectuée par SilicoLife. De plus, les coefficients stœchiométriques de la photosynthèse ont été corrigés et validés par des experts du domaine du LPCV en amont de la thèse. Pour l'ensemble de ces raisons, le modèle iMgadiit\_SilicoLife a été utilisé comme modèle de départ pour cette thèse.

## 1.4 Objectifs de la thèse

Ce travail de thèse a pour objectif principal d'intégrer les connaissances les plus avancées du métabolisme lipidique de *M. gaditana* provenant des experts du LPCV, de TotalEnergies et de la littérature, dans le modèle métabolique reconstruit par SilicoLife, puis d'exploiter ce modèle pour approfondir notre compréhension actuelle du métabolisme lipidique de cette microalgue. Pour cela, nous mettrons particulièrement l'accent sur l'utilisation des profils lipidiques expérimentaux détaillés de *M. gaditana* et de ses mutants, disponibles au laboratoire.

Cet objectif repose sur la réalisation de deux sous objectifs :

- (1) Intégrer des données de composition de biomasse, notamment de données lipidomiques détaillées, au modèle métabolique de *M. gaditana* sous forme de fonctions de biomasse.

Pour atteindre cet objectif, la conception d'un outil bioinformatique capable de générer une fonction de biomasse à partir de données fournies par l'utilisateur est nécessaire. La définition de la fonction de biomasse étant une problématique récurrente lors du développement et de l'utilisation d'un GEM, l'outil développé devra permettre d'ajouter une fonction de biomasse à n'importe quel GEM.

- (2) Permettre aux biologistes du LPCV et de TotalEnergies d'exploiter le modèle métabolique de *M. gaditana*. Les biologistes doivent pouvoir facilement visualiser le contenu du modèle, modifier les compositions de biomasse selon les profils lipidiques mesurés, et réaliser des simulations de flux dans différentes conditions (différentes conditions de culture, pour différentes souches).

Le **Chapitre 2** de ce manuscrit de thèse présente le développement et la validation du nouveau modèle métabolique de *M. gaditana* ainsi que son application pour l'analyse d'un mutant publié généré par le LPCV et TotalEnergies. Le **Chapitre 3** présente BioModTool, un package python développé au cours de la thèse permettant d'ajouter de nouvelles fonctions de biomasse aux modèles métaboliques. Enfin le **Chapitre 4** reprend les outils développés et résultats obtenus lors de cette thèse et propose de nouvelles perspectives d'applications.

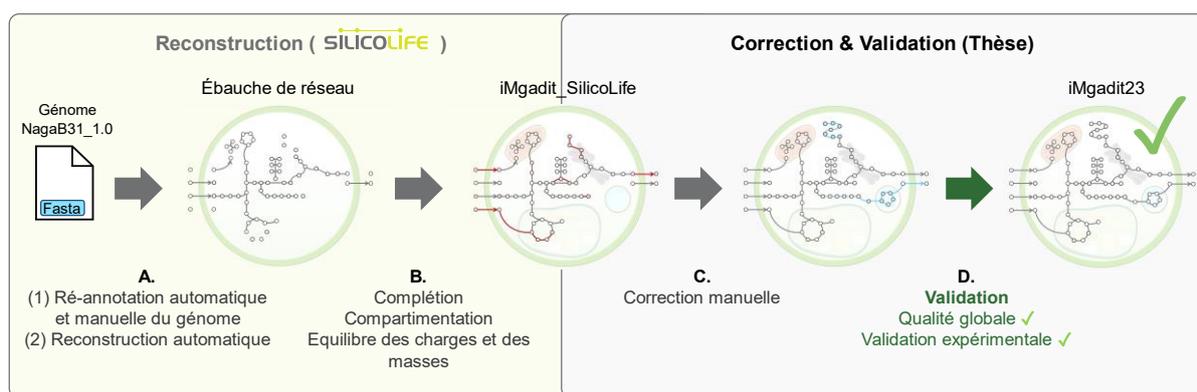
## 2 Chapitre 2

# iMgadit : modèle métabolique de *M. gaditana*

## 2.1 Préambule

*Microchloropsis gaditana* est une microalgue oléagineuse prometteuse pour la production de lipides pour diverses applications biotechnologiques. Afin d'optimiser les conditions de culture de *M. gaditana* et de développer de nouvelles souches aux profils lipidiques intéressants, une compréhension approfondie du métabolisme, notamment du métabolisme lipidique est nécessaire. Les modèles métaboliques à l'échelle du génome (GEM) constituent des outils puissants pour étudier le métabolisme d'un organisme d'intérêt. Un modèle métabolique, au métabolisme lipidique détaillé, de *M. gaditana* a donc été reconstruit pour répondre à ces problématiques.

La reconstruction de la première version du modèle métabolique de *M. gaditana* a été générée en amont de la thèse par l'entreprise SilicoLife (<https://www.silicolife.com/>). Dans la suite de ce manuscrit, le terme "iMgadit\_SilicoLife" sera employé pour désigner cette première version du modèle. Le modèle a ensuite été corrigé manuellement à partir de données et de connaissances disponibles en interne et dans la littérature. La Figure 2.1 présente un résumé du processus de reconstruction et de correction du modèle.



**Figure 2.1. Processus de reconstruction, correction et validation du modèle métabolique de *M. gaditana*.**

Une première version du GEM iMgadit\_SilicoLife a été reconstruite par SilicoLife, impliquant des étapes de réannotation du génomes, de reconstruction automatique d'une ébauche de réseau (A) et une première étape de correction (B). Le modèle a ensuite été manuellement corrigé (C) et validé (D) au cours de la thèse. Ces étapes de corrections et validations ont été répétées dans un processus itératif.

Ce chapitre est divisé en trois sections. (1) La présentation du processus de reconstruction du modèle métabolique ainsi que des choix de modélisation effectués. (2) La présentation de la version "publiée" du modèle, iMgadiT23, sa validation et son application à l'étude du phénotype d'un mutant du laboratoire. Cette partie fait l'objet d'un article (soumis). (3) Une discussion des deux parties précédentes.

## 2.2 Reconstruction et correction manuelle du modèle

### 2.2.1 Reconstruction de la première version du modèle par SilicoLife

La reconstruction de iMgadiT\_SilicoLife a été réalisée en quatre étapes (Figure 2.1). Dans un premier temps le génome de *M. gadiTana B-31* (NagaB31\_1.0, numéro d'accès GenBank GCA\_000569095.1, Corteggiani Carpinelli et al. 2014) a été réannoté afin d'assigner des fonctions métaboliques aux gènes. Les réactions biochimiques ont ensuite été compilées afin de générer une première ébauche de réseau métabolique. Ce réseau a ensuite été corrigé une première fois, impliquant notamment une étape de complétion durant laquelle des réactions ont été ajoutées au modèle afin de rendre fonctionnelles les différentes voies métaboliques. Les réactions et métabolites du réseau ont ensuite été répartis dans différents compartiments: le milieu extracellulaire, le cytosol, le chloroplaste (séparé en stroma et lumen), le réticulum endoplasmique (RE), la mitochondrie (séparée en une matrice et un espace intermembranaire (EIM)) et le peroxysoxe. Une fonction de biomasse a été ajoutée au modèle et une première étape de validation a été réalisée. Les caractéristiques intrinsèques de qualité ont été vérifiées avec MEMOTE (Lieven et al. 2020). Et une première étape de validation préliminaire à partir de taux de croissances de la littérature a été effectuée par SilicoLife.

Une description plus détaillée du processus de reconstruction de SilicoLife est présentée dans Pereira et al. 2018.

### 2.2.2 Correction manuelle du modèle

#### 2.2.2.1 Méthodes

##### **Environnement logiciel**

Toutes les modifications apportées au modèle ont été exécutées dans un environnement Conda (Anaconda 2016) spécifique, en utilisant la bibliothèque open-source COBRAPy (version 0.21.0) (Ebrahim et al. 2013) avec Python (version 3.6.13) et le solveur CPLEX (version 12.8.0.0). Après chaque modification, la suite de tests MEMOTE (Platform Windows, Memote Version 0.12.0) a été employée,

en ligne de commande, pour garantir le suivi de la qualité du modèle (Lieven et al. 2020). Une carte métabolique détaillée représentant l'intégralité du contenu du modèle a été dessinée à l'aide de l'application web Escher (King et al. 2015) et régulièrement actualisée à mesure des corrections apportées au modèle.

### **Suivis des modifications du modèle**

Chaque correction apportée au modèle métabolique a été réalisée au sein d'un script Jupyter Notebook dédié (Kluyver et al. 2016). Un identifiant unique a été assigné à chaque correction afin de permettre de croiser les informations des différents fichiers de suivi des corrections. Les informations relatives aux réactions, métabolites et gènes ajoutés, supprimés ou modifiés ont été enregistrés dans des fichiers CSV (added\_reactions.csv, removed\_reactions.csv, modified\_reactions.csv, new\_id\_reactions.csv, added\_metabolites.csv etc.). Toutes les corrections apportées au modèle ont également été consignées dans un rapport interne, associé à un fichier Excel détaillé. Ce fichier Excel documente de manière exhaustive et justifie l'ensemble des modifications apportées au modèle. Il détaille notamment les informations suivantes:

1. Description de la modification effectuée:
  - L'identifiant unique de la modification.
  - Le type de modification (Ajout de réaction, Correction de GPR etc.).
  - La voie métabolique concernée par la correction.
  - Une description succincte de la modification effectuée.
  - La liste des identifiants des réactions et/ou des métabolites concernés par la modification.
  - Les arguments justifiant cette modification (arguments des experts et/ou issus de la littérature).
  - Les références bibliographiques associées à la modification.
  - La version du modèle sur laquelle a été effectuée la modification.
2. Validation de la modification:
  - Les résultats des tests MEMOTE les plus pertinents.
  - Le résultat du test de cohérence stœchiométrique excluant les photons (la présence de réactions impliquant les photons faisant échouer le test de cohérence stœchiométrique en raison de leur absence de formule chimique).
  - Le résultat des tests de production de chaque métabolite consommé dans la réaction de biomasse.

Afin d'assurer la traçabilité, les modifications ont été consignées dans un dépôt GitHub (privé), permettant également d'assurer le suivi des différentes versions du modèle.

Pour des raisons de confidentialité ces différents documents ne peuvent pas être fournis.

### 2.2.2.2 Définitions et conventions

Les modèles métaboliques sont composés de différents sous-composants, notamment des gènes, des métabolites et des réactions. Les métabolites et les réactions sont associés à un compartiment donné. Ainsi, une entité chimique (par exemple l'ATP) présente dans divers compartiments cellulaires *in vivo* est associée, dans le modèle, à plusieurs métabolites (atp\_c, atp\_m, atp\_h etc., le suffixe représentant le compartiment). Il en va de même pour les réactions.

Différents attributs tels qu'un identifiant unique, un nom, des annotations et des notes sont associés au modèle et à chaque sous-composant. Ces différents attributs et les conventions correspondantes sont présentés ci-dessous (selon le formalisme de la librairie COBRApy (Ebrahim et al. 2013)).

**- Identifiant (ID) :** Les identifiants sont des chaînes de caractères alphanumériques (les underscores (symbole : `_`) peuvent également être utilisés). Un identifiant doit permettre d'identifier de manière unique un métabolite, une réaction ou un gène au sein du modèle.

- **Modèle :**

Deux conventions sont généralement utilisées pour identifier les modèles métaboliques (Carey et al. 2020).

- Convention historique : i + initiales des auteurs + nombre de gènes dans le modèle.
- Convention recommandée : i + indicateur de l'espèce + indicateur d'itération (par exemple l'année de publication du modèle). C'est cette convention qui a été sélectionnée pour la version "publiée" du modèle : iMgadit23.

- **Gènes :**

Les identifiants des gènes correspondent généralement aux numéros d'accèsion GenBank ou aux identifiants des gènes dans une base de données spécifique de l'organisme (Ravikrishnan et Raman 2015).

- **Métabolites et réactions biochimiques :**

BiGG est une base de données spécifique aux GEM (King et al. 2016; Norsigian et al. 2020). Les identifiants BiGG (BiGG ID) présentent les avantages d'être uniques, bien définis, lisibles par l'homme et mémorisables. Des spécifications et des directives relatives à la définition de ces identifiants sont fournies par BiGG. Les BiGG ID (1) peuvent contenir entre 3 et 60 caractères bien que les identifiants courts soient privilégiés, (2) ne doivent contenir que des lettres majuscules et minuscules, des chiffres

et des underscores et (3) ne doivent pas commencer par un underscore. (4) Les identifiants des réactions sont préférentiellement en majuscules et ceux des métabolites en minuscule. (5) Le compartiment du métabolite ou de la réaction doit être indiqué via les deux derniers caractères de l'identifiant, à savoir "\_x" où x correspond à l'abréviation du compartiment. (6) Les réactions d'échange, de demande et de biomasses sont respectivement identifiées par les préfixes "EX\_", "DM\_" et "BIOMASS\_".

- **Nom** : Les noms sont des chaînes de caractères qui ne sont pas soumises à des règles strictes en ce qui concerne leur définition. Les noms permettent d'explicitier l'identifiant et correspondent souvent au nom chimique du métabolite ou de la réaction. Ils peuvent englober différents alias, synonymes et abréviations.

- **Annotation** : Les annotations sont des dictionnaires Python permettant d'ajouter des références croisées à diverses bases de données (telles que KEGG (Kanehisa et Goto 2000b), MetaNetX (Bernard et al. 2014; Moretti et al. 2021) etc.) et des termes d'ontologie (SBO, Systems Biology Ontology) (Courtot et al. 2011). La clé du dictionnaire correspond à l'identifiant de la base de données, la valeur associée à l'identifiant de l'entité (par exemple du métabolite) dans la base de données. L'utilisation d'identifiants répertoriés dans Identifiers.org (<https://registry.identifiers.org/registry>) permet la création d'URL vers les bases de données spécialisées lors de l'enregistrement du modèle au format SBML (Systems Biology Markup Language). Par exemple, les annotations {'bigg.metabolite': 'atp', 'kegg.compound': 'C00002'} permettent la création de liens vers l'ATP dans les bases de données BiGG et KEGG Compound dans le format SBML du modèle:

- BiGG: <http://bigg.ucsd.edu/models/universal/metabolites/atp>

- KEGG Compound: <https://identifiers.org/kegg.compound/C00002>

- **Notes** : Les notes sont des dictionnaires permettant d'ajouter des informations supplémentaires telles que des références bibliographiques, des remarques etc. de manière à être lisible par l'homme.

### 2.2.2.3 Nomenclatures des identifiants dans iMgadiT

- Modèle

Le modèle est identifié comme i + MgadiT (indicateur de l'espèce).

Les différentes versions du modèle sont différenciées par l'utilisation d'un indicateur d'itération différent :

- Version reconstruite par SilicoLife : iMgadiT\_SilicoLife
- Version soumise à publication en 2023 : iMgadiT23

D'autres versions du modèle, contenant des informations confidentielles sont également disponibles. Dans ce manuscrit, la dénomination iMgadit, sans indicateur d'itération fait (indifféremment) référence à iMgadit23 et/ou à une version confidentielle.

- Gènes

Les numéros d'accession GenBank du génome de *M. gaditana B-31* (numéro d'accès GenBank GCA\_000569095.1) ont été utilisés comme identifiants des gènes dans iMgadit.

- Compartiments

Le modèle comporte huit compartiments. Les abréviations de ces compartiments sont récapitulées dans la Table 2.1.

**Table 2.1. Dénomination des compartiments présents dans iMgadit.**

Abréviation	Compartiment
c	cytosol
e	milieu extracellulaire
i	espace intermembranaire de la mitochondrie
m	matrice mitochondriale
h	stroma du chloroplaste
l	lumen du chloroplaste
p	peroxysome
r	réticulum endoplasmique

- Métabolites et réactions

La majorité des identifiants des métabolites et des réactions du modèle métabolique iMgadit\_SilicoLife sont issus de BiGG. Cependant les identifiants BiGG sont peu adaptés à la modélisation des lipides. Une nouvelle nomenclature pour les métabolites et les réactions du métabolisme lipidique a donc été mise en place afin de définir des identifiants compréhensibles par l'homme.

#### 2.2.2.4 Nomenclatures des lipides

Dans cette section, les nomenclatures des lipides classiquement utilisées par la communauté scientifique seront présentées. A partir de ces nomenclatures et des spécifications de BiGG, une nouvelle nomenclature pour les identifiants (des métabolites et des réactions) du métabolisme lipidique est proposée. Cette nomenclature a été utilisée dans iMgadit.

## Nomenclatures des acides gras :

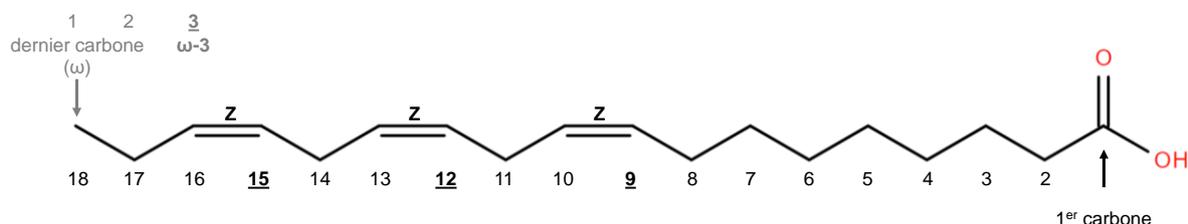
Plusieurs nomenclatures sont fréquemment utilisées pour les acides gras (IUPAC 1978; Liebisch et al. 2020).

- La **nomenclature normalisée** provenant de la chimie organique. Cette nomenclature décrit le nombre d'atomes de carbone via un préfixe en terminologie grecque (1: meth-, 2: eth-, 3: prop-, ..., 14: -tetradeca, 16: -hexadeca, 18: -octadeca etc.), la structure de la chaîne carbonée (nombre, position et configuration (*cis*, *trans*) des doubles liaisons) et la nature de la fonction chimique qui est indiquée par un suffixe (acide carboxylique "-oïque").
- La **nomenclature abrégée C:D** où C représente le nombre d'atomes de carbone et D est le nombre de doubles liaisons.
- La **nomenclature usuelle** qui se réfère souvent aux lipides en fonction de leur source d'origine ou de leur découverte.

Deux nomenclatures sont souvent utilisées (ensemble ou séparément) pour spécifier la position et la géométrie des doubles liaisons des acides gras insaturés.

- La **nomenclature chimique "Z/E"**, où Z (de l'allemand *zusammen*) indique une double liaison *cis* (c'est-à-dire lorsque les hydrogènes sont du même côté de la double liaison), et E (*entgegen*) désigne une double liaison *trans* (les hydrogènes sont de part et d'autre de la double liaison).
- La **numérotation physiologique**, qui indique la position de la première double liaison à partir du premier carbone (carbone " $\alpha$ " ou " $\Delta$ ") ou du dernier carbone (carbone " $\omega$ " ou " $n$ ").

Ces différentes nomenclatures sont illustrées à partir de la structure de l'acide gras 18:3 (9Z,12Z,15Z) (Figure 2.2). Dans la suite de ce manuscrit la nomenclature abrégée chimique sera utilisée.



**Figure 2.2. Structure de l'acide gras C18:3 (9Z,12Z,15Z) (nomenclature abrégée chimique).**

Nomenclature normalisée: acide 9Z,12Z,15Z-octadécatriénoïque ; Nomenclature usuelle: acide  $\alpha$ -linoléinique ; Nomenclature abrégée physiologique : C18:3  $\omega$ -3.

### Choix de la nomenclature pour le modèle :

Dans BiGG, dont les identifiants ont été utilisés dans iMgadiit\_SilicoLife, les identifiants des acides gras à chaînes courtes, moyennes et longues correspondent à des acronymes de la nomenclature normalisée (Table 2.2) (King et al. 2016; Norsigian et al. 2020). Cette nomenclature présente le désavantage de ne pas être facilement décryptable. D'autre part, pour certains acides gras, notamment pour les acides gras à très longue chaîne carbonée, d'autres nomenclatures sont également utilisées, créant de la disparité au sein des identifiants. Par exemple dans BiGG l'identifiant de l'acide gras 20:4-CoA (arachidonyl-CoA) est "arachdcoa", alors que le 20:1-CoA est identifié "c201coa". Cette disparité peut amener à la duplication de métabolites dans les GEM. Par exemple dans iMgadiit\_SilicoLife, le 3-hydroxyoctadecanoyl-CoA (un intermédiaire réactionnel de la synthèse et de la dégradation des acides gras) était dupliqué sous les identifiants "3hC18\_0\_CoA" et "3hodcoa". De plus, tous les acides gras et dérivés d'acide gras ne sont pas répertoriés dans BiGG.

**Table 2.2. Exemples des différentes nomenclatures des acides gras.**

Nomenclature abrégée chimique	Nomenclature usuelle	Nomenclature normalisée chimique	BiGG ID (ID dans iMgadiit_SilicoLife)	ID dans iMgadiit23
C16:0	acide palmitique	acide hexadécanoïque ( <u>hexadecanoic acid</u> en anglais)	hdca	c160
C18:0	acide stéarique	acide <u>octadécanoïque</u>	ocdca	c180
C18:1	acide oléique	acide <u>octadécènoïque</u> (9Z)	ocdcea	c181
C18:2	acide linoléique	acide octadécadiénoïque (9Z,12Z)	ocdcya	c182
C18:3	acide linoléique (linolenic en anglais)	acide octadécatriénoïque (9Z,12Z,15Z)	lnlncg	c183

Afin d'uniformiser la nomenclature au sein de iMgadiit et de rendre celle-ci plus claire et compréhensible pour l'ensemble des utilisateurs, les identifiants des acides gras et de leurs dérivés ont été standardisés. La nomenclature retenue est basée sur la nomenclature abrégée chimique, adaptée afin de respecter les spécifications des identifiants BiGG.

ID = "classe de la molécule" + "chaîne carbonée selon la nomenclature abrégée CD (suppression des ":")"

Des exemples de différentes classes d'acide gras et dérivés d'acide gras sont présentés dans la (Table 2.3). Les identifiants BiGG ont été ajoutés dans les annotations des métabolites.

**Table 2.3. Nomenclature des acides gras.**

Classe	Abréviation (préfixe)	Suffixe	Exemple (16:0)
Acide gras	c	/	c160
Acide gras – protéine porteuse d'acyle (Acyl Carrier Protein)	c	ACP	c160_ACP
Acyl-CoA	c	coa	c160_coa
Acide gras-FAS1	c	FAS1	c160_FAS1
3-oxoacyl-CoA	3oc	coa	3oc160_coa
3-hydroxyacyl-CoA	3hc	coa	3hc160_coa
Trans-2-enoyl-CoA	c	2e_coa	c161_2e_coa

### Nomenclature des glycérolipides et stratégie de modélisation des TAG

Plusieurs nomenclatures existent également pour désigner les glycérolipides (Liebisch et al. 2020). Ces nomenclatures spécifient la classe du glycérolipide ainsi que le nombre de carbones et de doubles liaisons qu'il contient, cette spécification dépendant du niveau de détail disponible. Lorsque la composition en acide gras est connue cette dernière peut être détaillée en listant le positionnement des acides gras sur le squelette glycérol dans l'ordre "(sn-1/sn-2/sn-3)". Lorsque le positionnement n'est pas connu, la composition en acide gras est indiquée "(sn-1\_sn-2\_sn-3)". En l'absence de ces informations, le nombre total de carbones et d'insaturations est indiqué. Par exemple les PC(16:1/18:1) et PC(16:0/18:2) peuvent être identifiés PC(34:2).

#### Choix de la nomenclature pour le modèle:

Dans iMgadit, la nomenclature utilisée est basée sur la nomenclature "ID de la classe (sn-1/sn-2/sn-3)", adaptée pour respecter les contraintes de modélisation. Les identifiants des glycérolipides sont donc définis selon la norme suivante:

ID = "classe du glycérolipide" + "chaîne carbonée sn-1" + "\_" + "chaîne carbonée sn-2"

**N.B:** L'utilisation des underscores dans les identifiants des glycérolipides dans iMgadit est liée à des contraintes de modélisation et non à la non-connaissance de la position des différentes chaînes carbonées. Ainsi, le PC(16:1/18:1) est identifié pc161\_181 dans iMgadit.

Les acronymes utilisés pour les différentes classes de glycérolipides sont présentés dans la Table 2.4.

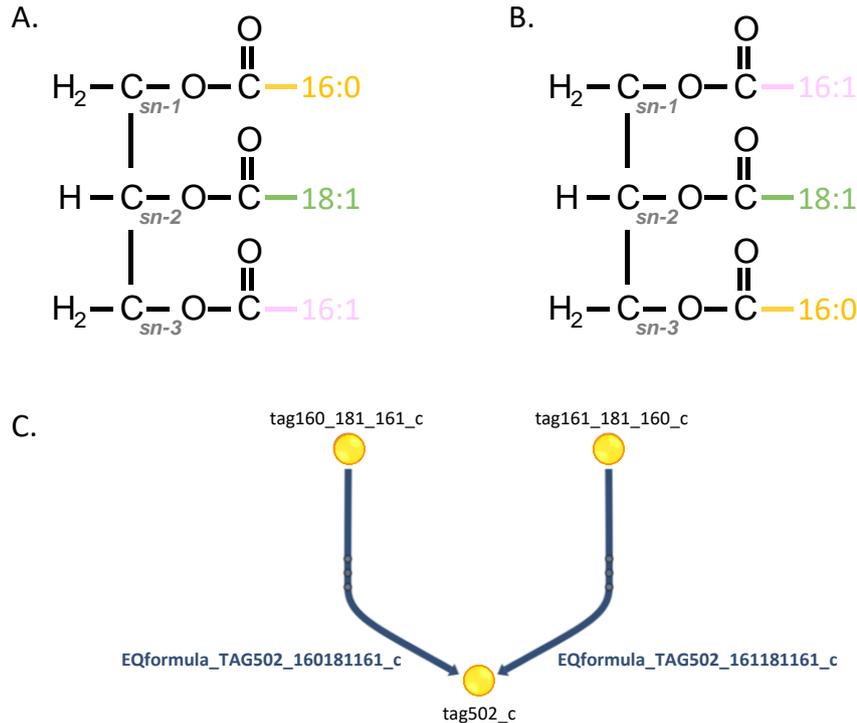
**Table 2.4. Nomenclature des différentes classes de glycérolipides.**

Classe	Abréviation
Phosphatidylcholine	pc
Phosphatidyléthanolamine	pe
Phosphatidylinositol	pi
Phosphatidylglycérol	pg
Monogalactosyldiacylglycérol	mgdg
Digalactosyldiacylglycérol	dgdg
Sulfoquinovosyldiacylglycérol	sqdg
Diacylglycérol-N,N,N-triméthylhomosérine	dgts
Monoacylglycérol	mag
Diacylglycérol	dag
Triacylglycérol	tag
acide lysophosphatidique	sn1_glyc3p
<i>sn-1</i> lyso-phosphatidylcholine	sn1_pc
<i>sn-1</i> lyso-diacylglycérol-N,N,N-triméthylhomosérine	sn1_dgts
<i>sn-2</i> lyso-monogalactosyldiacylglycérol	sn2_mgdg

Les TAG, comportant trois chaînes carbonées sont définis :

ID = tag + "chaîne carbonée *sn-1*" + "\_" + "chaîne carbonée *sn-2*" + "\_" + "chaîne carbonée *sn-3*"

De par la symétrie de la structure chimique des TAG, la molécule de TAG (*sn-1/sn-2/sn-3*) est identique à la molécule de TAG (*sn-3/sn-2/sn-1*), ce qui n'est pas le cas dans iMgadit où les tagsn1\_sn2\_sn3 et tagsn3\_sn2\_sn1 constituent des métabolites distincts en raison de leurs voies de synthèse qui sont distinctes (Figure 2.3 A et B). Afin de pallier cette limitation, des pseudo-réactions d'équivalence (identifiées par le préfixe "EQformula\_") ont été ajoutées au modèle. Ces pseudo-réactions convertissent des TAG identiques en un TAG unique (identifié par le nombre total de carbones et de doubles liaisons). Par exemple, pour représenter l'équivalence entre les tag160\_181\_161 et tag161\_181\_180, deux réactions convertissant respectivement ces TAG en pseudo-TAG tag502 (50 carbones et 2 insaturations) ont été ajoutées au modèle (EQformula\_TAG502\_160181161\_c et EQformula\_TAG502\_161181160\_c) (Figure 2.3C).



**Figure 2.3. Exemple de la stratégie de modélisation des TAG dans iMgadiT.**

(A) Structure du tag160\_181\_161 (TAG(16:0/18:1/16:1)). (B) Structure du tag161\_181\_160 (TAG(16:1/18:1/16:0)). (C) Réactions EQformula ajoutées au modèle.

### 2.2.2.5 Modélisation du métabolisme lipidique

L'intégralité des voies de synthèse et de dégradation des acides gras et des glycérolipides présentes dans iMgadiT a été manuellement corrigée à l'aide de connaissances d'experts du domaine du LPCV et issues de recherches bibliographiques. La méthodologie employée pour corriger le métabolisme lipidique est explicitée ci-dessous à partir de l'exemple de la synthèse chloroplastique des glycérolipides.

Dans un premier temps, les réactions génériques du métabolisme des lipides ont été déterminées (Table 2.5). La synthèse des glycérolipides dans le chloroplaste implique l'acétylation du glycérol-3-phosphate (glyc3p) par des glycérol-3-phosphate acyl-ACP acyltransférases (ATS1 et ATS2) produisant un acide phosphatidique, précurseur de la synthèse : (1) du phosphatidylglycérol (PG) via la CDPDAG synthase (CDPDAGS), la phosphatidylglycérol phosphate (PGP) synthase (PGSA), et la PGP phosphatase (PGPP), (2) du diacylglycérol (DAG) via l'acide phosphatidique phosphatase (PAP). Le DAG est ensuite utilisé pour la synthèse de sulfoquinovosyldiacylglycérol (SQDG), de monogalactosyldiacylglycérol (MGDG) et de digalactosyldiacylglycérol (DGDG) respectivement via la SQDG Synthase (SQDGS), MGDG Synthase (MGDGS) and DGDG Synthase (DGDGS) (Li-Beisson et al. 2019). Les gènes codant les enzymes catalysant les différentes réactions ont été ajoutés. Étant donné

l'absence de caractérisation de la spécificité des enzymes du métabolisme lipidique chez *M. gaditana*, les gènes ont été associés aux réactions du métabolisme lipidique indépendamment des profils en acides gras.

Une fois la liste des réactions établie, des données lipidomiques spécifiques de *M. gaditana* ont été utilisées afin de déterminer l'ensemble des arrangements *sn-1/sn-2* (et *sn-1/sn-2/sn-3* pour les TAG) existants. Les réactions nécessaires au métabolisme de l'ensemble de ces lipides spécifiques ont été ajoutées au modèle. Ainsi, les réactions présentées dans la Table 2.5 ont été ajoutées au modèle pour l'ensemble des couples listés dans la Table 2.6. Certains de ces lipides ne sont pas présents, ou sont présents en trop faible quantité pour être dosés dans les analyses lipidomiques, mais constituent des intermédiaires réactionnels pour la synthèse d'autres lipides. La même stratégie a été suivie pour l'ensemble du métabolisme lipidique.

**Table 2.5. Réactions génériques décrivant la synthèse chloroplastique des glycérolipides dans iMgadii.**

"sn1" et "sn2" indiqués en bleu et vert clairs, ne sont présents que dans les noms génériques et sont substitués par les chaînes carbonées correspondantes dans les identifiants du modèle. Par exemple le métabolite *csn1*\_ACP\_h correspond aux métabolites *c140*\_ACP\_h, *c160*\_ACP\_h etc. dans le modèle. Lorsque "sn1" ou "sn2" est indiqué en noir, il figure dans l'identifiant du modèle, par exemple *sn1\_glyc3p**sn1*\_h correspond aux métabolites *sn1\_glyc3p140*\_h, *sn1\_glyc3p160*\_h etc.

Identifiants des métabolites: glycerol-3-phosphate (*glyc3p*), cytidine triphosphate (*ctp*), diphosphate (*ppi*), proton (*h*), UDP-alpha-D-galactose (*udpgal*), uridine diphosphate (*udp*), UDP-6-sulfoquinovose (*udpsq*).

Identifiant	Nom (anglais)	Equation	GPR
ATS1_ <i>sn1</i> _h	Glycerol-3-phosphate acyl-ACP acyltransferase	<i>csn1</i> _ACP_h + <i>glyc3p</i> _h --> ACP_h + <i>sn1_glyc3p</i> <i>sn1</i> _h	Naga_100106g21
ATS2_ <i>sn2</i> _h	Lysophosphatidic acid acyltransferase	<i>csn2</i> _ACP_h + <i>sn1_glyc3p</i> <i>sn1</i> _h --> ACP_h + <i>pa</i> <i>sn1</i> _ <i>sn2</i> _h	Naga_100007g86 or Naga_100015g9
CDPDAGS <i>sn1sn2</i> _h	CDPDAG synthase	<i>ctp</i> _h + <i>h</i> _h + <i>pa</i> <i>sn1</i> _ <i>sn2</i> _h --> <i>cdpdags</i> <i>sn1</i> _ <i>sn2</i> _h + <i>ppi</i> _h	Naga_100003g10 or Naga_100002g146
PGSAs <i>sn1sn2</i> _h	Phosphatidylglycerol phosphate synthase	<i>cdpdag</i> <i>sn1</i> _ <i>sn2</i> _h + <i>glyc3p</i> _h --> <i>cmp</i> _h + <i>h</i> _h + <i>pgp</i> <i>sn1</i> _ <i>sn2</i> _h	Naga_100081g13 or Naga_101156g2 or Naga_100114g5
PGPP <i>sn1sn2</i> _h	Phosphatidylglycerol phosphate phosphatase	<i>h2o</i> _h + <i>pgp</i> <i>sn1</i> _ <i>sn2</i> _h --> <i>pg</i> <i>sn1</i> _ <i>sn2</i> _h + <i>pi</i> _h	
PAP <i>sn1sn2</i> _h	Phosphatidic acid phosphatase	<i>h2o</i> _h + <i>pa</i> <i>sn1</i> _ <i>sn2</i> _h --> <i>dag</i> <i>sn1</i> _ <i>sn2</i> _h + <i>pi</i> _h	
MGDGS <i>sn1sn2</i> _h	MGDG synthase	<i>dag</i> <i>sn1</i> _ <i>sn2</i> _h + <i>udpgal</i> _h --> <i>h</i> _h + <i>mgdgs</i> <i>sn1</i> _ <i>sn2</i> _h + <i>udp</i> _h	Naga_100092g5
DGDGS <i>sn1sn2</i> _h	DGDG synthase	<i>mgdgs</i> <i>sn1</i> _ <i>sn2</i> _h + <i>udpgal</i> _h --> <i>dgdgs</i> <i>sn1</i> _ <i>sn2</i> _h + <i>h</i> _h + <i>udp</i> _h	Naga_100010g107
SQDGS <i>sn1sn2</i> _h	SQDG synthase	<i>dag</i> <i>sn1</i> _ <i>sn2</i> _h + <i>udpsq</i> _h --> <i>h</i> _h + <i>sqdgs</i> <i>sn1</i> _ <i>sn2</i> _h + <i>udp</i> _h	Naga_100010g28 or Naga_100208g9 or Naga_100008g19

**Table 2.6. Liste des couples sn-1/sn-2 identifiés dans les glycérolipides chloroplastiques de *M. gaditana*.**

sn-1	sn-2
14:0	16:0
14:0	16:1
16:1	16:0
16:1	16:1
20:5	16:0
20:5	16:1
20:5	20:5
20:5	20:4
20:5	14:0

### 2.2.2.6 Modélisation de la biomasse

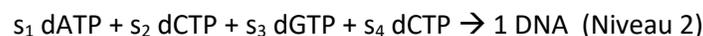
Comme présenté en introduction, il existe plusieurs formalismes pour modéliser la biomasse, notamment pour décrire sa composition lipidique. Des glycérolipidomes détaillés étant disponibles au LCPV, le profil en glycérolipides ainsi que les profils en acides gras de chaque classe de glycérolipides sont pris en compte dans la fonction de biomasse. Ainsi, dans iMgadit, la biomasse est organisée en trois niveaux.

**Le premier niveau** consomme les macromolécules principales de la cellule, telles que l'ADN, l'ARN, les protéines, les glucides, les lipides, etc. :



Où  $s_i$  (souvent exprimé en  $\text{mmol.gDW}^{-1}$ ) correspond à l'abondance de la macromolécule (ou du pool de molécules)  $i$  dans la biomasse cellulaire. ADN, ARN, PROTÉINES, etc. sont des pseudo-métabolites représentant les macromolécules.

**Le niveau 2** correspond à la consommation des monomères (les désoxynucléotides triphosphates (dNTPs), les acides aminés etc.), symbolisant la polymérisation des macromolécules. À titre d'exemple, l'ADN est un pseudo-métabolite produit par une pseudo-réaction qui consomme les quatre dNTPs :

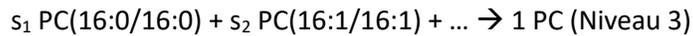


Où  $s_i$  désigne la fraction molaire de chaque métabolite au sein de la macromolécule (ici en  $\text{mol.mol}_{\text{ADN}}^{-1}$ ).

**Le niveau 2 – lipides** consomme les différentes classes de glycérolipides :



Où  $s_i$  représente la fraction molaire de la classe de lipides dans l'ensemble des lipides ( $\text{mol}_{\text{classe}} \cdot \text{mol}_{\text{LIPIDES}}^{-1}$ ). PC (phosphatidylcholine), PE (phosphatidyléthanolamine), etc. sont des pseudo-métabolites représentant chaque classe de lipides qui doivent à leur tour être produits par des pseudo-réactions supplémentaires de **niveau 3**. Par exemple, pour PC :



Où  $s_i$  représente la fraction molaire de chaque lipide spécifique dans sa classe (ici en  $\text{mol} \cdot \text{mol}_{\text{PC}}^{-1}$ ).

Chaque fonction de biomasse de iMgadiT est donc composée de 17 pseudo-réactions (Table 2.7). L'organisation de la fonction de biomasse en différents niveaux permet de clarifier sa composition.

**Table 2.7. Structure des fonctions de biomasse dans iMgadiT.**

*Différentes fonctions de biomasses ayant été ajoutées au modèle, chacune d'elle est identifiée par un suffixe unique (par exemple "biomass\_WT526").*

Pseudo-réaction	Niveau	Unité des coefficients stœchiométriques
BIOMASS_suffix_c	Niveau 1	$\text{mmol} \cdot \text{g}_{\text{DW}}^{-1}$
DNA_suffix_c	Niveau 2	$\text{mol} \cdot \text{mol}_{\text{ADN}}^{-1}$
RNA_suffix_c	Niveau 2	$\text{mol} \cdot \text{mol}_{\text{ARN}}^{-1}$
CARBOHYDRATES_suffix_c	Niveau 2	$\text{mol} \cdot \text{mol}_{\text{CARBOHYDRATES}}^{-1}$
PROTEINS_suffix_c	Niveau 2	$\text{mol} \cdot \text{mol}_{\text{PROTEINES}}^{-1}$
PIGMENTS_suffix_c	Niveau 2	$\text{mol} \cdot \text{mol}_{\text{PIGMENTS}}^{-1}$
LIPIDS_suffix_c	Niveau 2 - lipides	$\text{mol} \cdot \text{mol}_{\text{LIPIDES}}^{-1}$
PC_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{PC}}^{-1}$
PE_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{PE}}^{-1}$
PI_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{PI}}^{-1}$
PG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{PG}}^{-1}$
MGDG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{MGDG}}^{-1}$
DGDG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{DGDG}}^{-1}$
SQDG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{SQDG}}^{-1}$
DGTS_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{DGTS}}^{-1}$
DAG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{DAG}}^{-1}$
TAG_suffix_c	Niveau 3	$\text{mol} \cdot \text{mol}_{\text{TAG}}^{-1}$

Le LPCV possède une plateforme de lipidomique spécialisée dans l'analyse des glycérolipides. Des glycérolipidomes de *M. gaditana* sauvage ainsi que de différentes souches mutantes cultivées dans diverses conditions sont disponibles. La technologie employée ne permet pas la caractérisation individuelle de chaque acide gras présent au sein des glycérolipides, cependant, elle permet de déterminer le nombre total de carbones et d'insaturations présents dans les acides gras composant un lipide donné. La majorité des lipides correspondent de manière exclusive à une unique entité lipidique du modèle, néanmoins, certains lipides peuvent être associés à plusieurs entités lipidiques. Afin de résoudre cette problématique d'ambiguïté et d'éviter d'opter de manière arbitraire pour une entité lipidique spécifique plutôt qu'une autre, des réactions d'équivalence ont été ajoutées au modèle (Table 2.8).

**Table 2.8. Détails des pseudo-réactions EQformula présentes dans le modèle.**

Lipides (format LPCV)	Lipides spécifiques correspondants	Réaction Eqformula ajoutée à iMgadiT	Equation de la réaction
DGTS 32:1	DGTS 16:0/16:1	EQformula_DGTS321_160161_c	dgts160_161_c --> dgts321_c
	DGTS 16:1/16:0	EQformula_DGTS321_161160_c	dgts161_160_c --> dgts321_c
DGTS 36:4	DGTS 20:3/16:1	EQformula_DGTS364_203161_c	dgts203_161_c --> dgts364_c
	DGTS 20:4/16:0	EQformula_DGTS364_204160_c	dgts204_160_c --> dgts364_c
DGTS 36:5	DGTS 20:4/16:1	EQformula_DGTS365_204161_c	dgts204_161_c --> dgts365_c
	DGTS 20:5/16:0	EQformula_DGTS365_205160_c	dgts205_160_c --> dgts365_c
DGTS 40:7	DGTS 20:3/20:4	EQformula_DGTS407_203204_c	dgts203_204_c --> dgts407_c
	DGTS 20:4/20:3	EQformula_DGTS407_204203_c	dgts204_203_c --> dgts407_c
	DGTS 20:3/20:5	EQformula_DGTS408_203205_c	dgts203_205_c --> dgts408_c
DGTS 40:8	DGTS 20:4/20:4	EQformula_DGTS408_204204_c	dgts204_204_c --> dgts408_c
	DGTS 20:5/20:3	EQformula_DGTS408_205203_c	dgts205_203_c --> dgts408_c
DGTS 40:9	DGTS 20:4/20:5	EQformula_DGTS409_204205_c	dgts204_205_c --> dgts409_c
	DGTS 20:5/20:4	EQformula_DGTS409_205204_c	dgts205_204_c --> dgts409_c
PC 32:1	PC 16:0/16:1	EQformula_PC321_160161_c	pc160_161_c --> pc321_c
	PC 16:1/16:0	EQformula_PC321_161160_c	pc161_160_c --> pc321_c
PC 34:1	PC 16:0/18:1	EQformula_PC341_160181_c	pc160_181_c --> pc341_c
	PC 16:1/18:0	EQformula_PC341_161180_c	pc161_180_c --> pc341_c
PC 34:2	PC 16:0/18:2	EQformula_PC342_160182_c	pc160_182_c --> pc342_c
	PC 16:1/18:1	EQformula_PC342_161181_c	pc161_181_c --> pc342_c
PC 34:3	PC 16:0/18:3	EQformula_PC343_160183_c	pc160_183_c --> pc343_c
	PC 16:1/18:2	EQformula_PC343_161182_c	pc161_182_c --> pc343_c
PC 36:4	PC 20:3/16:1	EQformula_PC364_203161_c	pc203_161_c --> pc364_c
	PC 20:4/16:0	EQformula_PC364_204160_c	pc204_160_c --> pc364_c
PC 36:5	PC 20:4/16:1	EQformula_PC365_204161_c	pc204_161_c --> pc365_c
	PC 20:5/16:0	EQformula_PC365_205160_c	pc205_160_c --> pc365_c
PE 40:7	PE 20:3/20:4	EQformula_PE407_203204_c	pe203_204_c --> pe407_c
	PE 20:4/20:3	EQformula_PE407_204203_c	pe204_203_c --> pe407_c

PE 40:8	PE 20:3/20:5	EQformula_PE408_203205_c	pe203_205_c --> pe408_c
	PE 20:4/20:4	EQformula_PE408_204204_c	pe204_204_c --> pe408_c
	PE 20:5/20:3	EQformula_PE408_205203_c	pe205_203_c --> pe408_c
PE 40:9	PE 20:4/20:5	EQformula_PE409_204205_c	pe204_205_c --> pe409_c
	PE 20:5/20:4	EQformula_PE409_205204_c	pe205_204_c --> pe409_c
PG 32:1	PG 16:0/16:1	EQformula_PG321_160161_c	pg160_161_c --> pg321_c
	PG 16:1/16:0	EQformula_PG321_161160_c	pg161_160_c --> pg321_c
PI 32:1	PI 16:0/16:1	EQformula_PI321_160161_c	pi160_161_c --> pi321_c
	PI 16:1/16:0	EQformula_PI321_161160_c	pi161_160_c --> pi321_c
TAG 46:0	TAG 14:0/16:0/16:0	EQformula_TAG460_140160160_c	tag140_160_160_c --> tag460_c
	TAG 16:0/16:0/14:0	EQformula_TAG460_160160140_c	tag160_160_140_c --> tag460_c
TAG 46:1	TAG 14:0/16:1/16:0	EQformula_TAG461_140161160_c	tag140_161_160_c --> tag461_c
	TAG 16:1/16:0/14:0	EQformula_TAG461_160161140_c	tag160_161_140_c --> tag461_c
TAG 46:2	TAG 14:0/16:1/16:1	EQformula_TAG462_140161161_c	tag140_161_161_c --> tag462_c
	TAG 16:1/16:1/14:0	EQformula_TAG462_161161140_c	tag161_161_140_c --> tag462_c
TAG 48:1	TAG 16:0/16:0/16:1	EQformula_TAG481_160160161_c	tag160_160_161_c --> tag481_c
	TAG 16:1/16:0/16:0	EQformula_TAG481_161160160_c	tag161_160_160_c --> tag481_c
TAG 48:2	TAG 16:0/16:1/16:1	EQformula_TAG482_160161161_c	tag160_161_161_c --> tag482_c
	TAG 16:1/16:1/16:0	EQformula_TAG482_161161160_c	tag161_161_160_c --> tag482_c
TAG 50:2	TAG 16:0/18:1/16:1	EQformula_TAG502_160181161_c	tag160_181_161_c --> tag502_c
	TAG 16:1/18:1/16:1	EQformula_TAG502_161181160_c	tag161_181_160_c --> tag502_c
TAG 52:4	TAG 16:1/16:0/20:3	EQformula_TAG524_161160203_c	tag161_160_203_c --> tag524_c
	TAG 20:3/16:0/16:1	EQformula_TAG524_203160161_c	tag203_160_161_c --> tag524_c
TAG 52:5	TAG 16:0/16:0/20:5	EQformula_TAG525_160160205_c	tag160_160_205_c --> tag525_c
	TAG 16:0/16:1/20:4	EQformula_TAG525_160161204_c	tag160_161_204_c --> tag525_c
	TAG 16:1/16:1/20:3	EQformula_TAG525_161161203_c	tag161_161_203_c --> tag525_c
	TAG 20:3/16:1/16:1	EQformula_TAG525_203161161_c	tag203_161_161_c --> tag525_c
	TAG 20:4/16:1/16:0	EQformula_TAG525_204161160_c	tag204_161_160_c --> tag525_c
	TAG 20:5/16:0/16:0	EQformula_TAG525_205160160_c	tag205_160_160_c --> tag525_c
TAG 52:6	TAG 16:1/16:0/20:5	EQformula_TAG526_161160205_c	tag161_160_205_c --> tag526_c
	TAG 20:5/16:0/16:1	EQformula_TAG526_205160161_c	tag205_160_161_c --> tag526_c
TAG 54:5	TAG 20:3/16:1/18:1	EQformula_TAG545_203161181_c	tag203_161_181_c --> tag545_c
	TAG 20:4/16:0/18:1	EQformula_TAG545_204160181_c	tag204_160_181_c --> tag545_c

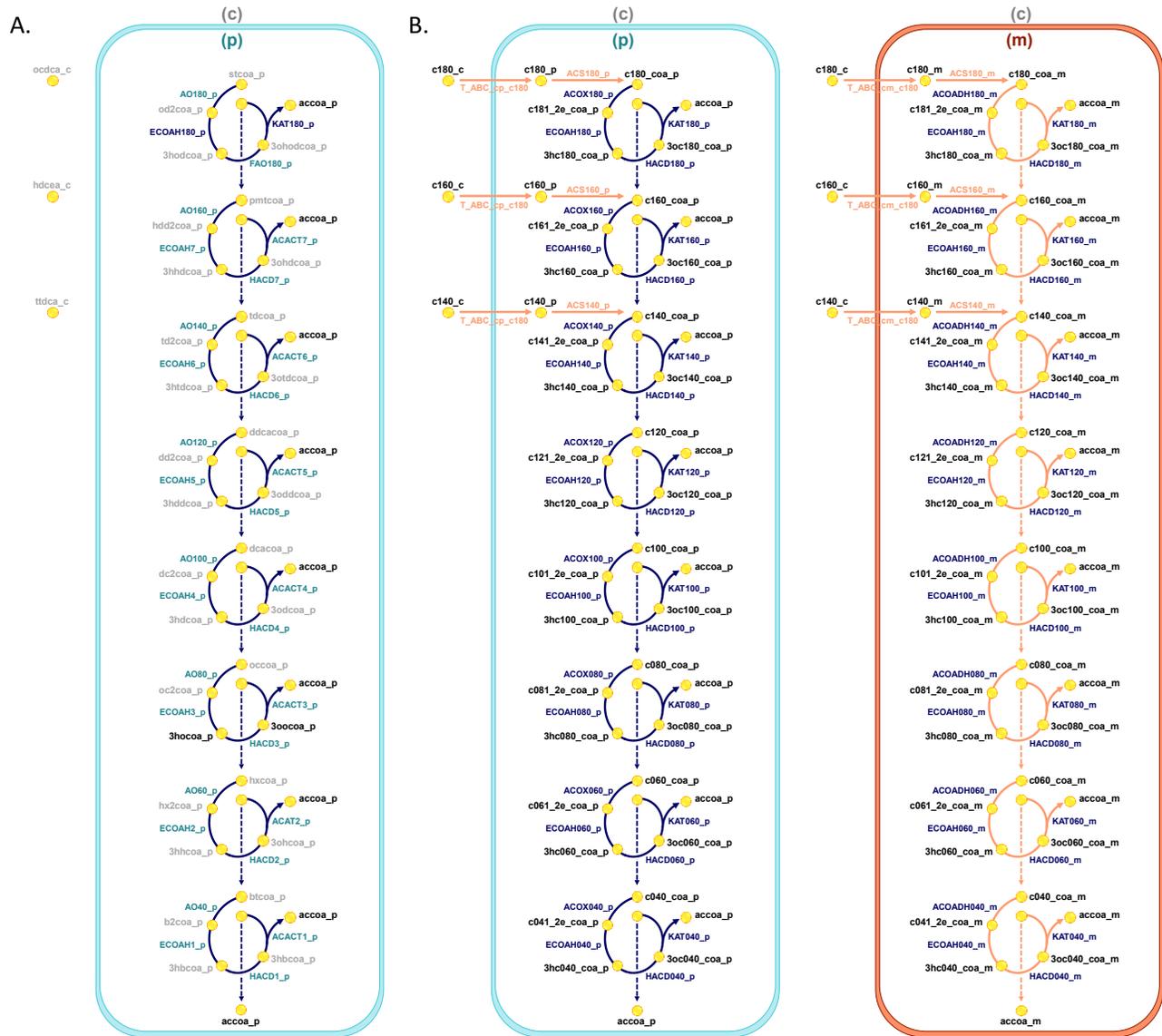
La composition de la biomasse de différentes souches de *M. gaditana* cultivées dans différentes conditions a été déterminée à partir de données disponibles au laboratoire et dans la littérature. Ces données ont permis la définition et l'ajout de sept fonctions de biomasse au modèle métabolique. Deux d'entre elles sont disponibles dans la version publique du modèle : iMgadi23.

## 2.2.3 Bilan des corrections réalisées au cours de la thèse

### 2.2.3.1 Correction du métabolisme lipidique de *M. gaditana*

Le modèle iMgadiLife comportait 605 réactions impliquées dans le métabolisme lipidique. Les voies de synthèse des acides gras et des glycérolipides ainsi que la  $\beta$ -oxydation étaient présentes dans le modèle. Ces voies métaboliques ont été intégralement corrigées. Cette correction a impliqué l'ajout de nombreuses réactions et de nombreux métabolites. La compartimentation de certaines réactions et de certains métabolites a été corrigée. De nombreuses réactions erronées ont été supprimées du modèle. L'ensemble des réactions et métabolites de ces voies a été renommé pour correspondre à la nouvelle nomenclature proposée dans ce manuscrit. Les voies de dégradation des glycérolipides et l'édition des acides gras ont été ajoutées au modèle. Finalement, le métabolisme lipidique est décrit par 1017 réactions dans la version publiée du modèle métabolique (iMgadi23).

L'impact de cette correction sur le contenu du modèle est illustré à partir de l'exemple de la  $\beta$ -oxydation. Dans iMgadiLife, la  $\beta$ -oxydation était localisée uniquement dans le peroxyosome. Seule la dégradation des acides gras saturés et monoinsaturés était décrite (78 réactions au total). De plus, ni l'import des acides gras libres du cytosol vers le peroxyosome ni leur activation par thioesterification au CoA par l'acyl-CoA synthase (ACS) n'étaient décrites. Ceci entraînait une non-fonctionnalité de la  $\beta$ -oxydation se traduisant par l'impossibilité d'avoir un flux lors des simulations. Par conséquent, des réactions de transport ainsi que des réactions ACS ont été incorporées au modèle. De plus, l'ensemble des réactions requises pour la dégradation des acides gras polyinsaturés (16:2, 18:2, 16:3, 18:3, 20:3, 20:4 et 20:5) a été ajouté. Par ailleurs, chez les espèces *Nannochloropsis* et *Microchloropsis*, la dégradation des acides gras via la  $\beta$ -oxydation peut se produire à la fois dans le peroxyosome et la mitochondrie (Poliner et al. 2015; Li-Beisson et al. 2019). La première étape de la  $\beta$ -oxydation diffère entre la mitochondrie et le peroxyosome. Dans le peroxyosome, la  $\beta$ -oxydation est initiée par la déshydrogénation de l'acyl-CoA en trans-2-enoyl-CoA via la réduction de dioxygène en peroxyde. Cette réaction est catalysée par l'acyl-CoA oxydase (ACOX). Dans la mitochondrie, la  $\beta$ -oxydation est initiée par l'acyl-CoA déshydrogénase (ACOADH) qui catalyse la déshydrogénation de l'acyl-CoA via la réduction du FAD en FADH<sub>2</sub> (Li-Beisson et al. 2019). Les réactions ACOADH ont été ajoutées à la mitochondrie, les autres réactions de la  $\beta$ -oxydation peroxyosomale ont été dupliquées dans la mitochondrie. Après correction, la  $\beta$ -oxydation est décrite par 143 réactions dans le peroxyosome, 141 dans la mitochondrie. Les identifiants des réactions et des métabolites ont également été corrigés afin de respecter la nouvelle nomenclature établie. La modélisation de la  $\beta$ -oxydation avant et après correction est représentée par la Figure 2.4.



**Figure 2.4. Modélisation de la  $\beta$ -oxydation des acides gras saturés dans iMgadit avant et après correction.** (A)  $\beta$ -oxydation des acides gras saturés dans iMgadit\_SilicoLife et (B) dans iMgadit23 après correction. Les identifiants des métabolites et des réactions de iMgadit\_SilicoLife ayant été renommés sont indiqués en gris et bleu clairs respectivement. La correction de la  $\beta$ -oxydation a nécessité l'ajout de nombreuses réactions (orange). Compartiments du modèle: cytosol (c), peroxysome (p) et matrice mitochondriale (m).

### 2.2.3.2 Bilan de l'ensemble des corrections

En supplément du métabolisme lipidique, d'autres voies ont été corrigées, notamment la photosynthèse, le cycle du glyoxylate, la photorespiration, le métabolisme de l'azote et les fonctions de biomasse.

La Table 2.9 présente un bilan des caractéristiques du modèle avant et après la thèse, ainsi que les corrections réalisées.

**Table 2.9. Bilan des corrections apportées à iMgadiT.**

		iMgadiT_SilicoLife	iMgadiT23
<b>Contenu du modèle</b>	Réactions	1890	2330
	Métabolites	1899	1977
	Gènes	864	889
<b>Modifications</b>	Réactions	ajoutées	1108
		supprimées	668
		renommées	302
	Métabolites	ajoutés	682
		supprimés	604
		renommés	224
Gènes	ajoutés		58
	supprimés		33

## 2.3 Résultats

**Article 1** : Deciphering lipid mutant phenotypes with a new genome-scale metabolic model of *Microchloropsis gaditana* encompassing detailed lipid metabolism.

Clémence Dupont-Thibert<sup>1,2</sup>, Sónia Carneiro<sup>3</sup>, Bruno Pereira<sup>3</sup>, Rafael Carreira<sup>3</sup>, Paulo Vilaça<sup>3</sup>, Sylvaine Roy<sup>1</sup>, Séverine Collin<sup>2</sup>, Eric Maréchal<sup>1</sup>, Elodie Billey<sup>2</sup>, Gilles Curien<sup>1</sup>, Maxime Durot<sup>2</sup>, Juliette Jouhet<sup>1</sup>

1 - Université Grenoble Alpes, CNRS, CEA, INRAE, Interdisciplinary Research Institute of Grenoble, IRIG-Laboratoire de Physiologie Cellulaire et Végétale, 17 Avenue des Martyrs, 38000 Grenoble, France

2 - TotalEnergies, OneTech, Centre de Recherche de Solaize CRES, Chemin du Canal, 69360 Solaize, France

3 - SilicoLife Lda. Rua do Canastreiro, 15 4715-387 Braga, Portugal

**N.B.1** : Cet article a été déposé sur bioRxiv (doi: <https://doi.org/10.1101/2023.12.06.570374>) et a été soumis pour publication.

**N.B.2** : La numérotation des figures et des tables est indépendante du manuscrit de thèse.

**N.B.3** : Les textes et figures supplémentaires de l'article sont présentés en annexe de ce manuscrit ([Annexe A](#)).

**N.B.4** : Le modèle et les scripts de cet article sont disponibles sur GitHub (<https://github.com/Total-RD/Mqaditana-GEM>)

# Deciphering lipid mutant phenotypes with a new genome-scale metabolic model of *Microchloropsis gaditana* encompassing detailed lipid metabolism.

Clémence Dupont-Thibert<sup>1,2</sup>, Sónia Carneiro<sup>3</sup>, Bruno Pereira<sup>3</sup>, Rafael Carreira<sup>3</sup>, Paulo Vilaça<sup>3</sup>, Sylvaine Roy<sup>1</sup>, Séverine Collin<sup>2</sup>, Eric Maréchal<sup>1</sup>, Elodie Billey<sup>2</sup>, Gilles Curien<sup>1</sup>, Maxime Durot<sup>2</sup>, Juliette Jouhet<sup>1</sup>

<sup>1</sup> Université Grenoble Alpes, CNRS, CEA, INRAE, Interdisciplinary Research Institute of Grenoble, IRIG-Laboratoire de Physiologie Cellulaire et Végétale, 17 Avenue des Martyrs, 38000 Grenoble, France

<sup>2</sup> TotalEnergies, OneTech, Centre de Recherche de Solaize CRES, Chemin du Canal, 69360 Solaize, France

<sup>3</sup> SilicoLife Lda. Rua do Canastreiro, 15 4715-387 Braga, Portugal

## Abstract

**Background:** The oleaginous microalga *Microchloropsis gaditana* (formerly *Nannochloropsis gaditana*) has gained large interest due to its potential to produce lipids for a wide range of biotechnological applications. To optimize *M. gaditana* growth conditions and develop new strains to enhance lipid synthesis and accumulation, a broad understanding of the organism metabolism is essential. Computational models such as genome-scale metabolic models constitute powerful tools for unravelling microorganism metabolism.

**Results:** In this work we present iMgadi23, a new genome-scale metabolic model for *M. gaditana*. Model covers 2330 reactions involving 1977 metabolites and associated with 889 genes. Pathways involved in membrane and storage glycerolipid biosynthesis and degradation have undergone thorough manual curation and have been comprehensively described based on current knowledge of *M. gaditana* lipid metabolism. Additionally, we developed a detailed 2D-pathway map of model content to provide a systems-level visualization of *M. gaditana* metabolism.

We demonstrated the predictive capabilities of iMgadi23, validating its ability to qualitatively and quantitatively capture *in vivo* growth phenotypes under diverse environmental and genetic conditions. Model was also able to capture the role of the Bubblegum acyl-CoA synthetase in remodeling *M. gaditana* lipid metabolism.

**Conclusion:** iMgadi23 and its 2D maps constitute valuable tools to increase understanding of *M. gaditana* metabolism and deciphering mutant phenotypes, specifically in the context of lipid metabolism. The model holds significant promise in predicting *M. gaditana* metabolic capabilities under varying genetic and environmental conditions, facilitating strain engineering, and optimizing cultivation processes for a broad range of industrial applications.

## 1. Introduction

Microalgae are photosynthetic microorganisms that have attracted significant attention from the biotechnology industry because of their ability to efficiently transform renewable resources (CO<sub>2</sub> and light) into biomass and other valuable molecules such as lipids. Given the wide range of lipids applications in diverse fields such as biofuels, green chemistry, health or food (Spolaore et al. 2006; Chu 2012; Rizwan et al. 2018), ongoing research involves the metabolic engineering of various microalgae strains and the optimization of culture processes to enhance lipid synthesis and accumulation (Sajjadi et al. 2018; Fu et al. 2019; Muñoz et al. 2021; Rock et al. 2021).

Microalgae constitute a polyphyletic group derived from several eukaryotic lineages (Keeling 2004; Petroutsos et al. 2014). Due to this complex evolutionary history, a wide variety of lipid compositions and lipid metabolism pathways have been reported within microalgal diversity (Kong et al. 2018; Li-Beisson et al. 2019). Strains of the genus *Nannochloropsis* (now divided into *Nannochloropsis* and *Microchloropsis* genera) belong to the Eustigmatophyceae class, and present a single secondary plastid with four limiting membranes resulting from a primary and a secondary endosymbioses (Petroutsos et al. 2014; Fawley et al. 2015). To date, five species of *Nannochloropsis*: *N. granulata*, *N. limnetica*, *N. oceanica*, *N. oculata*, and *N. australis* and two species of *Microchloropsis*: *M. gaditana* and *M. salina* have been identified (Fawley et al. 2015; Gong et al. 2020). *Nannochloropsis* and *Microchloropsis* accumulate high levels of triacylglycerol (TAG), making them promising feedstock for next generation biofuels. They also produce eicosapentaenoic acid (EPA, 20:5 20 carbon atoms : five unsaturations), a very long-chain polyunsaturated fatty acid (VLC-PUFA) which has been shown to provide many health benefits to humans (Ma et al. 2016; Al-Hoqani et al. 2017). The lipid molecules identified in both *Nannochloropsis* and *Microchloropsis* genera are similar (Mühlroth et al. 2017). Numerous gaps of knowledge regarding lipid metabolism pathways and their regulation need to be filled for these genera and more generally for microalga strains containing plastids with four limited membranes (Li-Beisson et al. 2019; Hoffmann and Shachar-Hill 2023). Filling these gaps is crucial to enhance lipid synthesis and accumulation.

Inclusion of accurate and species-specific description of lipid metabolic pathways into a genome-scale metabolic model (GEM) would constitute a valuable tool to further investigate lipid metabolism and to guide the production of these biotechnological targets. Derived from genome annotations and experimental data, GEMs are metabolic networks that collect all known metabolic information of a biological system, including the metabolites, reactions, genes, enzymes and associated gene-protein-

reaction (GPR) rules. Constraint-Based Reconstruction and Analysis (COBRA) methods perform systems-level analyses on GEMs to simulate how genetic and environmental factors affect phenotype, represented by flux distributions (Heirendt et al. 2019). For example, Flux Balance Analysis (FBA), is widely used to predict the phenotype that is expressed under certain culture conditions by calculating an optimal network state that optimizes an objective function (Orth et al. 2010). Biomass Objective Functions (BOFs), consuming all metabolites required to produce one gram of dry weight biomass (gDW), are commonly used as FBA objectives to simulate microorganisms growth. The resulting solution from an FBA is typically not unique, *i.e.*, multiple flux distributions can lead to the same growth rate. A common approach to decrease the possible solution space of FBA is to use the parsimonious version of FBA (pFBA). This approach finds a flux distribution with minimum absolute flux values among the alternative optima, assuming that the cell attempts to achieve its objective while allocating the minimum amount of resources (Lewis et al. 2010). In the last 25 years, COBRA methods have proven to be powerful tools for understanding and redesigning the metabolism of microbial strains including microalgae (Tibocha-Bonilla et al. 2018; Gu et al. 2019; Ye et al. 2022).

Given their potential for biotechnological applications, efforts have been done in the last few years to reconstruct GEMs of *Nannochloropsis* and *Microchloropsis* species. The first draft metabolic model of *Nannochloropsis sp.* was published by Pham in 2016. It was followed in 2017 by the publication of two fully reconstructed models of *M. gaditana* (iRJ1321) and *M. salina* (iNS937) (Shah et al. 2017; Loira et al. 2017). Finally, a model of *M. gaditana* automatically reconstructed using AuCoMe (Automated Comparison of Metabolism) was recently published (Belcour et al. 2023). Among these four GEMs, only iRJ1321 and iNS937 allow biomass production. In the Pham model, biomass reaction was added based on *Viridiplantae* data. However, not all precursors for this BOF can be synthesized by the model, making growth simulation impossible (Pham 2016). The Belcour model, originally reconstructed for conducting phylogenetic comparisons between species rather than predicting flux distributions, lacks any BOF. Even if iRJ1321 and iNS937 can be used to predict flux distributions, they are sub-optimal choices for the study of lipid metabolism due to the presence of non-functional or mis-compartmentalized pathways, and to missing components. In this study, we present iMgadit23, a new metabolic model for *M. gaditana* with a detailed description and modeling of lipid metabolism. Model was used to simulate growth behavior of the strain in different media and genetic conditions, and to decipher the phenotype of the *acsbg* mutant strain, which has an interesting lipid profile (Billey et al. 2021). To elucidate the practical applicability of iMgadit23, simulation results were validated with experimental data from literature.

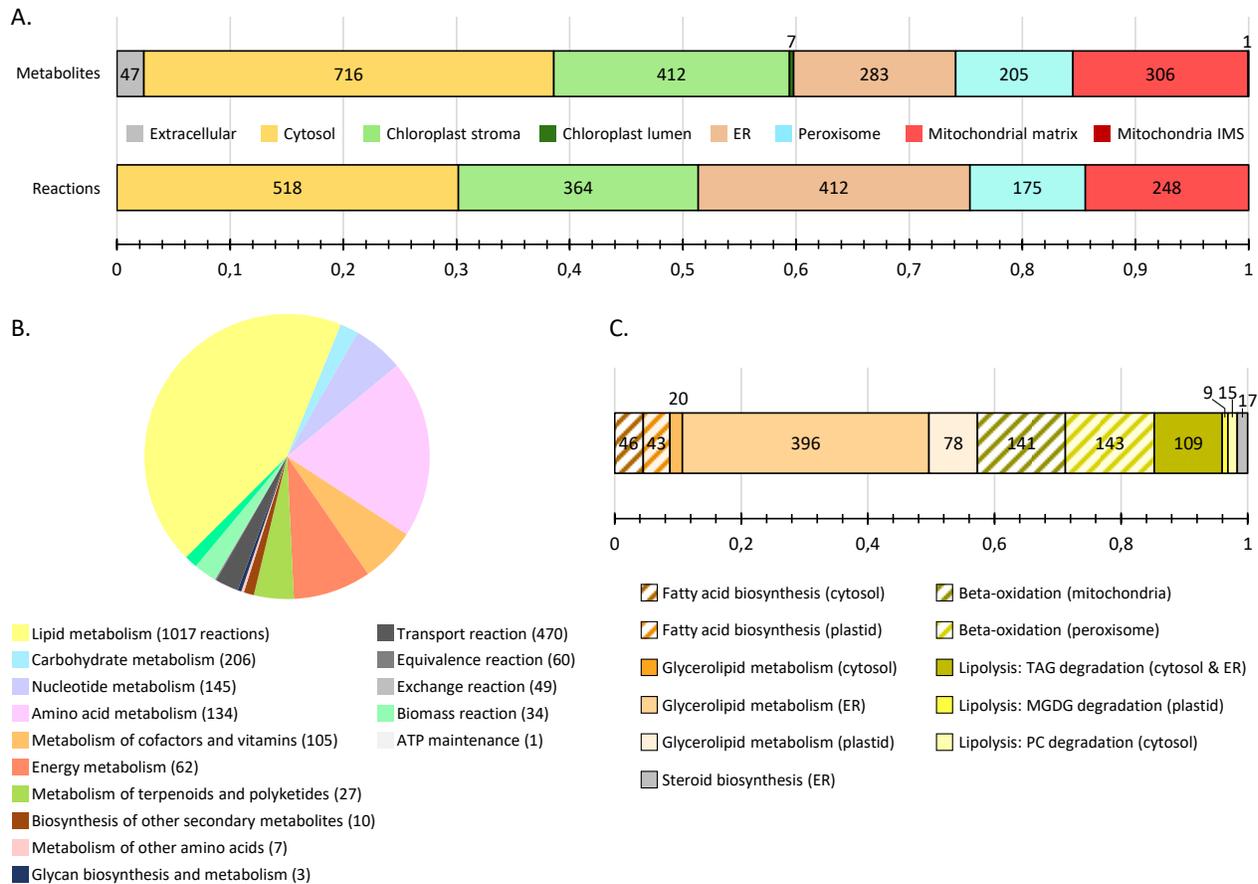
## 2. Results

### 2.1 *iMgadi23: a new genome-scale metabolic model for the oleaginous microalga M. gaditana.*

#### **General model properties**

GEM model was automatically reconstructed based on the genome sequence of *M. gaditana* B-31, before being manually curated and refined. The final curated model, *iMgadi23*, consists of 889 genes, 2330 reactions and 1977 metabolites (1155 chemical entities, *i.e.* when excluding metabolites duplicated in multiple compartments). The model includes eight compartments: extracellular (denoted by "e" in the model), cytosol (c), chloroplast stroma (h) and lumen (l), endoplasmic reticulum (ER, r in model), peroxisome (p), mitochondrial matrix (m) and intermembrane space (IMS, i in model) (Figure 1A). Model compartments were defined to encompass the intra-organelle space plus its membrane enclosure. For example, ER compartment covers both ER membrane and lumen. Most metabolites (36%) and reactions (30%) are located in the cytosol. While the chloroplast stroma contains 21% of all metabolites and reactions, the ER contains 24% of all reactions and 14% of all metabolites. All of the 2330 reactions are classified into 15 different subsystems mainly based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa and Goto 2000). These subsystems cover central and specialized metabolism (Figure 1B). Subsystems were divided into more detailed sub-subsystems available in *iMgadi23* notes.

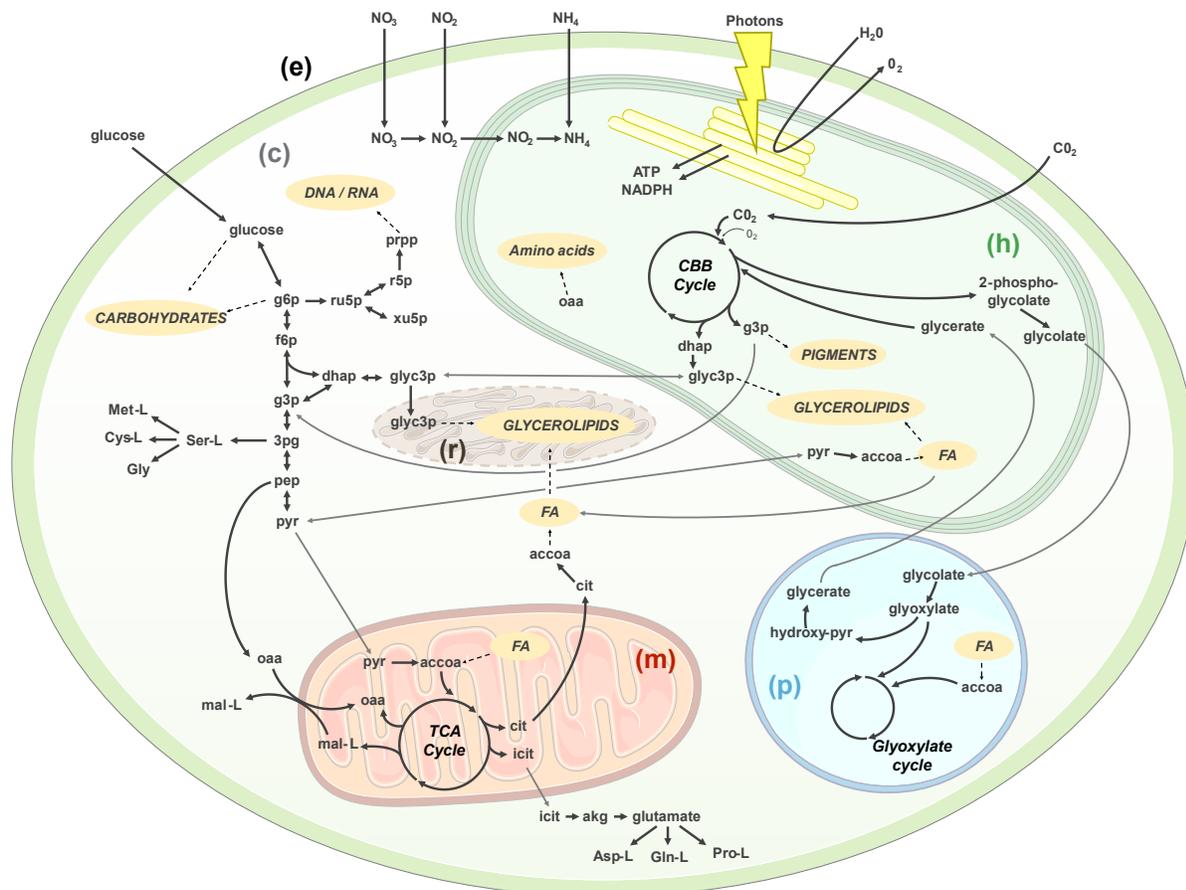
Given the promising lipid accumulation capacity of *M. gaditana* and its biotechnological potential, membrane and storage glycerolipid biosynthesis and degradation pathways were described in detail in *iMgadi23*. In total, lipid metabolism encompasses 1017 reactions (44% of all model reactions). Among them, 600 reactions, localized either in plastid, cytosol, or ER are involved in lipid biosynthesis (fatty acid biosynthesis, glycerolipid metabolism and steroid biosynthesis). The remaining 417 reactions are involved in fatty acids (FAs) degradation ( $\beta$ -oxidation in mitochondria and peroxisome) and glycerolipids degradation (TAG, MGDG and PC) in cytosol, plastid and ER (Figure 1C).



**Figure 1. General properties of iMgadi23 metabolic model.**

(A) Distribution of metabolites and reactions in the eight compartments: extracellular, cytosol, chloroplast stroma and lumen, endoplasmic reticulum (ER), peroxisome, mitochondrial matrix and intermembrane space (IMS). Transport reactions included in the model and describing metabolites movements between two compartments are not represented. (B) Distribution of reactions into 15 subsystems (KEGG metabolic pathways), number of reactions in each subsystem is indicated between parentheses. (C) Distribution of lipid metabolism reactions by compartmentalized pathways; compartments are indicated between parentheses.

Central metabolism as modeled in iMgadi23 is represented in (Figure 2). An exhaustive 2D-pathway map was drawn and is available as JSON file compatible with Escher (Supplemental File S1) (King et al. 2015). This map provides a systems-level representation of *M. gaditana* metabolism and constitutes a valuable tool to visualize and analyze biological data including reaction data (metabolic fluxes), metabolite data (metabolomics), and genomic data (transcriptomics data). An overview of this map content is presented in supplemental file (Supplemental Text Figure S1).



**Figure 2. Central metabolism as modeled in iMgadit23.**

The main model compartments are represented: cytosol (c), chloroplast stroma (h), extracellular medium (e), mitochondria matrix (m), peroxisome (p) and ER (r).

Metabolites: 3-phosphoglycerate (3pg), acetyl-CoA (accoa), L-aspartate (asp-L),  $\alpha$  ketoglutarate (akg), citrate (cit), L-cysteine (cys-L), fatty acids (FA), fructose-6-phosphate (f6p), glucose-6-phosphate (g6p), glyceraldehyde-3-phosphate (g3p), L-glutamine (gln-L) glycerol-3-phosphate (glyc3p), glycine (gly), dihydroxyacetone phosphate (dhap), isocitrate (icit), malate (mal-L), methionine (met-L), oxaloacetate (oaa), phosphoenolpyruvate (pep), phospho-ribose 1-diphosphate (prpp), proline (pro-L), pyruvate (pyr), ribose 5-phosphate (r5p), ribulose 5-phosphate (ru5p), serine (ser-L), xylulose 5-phosphate (xu5p). Pathways: Calvin–Benson–Bassham (CBB) cycle, tricarboxylic acid cycle (TCA) cycle.

### **Lipid metabolism**

Based on current knowledge of lipid metabolism of *M. gaditana* and other photosynthetic organisms, FAs and glycerolipids biosynthesis and degradation, including demonstrated and hypothetical pathways, were added to iMgadit23. These pathways are schematized in Figure 3. All reactions were duplicated to represent lipid metabolism diversity in terms of glycerolipid class and acyl chains profiles (Jouhet et al. 2017). To the authors knowledge, iMgadit23 represents the most up-to-date and accurate GEM of *Nannochloropsis* and *Microchloropsis* both in term of reaction content and localization.

- Lipid biosynthesis

FAs bound to acyl carrier proteins (ACPs) are *de novo* synthesized by the multi-subunit bacterial type II FA synthase (FASII) complex in chloroplast stroma (Figure 3A), generating essentially 16:0-ACP and 18:0-ACP which can undergo desaturation by a stromal acyl-ACP desaturase, leading to the production of 16:1-ACP and/or 18:1-ACP (Figure 3B). FA-ACPs can either be used to esterify glycerol-3-phosphate (glyc3p in iMgadi23) to produce precursors for four glycerolipids: phosphatidylglycerol (PG), sulfoquinovosyldiacylglycerol (SQDG), monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) (Figure 3C), or be converted into free FAs to be exported to the cytosol (Figure 3D) (Li-Beisson et al. 2019). FA elongation pathway by FASI complex was also added in iMgadi23 cytosol (Figure 3E). Indeed, transcription of the cytosolic type I FAS was reported in *M. gaditana* in stress conditions such as high light exposition, suggesting possibility of cytosolic *de novo* FA synthesis (Alboresi et al. 2016). Cytosolic free FA are then thioesterified to CoA by acyl-CoA synthetases (ACs), forming acyl-CoAs which are then used as initial precursors for all membrane phosphoglycerolipids in the ER *i.e.*, phosphatidylinositol (PI), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylcholine (PC); triacylglycerol (TAG) and the betaine lipids diacylglycerol-3-O-(N,N,N-trimethyl)-homoserine (DGTS) (Figure 3 F and G) (Li-Beisson et al. 2019). 16:0-CoA is also the initial substrate for a succession of elongations and desaturations forming at first 18:0-CoA and ending up with very-long-chain polyunsaturated FAs (VLC-PUFAs) up to 20:5 (EPA) (Figure 3H) (Vieler et al. 2012; Dolch et al. 2017). Re-importation of VLC-PUFAs into the plastid is required for the assembly of MGDG, DGDG, SQDG and PG. This process, known as the "omega pathway", remains to be fully characterized (Figure 3I) (Petroustos et al. 2014; Dolch et al. 2017).

- Lipolysis and acyl editing

Since growth and survival under fluctuating environmental conditions require permanent remodeling or turnover of membrane lipids as well as rapid mobilization of storage lipids, lipid degradation and acyl editing are important processes in microalgae. The process of acyl editing refers to a de-acylation of an acyl group from a given lipid and re-acylation on another lipid. It allows for fatty acid desaturation and modification prior to being removed and incorporated into other glycerolipids. Most of what is currently known about glycerolipid degradation and acyl editing has been determined in plants and much less is known in algae (Kong et al. 2018; Hoffmann and Shachar-Hill 2023). Few data are available for *M. gaditana* and other *Nannochloropsis* and *Microchloropsis* species. Based on current knowledge for *M. gaditana*, algae and plants, TAG, DAG and MAG lipolysis (Figure 3J) and the hydrolysis of PC by Phospholipase D,

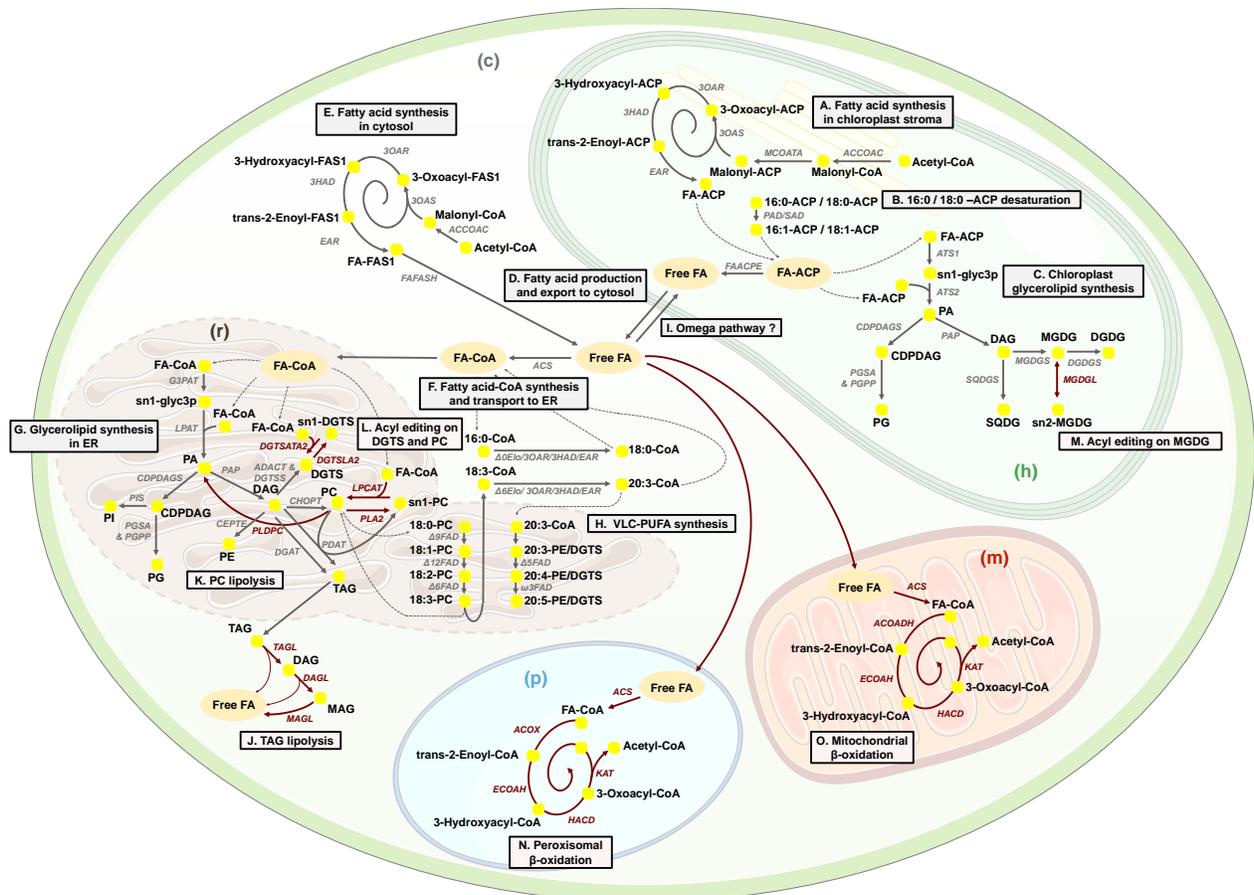
producing PA and a free choline group (Figure 3K) were added to our model. Acyl editing on PC, DGTS and MGDG substrates were also added (Figure 3 L and M). Reactions addition regarding glycerolipid degradation and acyl editing are further justified in Supplemental Material and Methods.

- β-oxidation

In *Nannochloropsis* and *Microchloropsis* species, FA degradation via β-oxidation may occur both in peroxisome and mitochondrion (Figure 3 N and O) (Poliner et al. 2015; Li-Beisson et al. 2019). It begins with the dehydrogenation of acyl-CoA to trans-2-enoyl-CoA either via the reduction of dioxygen (O<sub>2</sub>) to generate peroxide (ACOX in peroxisome) or via the reduction of FAD to FADH<sub>2</sub> (ACOADH in mitochondria). In both compartments, trans-2-enoyl-CoA is then hydrated, dehydrogenated, and cleaved into an acetyl-CoA and an acyl-CoA, two-carbons shorter compared to the initial substrate. For unsaturated FAs, two additional enzymes are required, cis-Δ<sup>3</sup>-Enoyl CoA isomerase that converts a cis-Δ<sup>3</sup> bond to a trans-Δ<sup>2</sup> bond (32ECOAI reactions) and the 2,4 Dienoyl CoA reductase that reduces 2,4-dienoyl-CoA into trans-Δ<sup>3</sup>-enoyl-CoA (24E2COAR reactions).

- Lipid metabolites and reactions identifiers

In iMgadi23, lipid nomenclature was chosen to facilitate understanding. Lipids are identified by their class acronym and detailed acyl chain composition following *sn-1*, *sn-2* (and *sn-3* for TAGs) order. Acyl chains are labeled according to the format "number of carbons : number of double bonds in acyl chain". For examples, hexadecanoic acid also known as palmitic acid or FA 16:0 is identified as "c160", and TAG 16:0/18:1/16:1 is identified as tag160\_181\_161. Due to the symmetry of the TAG chemical structure, the TAG molecule *sn-1 sn-2 sn-3* is identical to the TAG molecule *sn-3 sn-2 sn-1*, which is not the case in iMgadi23 where tag\_sn1\_sn2\_sn3 and tag\_sn3\_sn2\_sn1 constitute distinct metabolites due to their distinct synthesis pathways. To overcome this limitation, "equivalence" pseudo-reactions were added to the model, converting identical TAGs into a unique pseudo-TAG. For example, to represent equivalence between tag160\_181\_161 and tag161\_181\_160, two reactions converting respectively these TAGs into pseudo-TAG tag502 (50 carbons and 2 unsaturations) were added to the model (reactions were named EQformula\_TAG502\_160181161\_c and EQformula\_TAG502\_161181160\_c). In total iMgadi23 contains 24 "EQformula\_TAG" reactions. Same modeling methodology was applied for all lipid classes.



**Figure 3. Lipids metabolism as modeled in iMgadi23.**

Grey and red arrows represent respectively lipid synthesis (A to I); lipid degradation and acyl editing (L to O).

(A) Fatty acids (FAs) are initially synthesized in chloroplast stroma, starting with an ATP-dependent carboxylation of acetyl-CoA by an acetyl-CoA carboxylase (ACCOAC), producing malonyl-CoA which is subsequently converted into malonyl-acyl carrier protein (ACP) by the malonyl CoA-ACP malonyltransferase (MCOATA). Then, 3-oxoacyl-ACP synthase III catalyzes the condensation of malonyl-ACP with acetyl-ACP (3OAS). 3-oxoacyl-ACP is reduced by a 3-oxo-acyl reductase (3OAR) into 3-hydroxyacyl-ACP which serves as a substrate for an enoyl-dehydratase. 2,3-trans-enoyl-ACP produced is finally reduced by an enoyl-reductase (EAR) into an acyl-ACP containing two additional carbons compared to the initial substrate, up to a chain length of 14, 16 or 18 carbons. (B) Saturated FAs 16:0-ACP and 18:0-ACP can be mono-unsaturated by a stromal acyl-ACP desaturase forming 16:1-ACP and 18:1-ACP (PAD161\_h and SAD181\_h reactions). The FA-ACPs have two fates: they can (C) be acylated to glycerol-3-phosphate (glyc3p) by glycerol-3-phosphate acyl-ACP acyltransferases (ATS1 and ATS2) to produce precursors of four glycerolipids species: PG, SQDG, MGDG and DGDG (respectively via CDPDAG synthase (CDPDAGS), PG phosphate synthase (PGSA), PG phosphate phosphatase (PGPP), phosphatidic acid phosphatase (PAP), SQDG Synthase (SQDGS), MGDG Synthase (MGDGS) and DGDG Synthase (DGDGS)), or (D) be converted to free FAs by an acyl-ACP hydrolase (FAACPE) and exported to the cytosol. (E) FAs can also be de novo synthesized and elongated in cytosol. (F) Cytosolic free FAs can be thioesterified to CoA by acyl-CoA synthetase (ACS) and (G) be used in ER as initial precursors for synthesis of TAG via the Kennedy pathway by the glycerol 3-phosphate acyltransferase (G3PAT), the lysophosphatidic acid acyltransferase (LPAT), the PAP, the phospholipid:diacylglycerol acyltransferase (PDAT) and the diacylglycerol acyltransferase (DGAT). Betaine lipid DGTS is formed by a 2-step reaction of DAG with S-adenosylmethionine (ADACT and DGTSS). The membrane phosphoglycerolipids PI, PG, PE, and PC are respectively synthesized via CDPDAGS, PI Synthase (PIS), PGSA, PGPP, ethanolamine phosphotransferase (CEPT) and choline phosphotransferase (CHOPT). (H) Δ0-ELO, and Δ6-ELO elongases as well as Δ9, Δ12, Δ6, Δ5 and Δ15/ω3 fatty acyl desaturase (FAD) are required for very-long-chain polyunsaturated fatty acids (VLC-PUFAs) synthesis. (I) VLC-PUFAs 20:4 and 20:5 are then imported to the plastid via the "omega pathway". (J) TAGs can be degraded via TAG-, DAG-, and MAG- lipases. (K) PC lipolysis can occur via Phospholipase D (PLDPC). (L) FAs can be edited on PC by Phospholipase A2 (PLA2) and lyso-PC acyl transferase (LPCAT) and DGTS can be edited by DGTS lipase A2 (DGTSLA2) and DGTS acyl transferase A2 (DGTSATA2). (M) Acyl-editing on MGDG (MGDG lipase (MGDGL)). Free FAs can be degraded via β-oxidation both in peroxisome (N) and mitochondria (O). Model compartments are denoted (c) cytosol, (h) chloroplast stroma, (m) mitochondrial matrix, (p) peroxisome and (r) ER.

### **Biomass Objective Function**

In GEMs, a BOF consumes resources essential for the growth of a given microorganism, weighted to represent mole fraction of each component in one gram of dry weight. In iMgadi23 BOF contains stoichiometric coefficients for 121 metabolites, including 78 lipids. These metabolites are consumed by six pseudo-reactions representing the different macromolecules (PROTEINS, DNA, RNA, LIPIDS, CARBOHYDRATES and PIGMENTS). Ten pseudo-reactions representing glycerolipid classes (TAG, DAG, PC, PE, PI, PG, DGTS, DQDG, MGDG and DGDG) were also added. Detailed lipidomic data, in terms of glycerolipid profiles and FA profiles in glycerolipid classes, of *M. gaditana* WT grown under photoautotrophy were used to calculate stoichiometric coefficients of lipids (Billey et al. 2021). *In fine*, biomass composition in iMgadi23:  $\text{CH}_{1.8}\text{O}_{0.49}\text{N}_{0.11}\text{P}_{0.005}\text{S}_{0.003}\text{Mg}_{0.001}$ , is consistent with the molecular formula of microalgae cell reported in literature:  $\text{CH}_{1.83}\text{O}_{0.48}\text{N}_{0.11}\text{P}_{0.01}$  (Ma et al. 2022). For detailed composition of BOF see Supplemental Material and Methods and Supplemental Files S2 and S3.

In addition to biomass composition, ATP requirements for both growth-associated maintenance (GAM, included in biomass pseudo-reactions) and non-growth-associated maintenance (NGAM, independent reaction) were respectively set to 29.89 and 2.2  $\text{mmol}_{\text{ATP.g}_{\text{DW}}^{-1}.\text{h}^{-1}}$  based on literature data for *Chlamydomonas reinhardtii* and *Nannochloropsis sp.* (Boyle and Morgan 2009; Zhang et al. 2014; Shah et al. 2017).

### *2.2 In silico predictions and validation*

#### **Validation of iMgadi23 quality with MEMOTE**

Overall quality of iMgadi23 was evaluated using MEMOTE, a test suite for standardized quality assessment of GEMs (Lieven et al. 2020). The general properties of iMgadi23 were compared to the other GEMs of *Nannochloropsis* and *Microchloropsis*. General model features are summarized in Table 1 and complete MEMOTE reports are available in Supplemental Files S4-S7. iMgadi23 constitutes a high-quality model whose features include a multi-compartmentalized network encompassing most relevant microalgae compartments. Excluding reactions converting photons into chemical energy, all model reactions are charge and mass balanced. Only a low fraction of model reactions (10%, 99 reactions) can carry unlimited flux under model default medium (see Material and Method section). 27.3% of model reactions (636 reactions) lack a Gene Protein Reaction rule (GPR), most of them (319) being transport reactions or lipid metabolism (282) due to gaps of knowledge of genes encoding lipid metabolism in *M. gaditana*. Regarding GPRs, two pseudo-genes were introduced in iMgadi23: (1) "Spontaneous" which was associated to spontaneous reactions such as  $\text{CO}_2$ ,  $\text{O}_2$  or photon diffusion through membranes and (2)

"NO\_GENE" associated to artificial reactions linked to modeling constraints (Biomass, NGAM, Exchange and Equivalence reactions). "OR" operator was used in GPRs to associate multiple genes to a common reaction. A total of 220 model metabolites are orphan (metabolites that are only consumed but not produced by reactions in the model) or dead-end (metabolites that can only be produced but not consumed by reactions in the model), resulting in a relatively high fraction of blocked reactions (19%, 443 reactions). Blocked reactions are generally caused by network gaps resulting from lack of metabolism knowledge. Nevertheless, in iMgadi23, blocked reactions do not prevent biomass production and are significantly fewer than in iNS937 and iRJ1321 where respectively 814 reactions (34.71%) and 930 reactions (48.49%) are blocked.

Detailed MEMOTE results also highlight high quality of lipid modeling in iMgadi23: among 1017 reactions describing lipid metabolism, only 61 (6%) are blocked. Moreover, some of these reactions are blocked because all lipids are not consumed in biomass reaction. When adding an artificial BOF reaction consuming all glycerolipids synthesized in model, number of lipid reactions blocked drops to 38. On the contrary, among 536 reactions representing lipid metabolism in iRJ321, 254 (47%) are blocked. Likewise, in iNS037, lipid metabolism is described by 827 reactions, 233 of which (28%) being blocked.

Finally, a "Total model score" of 67% was calculated by MEMOTE for iMgadi23. Both model consistency and annotation are considered to calculate this score. Consistency score of iMgadi23 is 55.7% because the model failed stoichiometry consistency test. All the different published models tested with MEMOTE in this study are also stoichiometrically inconsistent due to light energy conversion into chemical energy, breaking mass conservation law in models. When removing both photosystem reactions (PSII\_h and PSI\_h) from iMgadi23, the model becomes stoichiometrically consistent, leading to a consistency score of 98.5% and a total score of 79.6%, within the standard GEMs quality. Indeed, in 2020, a quality control assessment of the 108 SBML-GEMs of the BiGG database using MEMOTE revealed an average consistency and total scores of 92.18% and 72.54% (Norsigian et al. 2020).

**Table 1. Properties of iMgadi23 and other *Nannochloropsis*, *Microchloropsis* GEMs.**

<sup>1</sup> Expected score 100%. <sup>2</sup> Expected score 0%

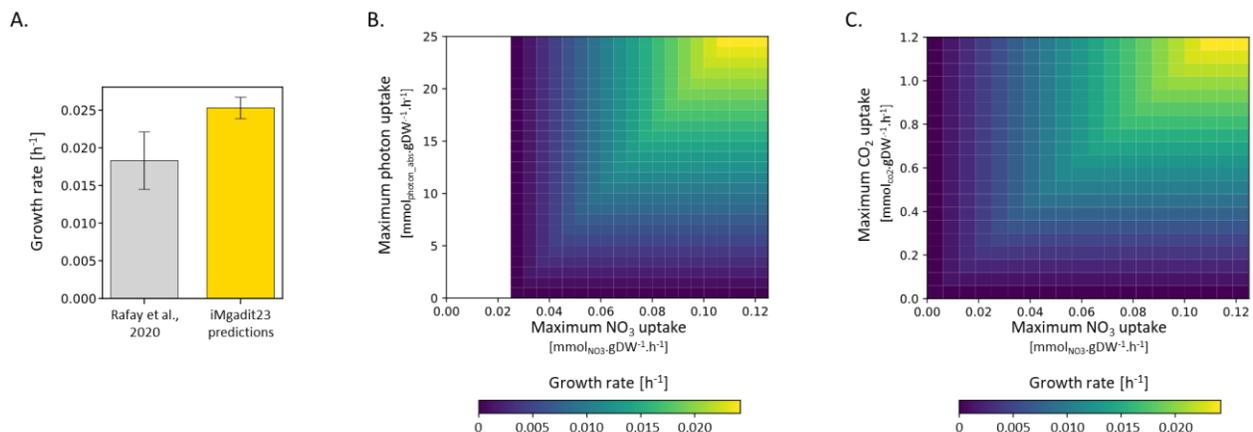
Model	Pham model (Pham 2016)	iNS937 (Loira et al. 2017)	iRJ1321 (Shah et al. 2017)	<i>Nannochloropsis_gaditana</i> (Belcour et al. 2023)	iMgadi23 (This study)
Strain	<i>N. sp</i>	<i>M. salina</i>	<i>M. gaditana</i>	<i>M. gaditana</i>	<i>M. gaditana</i>
Reactions	987	2345	1918	2729	2330
Metabolites	1024	1985	1862	3076	1977
Genes	383	934	1321	3474	889
Compartments	10	10	4	2	8
extracellular	✓	✓	✓	✓	✓
cytosol	✓	✓	✓	✓	✓
mitochondrion	✓	✓	✓		✓ 2 compartments: matrix & IMS
chloroplast	✓	2 compartments: stroma & lumen	✓		✓ 2 compartments: stroma & lumen
peroxisome	✓	✓			✓
glyoxysome					
endoplasmic reticulum	✓	✓			✓
Golgi apparatus	✓	✓			
lysosome		✓			
nucleus	✓	✓			
vacuole	✓				
cell wall	✓				
<b>MEMOTE results (%)</b>					
Total score	/	18	25.3	17.7	67
Mass Balance <sup>1</sup>	/	98.9	99.7	0	99.9
Charge Balance <sup>1</sup>	/	0	90.9	100	99.9
Metabolic connectivity <sup>1</sup>	/	100	100	99.7	100
Unbounded Flux in Default Medium <sup>1</sup>	/	58.9	97.9	58.8	89.9
Reactions without GPR <sup>2</sup>	/	22	10.4	3.7	27.3
Universally Blocked Reactions <sup>2</sup>	/	34.7	48.5	80.4	19
Orphan Metabolites <sup>2</sup>	/	9	11.6	26.1	5.1
Dead End Metabolites <sup>2</sup>	/	11.3	14.8	27.2	6.1
Stoichiometrically <sup>2</sup> Balanced Cycles <sup>2</sup>	/	29.1	1.6	8.1	4.6

### Quantitative validation

Three experimental maximum nitrate ( $\text{NO}_3$ ) uptake rates measured in *M. gaditana* cultivated under photoautotrophy (Rafay et al. 2020) were successively used to constrain  $\text{NO}_3$  uptake in iMgadi23 and pFBA was used to predict maximal growth rate. As presented in Figure 4A, model predicts an average maximum growth rate of  $0.0253 \pm 0.0014 \text{ h}^{-1}$  which is slightly higher than maximum growth rate reported experimentally:  $0.0183 \pm 0.0038 \text{ h}^{-1}$  (Rafay et al. 2020). Nevertheless, iMgadi23 predictions are in agreement with the maximal exponential growth rate values reported in literature for *M. gaditana* under photoautotrophic conditions with  $\text{NO}_3$  as nitrogen source which range from 0.01 to  $0.044 \text{ h}^{-1}$  with an average of  $0.027 \text{ h}^{-1}$  (Gentile and Blanch 2001; Ren and Ogden 2014; Kim et al. 2014; Kareya et al. 2020). iMgadi23 predictions are also consistent with a recently developed kinetic growth model of *M. gaditana*, predicting a maximal growth rate of  $0.0256 \text{ h}^{-1}$  (Lima et al. 2022).

pFBA simulation results were used to calculate the ratio of oxygen evolved per carbon dioxide fixed. Ratio of  $1.37 \text{ mol}_{\text{O}_2} \cdot \text{mol}_{\text{CO}_2}^{-1}$  predicted by iMgadi23 model is consistent with the experimental data reported for *Nannochloropsis sp.* under photoautotrophic conditions with  $\text{NO}_3$  as nitrogen source : ranging from 1.17 to  $1.95 \text{ mol}_{\text{O}_2} \cdot \text{mol}_{\text{CO}_2}^{-1}$  (Zhang et al. 2014) .

In microalgae, growth depends on numerous environmental parameters, notably availability of carbon, nitrogen sources and light intensity. Phenotypic phase plane analyses using pFBA approach were performed to provide an overview of how predicted optimal growth rate is impacted by changes in  $\text{CO}_2$ ,  $\text{NO}_3$  and photon availability (Figure 4.B and C).



**Figure 4. Predicted optimal growth rates of *M. gaditana* as a function of nitrogen and light.**

(A)  $\text{NO}_3$  uptake rates reported by Rafay et al. in 2020 were used to constrain iMgadi23. Maximum growth rates were predicted by pFBA. Phenotypic phase plane results providing a global view of how change in (B)  $\text{NO}_3$  and photon and (C)  $\text{NO}_3$  and  $\text{CO}_2$  impact optimal predicted growth (color is used as third dimension to represent growth rate).

### **Qualitative validation**

*M. gaditana* as well as other *Nannochloropsis* and *Microchloropsis* species are able to grow in various conditions including: photoautotrophy with different nitrogen sources, heterotrophy and mixotrophy on different carbon sources (Xu et al. 2004; Kilian et al. 2011; Gim et al. 2016; Loira et al. 2017). iMgadi23 was tested for its ability to capture *M. gaditana* growth in 12 different conditions: heterotrophic growth on seven different carbon sources, mixotrophic growth on glucose and photoautotrophic growth on four nitrogen sources. Except for heterotrophic growth on ethanol, which is not predicted by iMgadi23 due to ethanol metabolism absence in model, all predictions meet experimental phenotypes (Table 2). In addition, growth phenotype of three knockout (KO) mutant strains was predicted in different media conditions. Model predictions are consistent with *in vivo* phenotypes for nitrate reductase KO (Naga\_100699g1) and nitrite reductase KO (Naga\_100852g1) tested for growth on different nitrogen sources and alternative oxidase KO (Naga\_100568g3), (Kilian et al. 2011; Bo et al. 2021). Only the Bubblegum ACS *MgACSBG* (Naga\_100014g59) is predicted as non-essential by iMgadi23, whereas Billey et al., 2021 reported that this gene might be essential. *M. gaditana* genome encodes six annotated ACSs (Naga\_100014g59, Naga\_100012g66, Naga\_101051g1, Naga\_100047g8, Naga\_100649g1 and Naga\_100035g43) (Billey et al. 2021). Since the specificity of ACS enzymes for fatty acids remains to be characterized, all ACSs genes were associated to cytosolic ACS reactions in iMgadi23 with the logical connector "OR". Thus, inactivation of only one of ACS coding gene in iMgadi23 does not impact any ACS reaction functionality. Further analysis of ACS reactions is presented in "Application *in vivo*" section. Taking everything into account, iMgadi23 demonstrates good capacity for growth/no-growth predictions, with a sensitivity of 0.94, a specificity of 0.75 and an accuracy of 0.9. Good capacity of iMgadi23 to predict growth on different cultivation conditions both quantitatively and qualitatively highlight its usefulness to evaluate the biological capabilities of *M. gaditana* on different experimental conditions.

**Table 2. iMgadi23 growth predictions in different conditions.**

Growth predictions by iMgadi23 are compared to in vivo results (+ and - corresponding to growth and no growth, respectively) for *Nannochloropsis* species (*N. sp.*, *N. oculata*) and *Microchloropsis* species (*M. gaditana*). TFPN column corresponds to results of the confusion matrix and indicates model performance (see Material and Methods).

Condition	Medium	Gene KO	In vivo	In silico	TFPN	Species	References
Heterotrophy	Glucose		+	+	TP	<i>N. sp.</i>	(Fang et al. 2004)
	Ethanol		+	-	FN	<i>N. sp.</i>	(Fang et al. 2004)
	Xylose		+	+	TP	<i>N. oculata</i>	(Gim et al. 2016)
	Rhamnose		+	+	TP	<i>N. oculata</i>	(Gim et al. 2016)
	Fructose		+	+	TP	<i>N. oculata</i>	(Gim et al. 2016)
	Sucrose		+	+	TP	<i>N. oculata</i>	(Gim et al. 2016)
	Galactose		+	+	TP	<i>N. oculata</i>	(Gim et al. 2016)
Mixotrophy	Glucose		+	+	TP	<i>N. sp.</i>	(Fang et al. 2004)
Photoautotrophy	Nitrite		+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Nitrate		+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Ammonium		+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Urea levels		+	+	TP	<i>M. salina</i>	(Loira et al. 2017)
Knock-Out (Photoautotrophy)	Nitrite	Naga_100699g1 (Nitrate reductase)	+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Nitrate	Naga_100699g1 (Nitrate reductase)	-	-	TN	<i>N. sp.</i>	(Kilian et al. 2011)
	Ammonium	Naga_100699g1 (Nitrate reductase)	+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Nitrite	Naga_100852g1 (Nitrite reductase)	-	-	TN	<i>N. sp.</i>	(Kilian et al. 2011)
	Nitrate	Naga_100852g1 (Nitrite reductase)	-	-	TN	<i>N. sp.</i>	(Kilian et al. 2011)
	Ammonium	Naga_100852g1 (Nitrite reductase)	+	+	TP	<i>N. sp.</i>	(Kilian et al. 2011)
	Nitrate	Naga_100568g3 (Alternative oxidase)	+	+	TP	<i>M. gaditana</i>	(Bo et al. 2021)
	Nitrate	Naga_100014g59 ( <i>MgACSBG</i> )	-	+	FP	<i>M. gaditana</i>	(Billey et al. 2021)

### 2.3 Application in vivo: characterization of *acsbg* mutant

One of the strengths of GEMs is to be a support for researchers working on metabolism or molecular engineering. For this reason, a mutant strain described in the literature was chosen to evaluate iMgadi23. The *MgACSBG* gene was studied by Billey and co-authors (Billey et al. 2021). In their work, they report that despite numerous attempts, they were unable to obtain a *MgACSBG* knockout, suggesting that reactions catalyzed by the *MgACSBG* protein may be vital for *M. gaditana*. As previously mentioned, iMgadi23 failed to predict *MgACSBG* essentiality due to lack of specificity information regarding ACS enzymes.

Billey and co-authors generated point mutations in *MgACSBG* gene and characterized its impact on lipid profile. *MgACSBG#31* mutant from Billey et al. 2021 (referred to as "*acsbg* mutant" in our study) harbors a local mutation in *MgACSBG* leading to three amino acid changes in the *MgACSBG* protein. Structural

modeling of the mutant protein suggests that accessibility of FAs to the catalytic site is affected. Glycerolipid profile is highly impacted by the mutation. Mutant strain notably showed (1) an increase in DGTS, (2) a very large increase in TAGs, (3) an overall decrease of 16:1 in glycerolipids, (4) a specific increase of unsaturated 18-carbon acyls in PC and (5) a decrease of 20-carbon acyls in DGTS. Altogether, results presented by Billey et al. 2021 suggest a role of MgACSBG in the production of 16:1-CoA and 18:3-CoA. Detailed analysis of lipid composition being provided in the publication, this gene makes a good candidate for iMgadi23 evaluation.

### **ACS reactions deletion**

Essentiality analysis of each cytosolic ACS reaction of iMgadi23 was conducted to determine which reactions, when deleted *in silico*, significantly reduce or eliminate biomass production. Most reactions (ACS140\_c, ACS160\_c, ACS161\_c, ACS181\_c, ACS183\_c, ACS204\_c and ACS205\_c) are predicted as essential by iMgadi23. Deletion of ACS120\_c, ACS180\_c, ACS182\_c and ACS203\_c do not affect predicted growth. These results suggest that MgACSBG might be specific for one or more of the FAs: 14:0, 16:0, 16:1, 18:1, 18:3, 20:4 or 20:5.

### **Differential flux analyses**

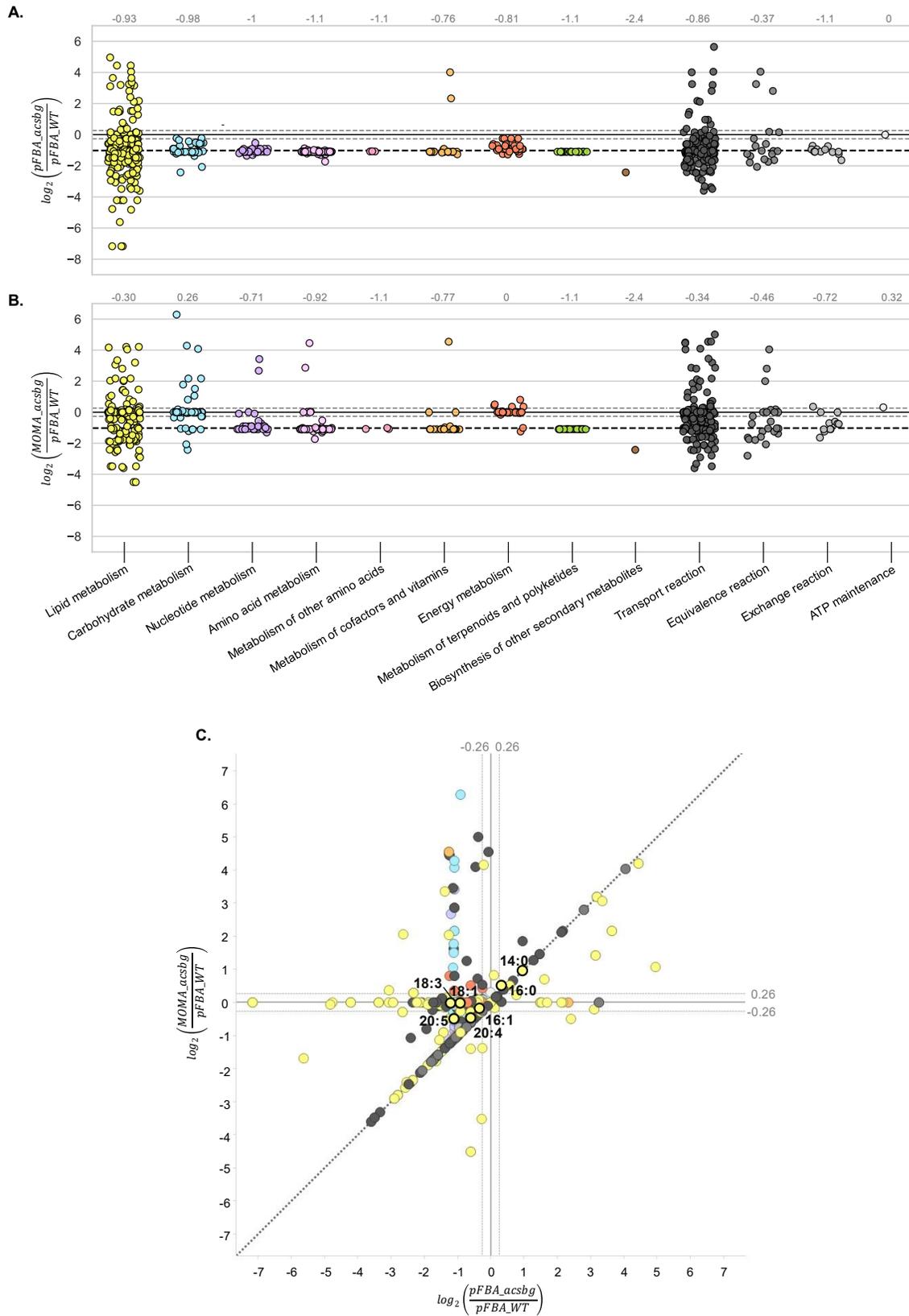
iMgadi23 was used to analyze phenotype of *acsbg* mutant by predicting how its lipid profile impact flux distribution. To do so, fully detailed lipidomic data from *acsbg* mutant including (1) total lipid content per gram of dry weight, (2) molar fraction of every glycerolipid classes and (3) molar fraction of the acyl profile of each class, were used to create a new BOF in iMgadi23.

At first, flux distributions were predicted by pFBA for WT and *acsbg* strains. For each simulation, appropriate BOF was constrained to experimental growth rate ( $0.030 \text{ h}^{-1}$  for WT,  $0.015 \text{ h}^{-1}$  for *acsbg*) and pFBA was conducted with minimization of photon uptake as objective function. To facilitate reading, these flux distributions will be respectively termed pFBA\_WT and pFBA\_*acsbg*. For each model reaction,  $\log_2(\text{pFBA\_acsbg}/\text{pFBA\_WT})$  was calculated (Figure 5.A). A positive  $\log_2$  fold change ( $\log_2(\text{FC})$ ) indicates that reaction flux is increased in *acsbg* compared to WT. On the contrary, a negative  $\log_2(\text{FC})$  indicates a decrease of the reaction flux in the mutant. Reactions with  $-0.26 \leq \log_2(\text{FC}) \leq 0.26$ , corresponding to a flux change lower than 20%, were considered as non-significantly impacted. See Material and Methods section for more details. Most reactions are downregulated in pFBA\_*acsbg* compared to pFBA\_WT (Figure 5A). In total, among 770 reactions predicted to have a flux in pFBA\_WT or pFBA\_*acsbg*, 606 have a significantly lower flux in mutant compared to WT. This downregulation tendency can be explained by experimental

strains growth rates ( $\mu_{WT} = 0.030 \text{ h}^{-1}$ ,  $\mu_{acsbg} = 0.015 \text{ h}^{-1}$ ) being applied as simulation constraints. Thus, many reactions are predicted to have a  $\log_2(\text{pFBA}_{acsbg}/\text{pFBA}_{WT})$  close to  $\log_2(\mu_{acsbg}/\mu_{WT}) = -1$ . Despite these simulation constraints, 50 reactions, mostly involved in lipid metabolism and transport, are predicted to be upregulated in  $\text{pFBA}_{acsbg}$  compared to  $\text{pFBA}_{WT}$ . Lipid metabolism and transport are the most impacted subsystems in terms of number of up and downregulated reactions but also in terms of  $\log_2(\text{FC})$  ranges.

Then, we applied minimization of metabolic adjustments (MOMA), another popular COBRA method developed to find a flux distribution for a perturbed condition by minimizing the Euclidean distance from a reference flux distribution (Segrè et al. 2002). MOMA is considered as more suitable to predict flux distributions in mutant organisms. *acsbg* BOF was constrained to its experimental value, and fluxes were predicted with MOMA using  $\text{pFBA}_{WT}$  as reference distribution. Predicted fluxes will be referred to as  $\text{MOMA}_{acsbg}$ .  $\log_2(\text{MOMA}_{acsbg}/\text{pFBA}_{WT})$  was calculated for each iMgadi23 reaction (Figure 5.B). Only flux change higher than 20% were considered significant. Among the 1009 reactions predicted to have a flux in  $\text{pFBA}_{WT}$  or  $\text{MOMA}_{acsbg}$ , 373 (including 79 reactions of lipid metabolism) are significantly downregulated, 80 (including 26 reactions of lipid metabolism) are significantly upregulated and 253 reactions have similar fluxes for both strains.

$\text{pFBA}$  and MOMA predictions for *acsbg* were then compared to identify reactions, robustly predicted to be up- or down- regulated in mutant by both approaches, with a focus on cytosolic ACS reactions (Figure 5.C). 31 reactions are upregulated in  $\text{pFBA}_{acsbg}$  and  $\text{MOMA}_{acsbg}$  compared to  $\text{pFBA}_{WT}$ , with 18 of these reactions involved in lipid metabolism (Figure 5.C). Interestingly, ACS140\_c and ACS160\_c are part of these significantly upregulated reactions. On the other hand, 366 reactions, including 74 reactions of lipid metabolism notably ACS204\_c and ACS205\_c, are downregulated both in  $\text{pFBA}_{acsbg}$  and  $\text{MOMA}_{acsbg}$ . ACS161\_c is significantly downregulated in  $\text{pFBA}_{acsbg}$  and non-significantly downregulated in  $\text{MOMA}_{acsbg}$  with  $\log_2(\text{MOMA}_{acsbg}/\text{pFBA}_{WT}) = -0.17$  (corresponding to flux decrease of 12%). For ACS181\_c and ACS183\_c, fluxes are significantly decreased in  $\text{pFBA}_{acsbg}$ , but  $\log_2(\text{MOMA}_{acsbg}/\text{pFBA}_{WT}) = 0$ . For both WT and *acsbg*, a null flux is predicted for ACS120\_c, ACS180\_c, ACS182\_c and ACS203\_c by  $\text{pFBA}$  and MOMA. All reactions significantly predicted as up regulated or downregulated in *acsbg* by both methods are detailed in Supplemental File S8.  $\text{pFBA}_{WT}$ ,  $\text{pFBA}_{acsbg}$ ,  $\text{MOMA}_{acsbg}$  and  $\log_2(\text{FC})$  for each model reaction are detailed in Supplemental File S9. The lipid metabolism results are also detailed in Figure 6.



**Figure 5. Results of the differential flux analyses between WT and *acsbg* strains. (Legend: see next page)**

Comparison of flux values for both strains for each model reaction grouped by subsystem predicted with (A) pFBA approach ( $\log_2(\text{pFBA\_acsbG}/\text{pFBA\_WT})$ ) and (B) MOMA approach ( $\log_2(\text{MOMA\_acsbG}/\text{pFBA\_WT})$ ). A positive  $\log_2(\text{FC})$  indicates an upregulated reaction whereas a negative value indicates a downregulated reaction. Average  $\log_2(\text{FC})$  of the subsystem is indicated in grey. Dashed black line indicates  $\log_2(\mu_{\text{acsbG}}/\mu_{\text{WT}}) = -1$ , dashed grey lines indicate significant thresholds of -0.26 and 0.26 corresponding to up/downregulation of 20% in acsbG compared to WT. (C) Comparison of  $\log_2(\text{FC})$  between both pFBA and MOMA approaches. Diagonal line represents identical predictions with both methods. Reactions predicted to be upregulated by both methods are in the upper right corner, reactions predicted to be downregulated by both methods are in the bottom left corner. Reactions predicted to be downregulated in pFBA\_acsbG but upregulated in MOMA\_acsbG are in the upper left corner. Reactions predicted to be upregulated in pFBA\_acsbG but downregulated in MOMA\_acsbG are in the bottom right corner. Dashed grey lines indicate significant threshold of -0.26 and 0.26. Cytosolic ACS reactions are highlighted in black.

Constraint-based modeling leaves open the possibility of alternate optimal solutions which mean that the same objective value can be achieved by a diverse set of flux distributions. Flux variability analysis (FVA) determines the possible range of flux values (*i.e.* minimum and maximum values for each reaction) which is allowable with the given objective value (Gudmundsson and Thiele 2010). To give an overview of flux variability notably within ACS reactions, we performed flux variability analyses (FVA) to determine the range of possible reaction fluxes that still satisfy the original optimal objective value predicted by pFBA (Table 3). FVA approach, with 100% optimality constraint, was applied for WT and *acsbG* strains by constraining appropriate BOF to experimental growth rate, with minimization of photon uptake as objective function. For each reaction, the fractional overlap in flux range between WT and *acsbG* was calculated as defined in Equation 5 (See Material and Methods). Interestingly, no overlap is predicted for ACS140\_c and ACS160\_c confirming their upregulation in *acsbG*. No overlap is also predicted for ACS161\_c, ACS181\_c, ACS183\_c and ACS205\_c reinforcing pFBA\_acsbG downregulation hypothesis. ACS204\_c has a fractional overlap of 0.34, suggesting that its predicted upregulation in pFBA\_acsbG can be due to variability of lipid metabolism.

To sum up regarding ACS reactions, all three approaches pFBA, MOMA and FVA agree in predicting an upregulation of thioesterification to CoA for long saturated FA chains 14:0 and 16:0, while a downregulation is observed for very long and unsaturated FA 20:5. Thioesterification to CoA of FA 16:1 is also predicted to be significantly downregulated by pFBA and FVA, while MOMA approach only suggests a non-significant decrease (12% decrease in *acsbG* flux compared to WT). Two of the three approaches (pFBA and FVA) predict a significant downregulation of ACS 18:3 in *acsbG*. Finally, pFBA and MOMA results also predict downregulation of FA 20:4 thioesterification to CoA. Nevertheless, FVA approaches predicts that WT and *acsbG* flux ranges for ACS 20:4 are overlapping ( $\xi = 0.34$ ) suggesting that downregulation predicted by pFBA and MOMA for *acsbG* could arise from lipid metabolism variability within iMgadi23. All together,

these results are consistent with the reported specificity of ACS proteins of the bubblegum family in the activation of long chains FAs and VLC-FAs with chain lengths from 16 to at least 24 carbons from studies in drosophila, rat and human (Min and Benzer 1999; Steinberg et al. 2000; Song et al. 2007). Moreover, these results are in close agreement with the findings reported in (Billey et al. 2021), where a significant accumulation of 16:0 acyl-CoA, a significant depletion of 16:1 and 18:3 acyl-CoA, and a minor depletion of 20:5 acyl-CoA were observed in the *acsbg* mutant. It is important to keep in mind that only lipidomic profile of *acsbg* mutant was used as BOF composition constraint in iMgadi23 and that no additional constraint was applied on ACS coding genes or reactions in model while conducting pFBA, MOMA and FVA. Still ACS reactions are predicted to be impacted in *acsbg* simulations with all three approaches.

**Table 3. FVA results for cytosolic ACS reactions.**

Reaction ID	WT [min,max] ( $\mu\text{mol.gDW}^{-1}.\text{h}^{-1}$ )	Acsbg [min,max] ( $\mu\text{mol.gDW}^{-1}.\text{h}^{-1}$ )	Fractional overlap
ACS120_c	[0,0]	[0,0]	1
ACS140_c	[0.21,0.21]	[0.41,0.41]	0
ACS160_c	[3.1,4.3]	[4.9,5.3]	0
ACS161_c	[3.9,4.0]	[3.1,3.1]	0
ACS180_c	[0,1.2]	[0,0.43]	0.36
ACS181_c	[4.5,4.5]	[2.4,2.4]	0
ACS182_c	[0,0]	[0,0]	1
ACS183_c	[5.1,5.1]	[2.2,2.2]	0
ACS203_c	[0,0]	[0,0]	1
ACS204_c	[0.03,0.38]	[0.02,0.15]	0.34
ACS205_c	[0.4,0.9]	[0.2,0.4]	0.07

To go further in deciphering *acsbg* phenotype, pFBA and MOMA predictions for the full lipid metabolism were compared (Figure 6). Respective pFBA and MOMA log<sub>2</sub>(FC) values are also recapitulated in Supplemental Text Tables S1 and S2 and Figures S2 and S3. As previously mentioned, cytosolic thioesterification of free FA to CoA by ACS is predicted to be upregulated for 14:0, 16:0 FAs and downregulated for 16:1 (significant only in pFBA\_acsbg), 20:4 and 20:5 FAs in *acsbg*. Similar results are predicted for FA-CoA transport to ER (Figure 6A). Plastidial 14:0-ACP and 16:0-ACP conversion to free FA (FAACPE reactions), followed by free FAs 14:0 and 16:0 export from chloroplast to cytosol (T\_F\_hc\_C

reactions) are also predicted to be upregulated in pFBA\_acsbg (for both FAs) and MOMA\_acsbg (only for 14:0) compared to pFBA\_WT (Figure 6B). Regarding, FAACPE160\_c and T\_F\_hc\_C160, fluxes are significantly increased in pFBA\_acsbg, but predicted increase is not significant with MOMA approach. Interestingly, while more 14:0-ACP and 16:0-ACP are converted to free FAs which are then exported from chloroplast, *de novo* synthesis of 14:0-ACP and 16:0-ACP is not predicted as upregulated (nor for any FA-ACP) in *acsbg* (Figure 6C). In pFBA\_acsbg, FA-ACP biosynthesis is even predicted to be downregulated. Impact of *acsbg* lipid profile on glycerolipid synthesis in ER is highly variable depending on *sn-1*, *sn-2* acyl chains (Figure 6D). However, TAGs biosynthesis via DGAT reactions is widely predicted to be upregulated both in pFBA\_acsbg and MOMA\_acsbg (Figure 6E and Supplemental File S10). Additionally, many LPAT and PAP reactions, involved in TAGs biosynthesis via the Kennedy pathway, are predicted to be upregulated, notably reactions incorporating 16:0 at the *sn-2* position of ER lipids and reactions incorporating 16:1 both at the *sn-1* and *sn-2* positions of ER lipids (Figure 6F). Finally, an overall downregulation tendency of plastid glycerolipid synthesis is predicted in *acsbg* mutant with both methods (Figure 6G). These results are consistent with proposed role for MgACSBG (Billey et al. 2021), highlighting that iMgadi23 constitutes a valuable tool to decipher mutant phenotype, notably regarding lipid metabolism. pFBA, MOMA and FVA results for all model reactions are available in Supplemental File S9.

- 2 Up regulated in pFBA\_acsbg and MOMA\_acsbg
- 1 Up regulated in one analysis (unchanged in the 2<sup>nd</sup>)
- 2 Down regulated in pFBA\_acsbg and MOMA\_acsbg
- 1 Down regulated in one analysis (unchanged in the 2<sup>nd</sup>)
- 0 Log2(FC) = 0 for pFBA and MOMA
- nan No conclusion
- Not in model

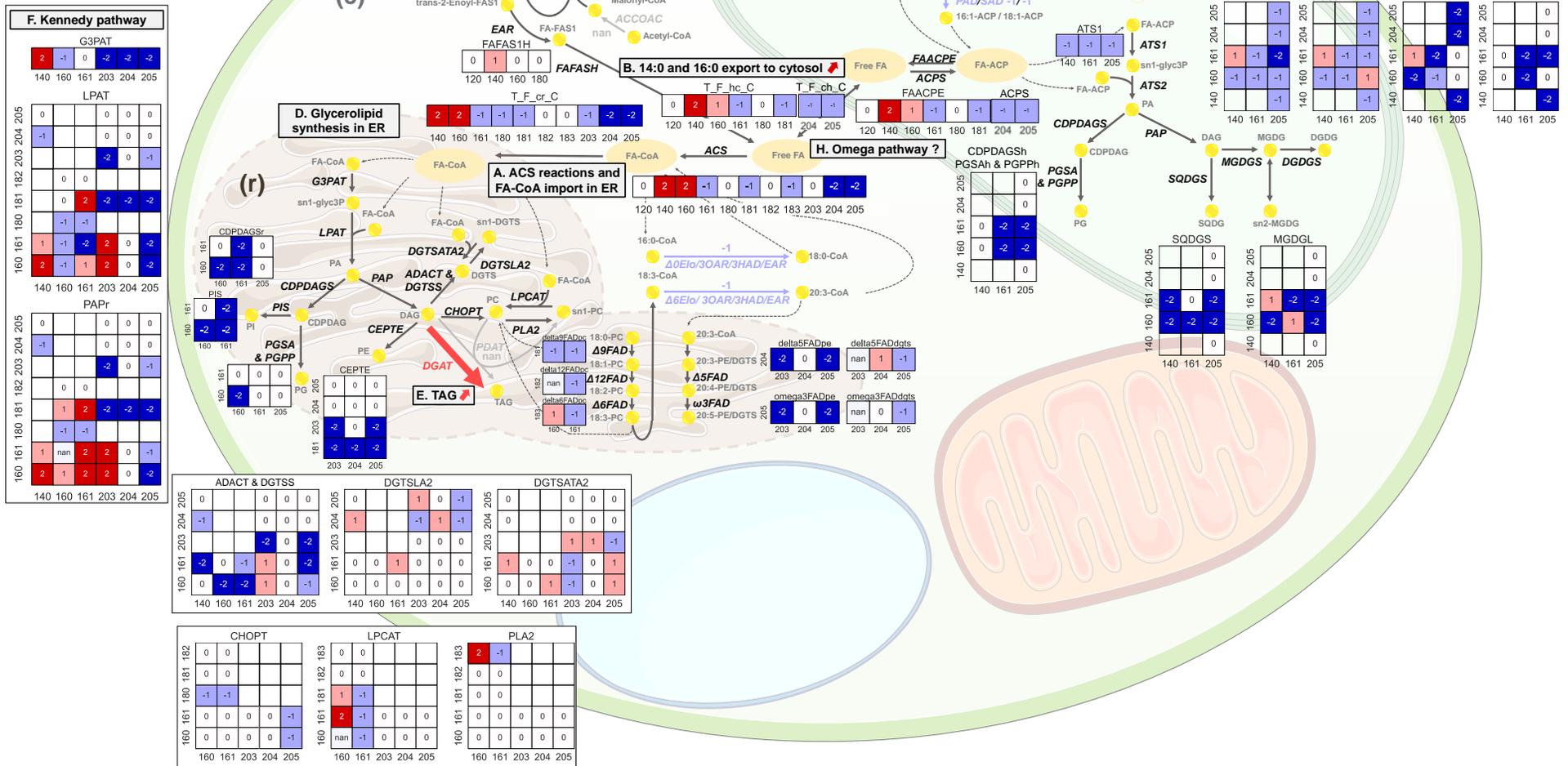


Figure 6. pFBA and MOMA predictions for lipid metabolism. (Legend: see next page)

For each reaction, results are shown for all acyl chains couples: *sn-1* (abscissa-axis) and *sn-2* (ordinate-axis). Red colors indicate reactions significantly upregulated in *acsbg* with both methods (dark red) or one method (light red). Blue colors indicate reactions significantly downregulated in *acsbg* with both methods (dark blue) or one method (light blue). Reactions for which same fluxes for WT and *acsbg* are predicted by both methods are indicated by 0. Reactions predicted up by one method but down by second method are notified as "nan" to indicate absence of conclusion. Finally, empty value indicates that reaction does not exist in *iMgadi23*. Main pathways of lipid metabolism are named and labelled from A to H. Cytosol, chloroplast and ER are indicated as (c), (h) and (r).

### 3. Discussion and perspectives

#### 3.1 *iMgadi23* model

GEMs are powerful tools to explore microorganisms metabolism. In this work, *iMgadi23* model of *M. gadi* was reconstructed and manually curated. Model ability to qualitatively and quantitatively predict *in vivo* growth phenotypes under diverse environmental and genetic conditions was validated. Our results highlight value of *iMgadi23* in guiding model-driven strain design and medium optimization to enhance growth or yields of target compounds.

A comprehensive 2D-pathway map of *iMgadi23* content is available in JSON format (Supplemental File S1). This map can serve as a platform to visualize and analyze flux distributions previously predicted using COBRA methods. The map can also facilitate the exploration of other relevant biological datasets including metabolic fluxes, lipidomics, metabolomics, and transcriptomics data (King et al. 2015). In addition, the map, together with *iMgadi23* GEM, enables the execution of interactive FBA simulations through Escher-FBA, a web-based open-source application (Rowe et al. 2018). Within this tool, users can interactively and easily adjust various FBA simulation parameters, including flux bounds, objective functions, and reaction knockouts and immediately visualize the impact on simulation results without programming skills.

#### 3.2 Modeling of the lipid metabolism

Due to diversity of existing lipid entities, metabolism tends to be the most complex part of metabolic models. One approach to get around lipid metabolism complexity, consists in considering that for each lipid metabolite *sn-1*, *sn-2* (and *sn-3* for TAGs) acyl chains are identical. This approach was used in *iRJ1321* model, greatly simplifying lipid metabolism (Shah et al. 2017). Using the example of TAG biosynthesis, which is of special interest for biotechnology applications, ten acyl chains (14:0, 16:0, 16:1, 18:0, 18:1, 18:2, 18:3\_ω3, 18:3\_ω6, 20:4 and 20:5) are considered in TAG metabolites in *iRJ1321*. Instead of considering all possible arrangements with repetition of three among ten acyl chains (resulting in 10<sup>3</sup> possible TAGs), only ten distinct TAGs with three identical acyl chains are described in *iRJ1321* (tag140\_c corresponding to TAG 14:0/14:0/14:0, tag160\_c for TAG 16:0/16:0/16:0 etc.). In *iMgadi23*, eight acyl chains were considered for TAGs (14:0, 16:0, 16:1, 18:0, 18:1, 20:3, 20:4 and 20:5). Also, rather than

considering only eight TAGs or describing all 512 ( $8^3$ ) putative possible TAGs, experimental data were used to determine existing TAGs (35 in total) in *M. gaditana* (Jouhet et al. 2017; Billey et al. 2021). Only reactions required for the metabolism of these 35 TAGs were added to model. For lipid synthesis in iMgadi23, unsaturation position within acyl chains was not taken into account. Consideration of these unsaturations position, notably regarding VLC-PUFAs could constitute an improvement of lipid metabolism description in our model.

In GEMs, the level of detail of lipid metabolism is often interrelated with the BOF lipid composition which depends on availability of detailed data. Indeed, exhaustive quantification of all lipid entities is not often available, resulting in BOF stoichiometric coefficients calculated from approximated lipid data. Here, data used for BOF lipid composition are based on FA quantification by gas chromatography (GC) and glycerolipid quantification by high-pressure liquid chromatography–tandem mass spectrometry (HPLC-MS/MS). Combination of both methods resulting in the precise quantification of the fatty acids present in each glycerolipid class (Jouhet et al. 2017; Billey et al. 2021). The strength of iMgadi23 is to be built and validated on such detailed lipidomic data that makes this model available for a deep analysis of lipid metabolism.

Finally, in term of reaction description, iMgadi23 constitutes an up-to-date overview of knowledge of *M. gaditana* lipid metabolism. Prediction of flux distribution together with model quality assessment with MEMOTE, revealed that among the different lipid pathways, all are functional except steroid biosynthesis. Due to gaps of knowledge of sterol metabolism its curation in our model would need further investigation. Sphingolipid metabolism is missing in iMgadi23. This pathway is described in both iNS937 and iRJ321 but non-functional, most reactions being blocked, likely due to the limited knowledge of sphingolipid metabolism in *Nannochloropsis and Microchloropsis*. Integrating functional sphingolipid metabolism into iMgadi23 would also require further biochemical and genetic characterizations. Gaps of knowledge for microalgal lipid synthesis, editing and degradation pathways still need to be filled notably for strains containing plastids with four limited membranes such as *M. gaditana* (Petroutsos et al. 2014; Hoffmann and Shachar-Hill 2023). As a result, in iMgadi23, some reactions were added although their existence in *M. gaditana* awaits experimental confirmation. For example, 18:0 and 18:1 synthesis in chloroplast has been described in higher plants but no clear evidence of its existence in microalgae has been reported so far. Similarly, acyl editing on PC, DGTS and MGDG is presumed to exist in *M. gaditana* although its actual existence requires validation (Hoffmann and Shachar-Hill 2023). Lipid metabolism characterization of *M. gaditana* and microalgae, in general, is an ongoing research topic. Continuously updating iMgadi23 with new knowledge could provide valuable insights into its quality and predictive capabilities.

Due to the precision in lipid metabolism modeling, including reaction description and compartmentation, and the high-quality data used for calculating the lipid composition of the BOF, iMgadiT23 stands as one of the most comprehensive models for lipid modeling in microalgae.

### 3.3 Analysis of *acsbg* mutant phenotype

Given the complexity and variability of lipid metabolism, characterizing and interpreting mutant strain phenotypes is very challenging, often requiring lot of experimental investigation that are costly and time consuming. In this study, we employed iMgadiT23 to elucidate the *acsbg* mutant phenotype. Two COBRA methods, pFBA and MOMA, were applied to predict impact of *acsbg* mutant lipid profile on reaction fluxes. The pFBA approach finds an optimal flux distribution which maximizes the objective function and minimizes the total sum of the absolute value of fluxes. On the other hand, MOMA minimizes the Euclidean distance from a reference flux distribution. Despite the difference in their mathematical objective, overall predictions for *acsbg* with pFBA and MOMA approaches consistently and effectively predicted the *acsbg* phenotype. In total 470 reactions are robustly predicted as up or downregulated by both methods. They notably predicted (1) the upregulation of cytosolic ACS 14:0 and 16:0; (2) the downregulation of ACS 20:5, 16:1 (ACS161\_c downregulation being significant only in pFBA\_acsbg); (3) the upregulation of many reactions of TAGs biosynthesis via the Kennedy pathway, notably reactions incorporating 16:0 and 16:1 at the *sn*-2 position of ER lipids; and (4) an increase export of free FAs 14:0 and 16:0 from chloroplast without increasing *de novo* FA-ACP synthesis. These predictions are consistent and partially confirm without any *a priori* the model presented by (Billey et al. 2021), particularly the proposed role for MgACSBG in the production of 16:1-CoA in the cytosol, prior to 16:1 incorporation at the *sn*-2 position of ER lipids via the Kennedy pathway. One important assumption for this proposed model is that positive feedback is exerted on the plastid, exporting more FAs. This assumption was validated by iMgadiT23 for 14:0 and 16:0 FAs. Nevertheless, a second role of MgACSBG in the production of 18:3-CoA was also proposed by (Billey et al. 2021) mainly based on their *in vivo* analysis of acyl-CoA profiles. This analysis revealed a significant depletion in 16:1-CoA and 18:3-CoA along with an accumulation of 16:0-CoA and 18:1-CoA in *acsbg* mutant. While pFBA and FVA predict a significant downregulation of ACS183\_c in *acsbg* (with a  $\log_2(\text{pFBA}_{\text{acsbg}}/\text{pFBA}_{\text{WT}}) = -1.2$ , which is lower than  $\log_2(\mu_{\text{acsbg}}/\mu_{\text{WT}}) = -1$ ), MOMA does not predict this downregulation ( $\log_2(\text{MOMA}_{\text{acsbg}}/\text{pFBA}_{\text{WT}}) = 0$ ). This could be partially explained by the steady state assumption inherent in COBRA methods used in this study. Accumulation or depletion of these metabolites could not be taken into account in the differential flux analyses, resulting in the absence of ACS183\_c downregulation in MOMA\_acsbg.

Regarding pFBA approach, on the one hand using experimental growth rates ( $\mu_{WT} = 0.030 \text{ h}^{-1}$ ,  $\mu_{acsbg} = 0.015 \text{ h}^{-1}$ ) to constrain BOF enables consideration of the mutant growth defect in predicted flux distributions. On the other hand, this approach has the drawback of potentially masking downregulated reactions.

Numerous energy-consuming reactions, such as NGAM, alternative oxidase (AOX) along with other futile cycles are activated in MOMA\_acsbg (these reactions have a flux in *acsbg* mutant but not in WT). Cytosolic FA 14:0 synthesis is activated in MOMA\_acsbg while 14:0 degradation is also activated. All these activations may be attributed to similar fluxes in photosynthesis, but with BOF demand being only half as much in *acsbg* compared to WT. Consequently, the excess energy not utilized for biomass production must be consumed. Interestingly, peroxisomal  $\beta$ -oxidation of FAs 20:5, 20:4, 20:3, 18:3, 16:1 and 14:0 was predicted to have a significant flux in MOMA\_acsbg. This activation of peroxisomal FAs degradation comes with the activation of glyoxylate cycle and photorespiration, as well as increased fluxes in TCA cycle, citrate/malate/oxaloacetate metabolism, serine/glycine metabolism and pentose phosphate pathway. *In silico*, degradation of 20:5, 20:4, 20:3, 18:3, 16:1 and 14:0 FAs suggests that these FAs are less needed for incorporation into glycerolipids in MOMA\_acsbg. MOMA, along with FBA and pFBA approaches, relies on steady state assumption *i.e.* total the total production of a metabolite must balance with its total consumption. Consequently, these methods cannot predict the accumulation of FAs, resulting in their degradation in MOMA\_acsbg. *In vivo*, the pool of available FA-CoAs is modified (Billey et al. 2021). Activation and upregulation of  $\beta$ -oxidation, glyoxylate cycle, photorespiration and TCA cycle is therefore a biologically plausible assumption that need to be experimentally validated.

*In silico* differential flux analyses between WT and *acsbg* strains demonstrate the utility of iMgadi23 as a valuable tool to decipher mutant phenotypes and ultimately increase our understanding of *M. gaditana* metabolism. Similar methodologies can be applied to elucidate the metabolic phenotype resulting from metabolic engineering or variation in culture parameters including light, temperature, nitrogen or phosphate availability.

#### 4. Conclusion

GEMs are powerful tools to explore the metabolism of biological systems. In this study, we reconstructed, manually curated and experimentally validated iMgadi23, a high-quality GEM for the industrially promising oleaginous microalga *M. gaditana*, with a special focus on lipid metabolism. To date, iMgadi23 is the most comprehensive GEM available for *Nannochloropsis* and *Microchloropsis* species, especially with respect to lipid metabolism. Our analyses showed that iMgadi23 can accurately predict growth

phenotypes under different media and genetic conditions, both qualitatively and quantitatively. Moreover, the *in silico* analysis of the *acsbg* mutant phenotype highlights the potential of iMgadi23 to elucidate the physiological behavior and metabolic states of *M. gaditana* in response to genetic or environmental changes. We believe that iMgadi23, together with its comprehensive 2D-pathway map, will be valuable tools for identifying novel metabolic engineering strategies aimed at enhancing lipid synthesis and accumulation in *M. gaditana*, thus opening new opportunities using this strain as a platform for various industrial applications.

## 5. Material and Methods

### 5.1 Draft model automatic reconstruction and manual curation

iMgadi23 was reconstructed based on the complete genome sequence of *M. gaditana* B-31 (GenBank accession number GCA\_000569095.1) (Corteggiani Carpinelli et al. 2014), according to the methodology presented in (Pereira et al. 2018). Briefly, model reconstruction was carried out in three different phases: (1) genome reannotation, (2) draft metabolic model building, (3) model completion and refinement by network gaps filling, compartmentalization of reactions and metabolites, addition of missing pathways and correction of inconsistencies based on literature and expert knowledge.

### 5.2 Biomass objective functions composition

iMgadi23 contains two independent BOFs, representing respectively biomass composition of WT and *MgACSBG#31* mutant strains (Billey et al. 2021). Each biomass objective function involves 17 pseudo-reactions and pseudo-metabolites representing BIOMASS, the different macromolecules (DNA, RNA, PROTEINS, CARBOHYDRATES, PIGMENTS and LIPIDS) and glycerolipid classes (DAG, TAG, PC, PE, PI, PG, DGTS, DGDG, MGDG, SQDG). To facilitate their identification, all pseudo-reactions and associated pseudo-metabolites involved in the BOF were respectively tagged “WT526” and “MgACSBG31”. Detailed lipidomic profiles in terms of glycerolipid classes and acyl chains profiles of glycerolipids reported for “untransformed WT” and “MgACSBG#31” strains from (Billey et al. 2021) were used to calculate lipid stoichiometric coefficients. Other stoichiometric coefficients of both BOFs are based on diverse experimental and literature data. Detailed biomass composition, original data, sources and final stoichiometric coefficients are detailed in Supplemental Material and Methods and Supplemental Files S2 and S3.

### 5.3 Model simulations and validation

All flux distributions were performed using open-source COBRApy package (Ebrahim et al. 2013; version 0.21.0) using cplex (version 12.8.0.0) as solver, with Python (version 3.6.13). When not otherwise stipulated, default medium was used for simulations, representing photoautotrophic growth with  $\text{NO}_3$  as nitrogen source with unlimited uptakes of  $\text{CO}_2$ ,  $\text{O}_2$ ,  $\text{H}_2\text{O}$ , protons, phosphate, sulphate, magnesium, mannose, and 6-deoxy-galactose. This entailed setting the lower bounds of the corresponding exchange reactions to  $-1000$ , as we used the usual convention of writing the exchange reactions in such way that production corresponds to positive fluxes and consumption to negative ones. Upper and lower bounds of photon and  $\text{NO}_3$  uptake were adjusted for each simulation. See Supplemental Material and Methods for more details.

#### **MEMOTE**

Model quality was assessed by MEMOTE Test Suite, which calculates an independent and comparable score based on standardized tests for diverse GEM features (Lieven et al. 2020). Command version of MEMOTE (Platform Windows, Memote Version 0.12.0) was run on a conda environment with Python (version 3.6.13). MEMOTE was also run for other *Nannochloropsis* and *Microchloropsis* GEMs discussed in this study (Shah et al. 2017; Loira et al. 2017; Belcour et al. 2023). All complete MEMOTE reports are available in Supplemental Files S4-S7.

#### **Quantitative validation**

Quantitative capability of iMgadi23 to predict growth was assessed using pFBA approach with WT BOF maximization as objective function.  $\text{NO}_3$  uptake rates reported by (Rafay et al. 2020; 0.119, 0.124 and 0.133  $\text{mmol}_{\text{NO}_3}.\text{gDW}^{-1}.\text{h}^{-1}$ ) were used to constrain EX\_no3\_ reaction. Unlimited photon uptake was allowed by setting EX\_photon\_abs\_ lower bound to  $-1000$ .

For phenotypic phase-planes analysis, couples of  $\text{NO}_3$ -photon, and  $\text{NO}_3$ - $\text{CO}_2$  values were used to constrain  $\text{NO}_3$ ,  $\text{CO}_2$  and photon exchange reactions lower bound. Respectively,  $\text{NO}_3$ , photon and  $\text{CO}_2$  uptake values (in  $\text{mmol}.\text{gDW}^{-1}.\text{h}^{-1}$ ) range in  $[0,0.12]$ ,  $[0,25]$  and  $[0,1.2]$ . Maximum growth rate was predicted by pFBA using WT BOF as objective function.

#### **Qualitative validation**

Qualitative growth/no-growth predictions were performed using pFBA using WT BOF as objective function. Detailed constraints applied for these simulations are available in Supplemental File S11 . Model

predictions were compared to *in vivo* growth data from literature (Xu et al. 2004; Kilian et al. 2011; Gim et al. 2016; Loira et al. 2017). *In silico* growth predictions were categorized as: TP, TN, FP and FN.

- TP (true positive: model simulation predicts growth while growth is observed *in vivo*),
- TN (true negative: model simulation predicts no growth while no growth is observed *in vivo*),
- FP (false positive: model simulation predicts growth while no growth is observed *in vivo*),
- FN (false negative: model simulation predicts no growth while growth is observed *in vivo*).

Sensitivity, specificity, and accuracy scores were calculated according to equations (1-3).

$$Sensitivity = \frac{TP}{TP + FN} \quad (Eq. 1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (Eq. 2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (Eq. 3)$$

#### 5.4 Differential flux analysis between WT and *acsbg* mutant strains

##### **pFBA WT and pFBA *acsbg* flux distributions**

pFBA was carried out employing the minimization of photon uptake as objective function, and constraining respectively WT BOF (BIOMASS\_biomass\_WT526) to 0.030 h<sup>-1</sup> and *acsbg* mutant BOF (BIOMASS\_biomass\_MgACSBG31) to 0.015 h<sup>-1</sup>; experimental growth rates reported by (Billey et al. 2021). Unlimited NO<sub>3</sub> uptake was allowed and photon uptake lower bound was constrained to -331 mmol<sub>photon</sub>.m<sup>-2</sup>.s<sup>-1</sup>.

##### **MOMA *acsbg* flux distribution**

MOMA was carried out employing using pFBA\_WT as reference flux distribution, and constraining *acsbg* mutant BOF to 0.015 h<sup>-1</sup>. Unlimited NO<sub>3</sub> uptake was allowed and photon uptake lower bound was constrained to -331 mmol<sub>photon</sub>.m<sup>-2</sup>.s<sup>-1</sup>. Lower and upper bounds of carbohydrates and amino acids exchange reactions were set to 0.

##### **Differential flux analysis**

For all aforementioned flux distributions, fluxes were rounded to 10<sup>-6</sup>. For each iMgadi23 reaction, the change in flux between *acsbg* and WT was calculated as follows:

$$\log_2(FC) = \log_2\left(\frac{\text{flux } acsbg}{\text{flux WT}}\right) \quad (Eq. 4)$$

Where  $flux_{acsbg}$  corresponds to pFBA\_acsbg or MOMA\_acsbg and  $flux_{WT}$  to pFBA\_WT.  $\log_2(FC)$  of -0.263 and 0.263 (corresponding to a fold change of 20% in flux) were used to filter reactions with insignificant flux change. When no flux was predicted for both *acsbg* and WT, reaction was considered as “both\_off”. When zero flux was predicted for WT and non-zero flux was predicted for *acsbg*, reaction was considered as activated “ON”. Finally, when non-zero flux was predicted to WT and zero flux was predicted for *acsbg*, reaction was considered as shut down “OFF”. Sign change in predicted fluxes was also considered by introduction of “neg\_to\_pos” and “pos\_to\_neg” labels.

### **FVA WT and FVA acsbg**

For FVA\_WT, WT BOF was constrained to 0.030 h<sup>-1</sup> and FVA with an optimality constraint of 100% was applied with minimization of photon uptake as objective function. Identically, for FVA\_acsbg, BIOMASS\_biomass\_MgACSBG31 was constrained to 0.015 h<sup>-1</sup> and FVA was applied with minimization of photon uptake as objective function, with an optimality constraint of 100%. For both simulations, unlimited NO<sub>3</sub> uptake was allowed and photon uptake lower bound was constrained to -331 mmol<sub>photon</sub>·m<sup>-2</sup>·s<sup>-1</sup>. To compare the resulting minimal and maximal reaction fluxes, for each reaction *i* within iMgadi23, the fractional overlap ( $\xi_i$ ) of the corresponding flux range between both strains was calculated as defined by (Simensen et al. 2022):

$$\xi_i = \frac{\max(0, \min(\alpha_{max,i}, \beta_{max,i}) - \max(\alpha_{min,i}, \beta_{min,i}))}{\max(|\alpha_{max,i} - \alpha_{min,i}|, |\beta_{max,i} - \beta_{min,i}|)} \quad (Eq. 5)$$

with  $\alpha_{min,i}$  and  $\alpha_{max,i}$  minimal and maximal fluxes for the given reaction *i* in FVA\_WT, and  $\beta_{min,i}$  and  $\beta_{max,i}$  minimal and maximal fluxes in FVA\_acsbg.

### *5.5 iMgadi23 distribution*

iMgadi23 is publicly available under a Creative Commons NonCommercial 4.0 International (CC BY-NC 4.0) license. Model and all scripts used for analyses presented in this study are available in a GitHub repository (<https://github.com/Total-RD/Mgadi23>), following the standard-GEM format (<https://github.com/MetabolicAtlas/standard-GEM>; Anton et al. 2023). Model reconstruction is provided in SBML, JSON, MAT and XLSX formats compatible with the metabolic engineering platforms and computational tools for metabolic model simulations (COBRA, Escher and others) (Supplemental Files S12-S15). Detailed 2D-pathway map of iMgadi23 content is available as JSON file compatible with Escher (King et al. 2015) (Supplemental File S1).

## References

Alboresi, Alessandro; Perin, Giorgio; Vitulo, Nicola; Diretto, Gianfranco; Block, Maryse; Jouhet, Juliette et al. (2016) Light Remodels Lipid Biosynthesis in *Nannochloropsis gaditana* by Modulating Carbon Partitioning between Organelles. In : *Plant physiology*, vol. 171, n° 4, p. 2468–2482. DOI: 10.1104/pp.16.00599.

Al-Hoqani, Umaima; Young, Rosanna; Purton, Saul (2017) The biotechnological potential of *Nannochloropsis*. In : *Perspectives in Phycology*, vol. 4, n° 1, p. 1–15. DOI: 10.1127/pip/2016/0065.

Anton, Mihail; Almaas, Eivind; Benfeitas, Rui; Benito-Vaquerizo, Sara; Blank, Lars M.; Dräger, Andreas et al. (2023) standard-GEM: standardization of open-source genome-scale metabolic models. In : *Preprint bioRxiv*. DOI: 10.1101/2023.03.21.512712.

Belcour, Arnaud; Got, Jeanne; Aite, Méziane; Delage, Ludovic; Collén, Jonas; Frioux, Clémence et al. (2023) Inferring and comparing metabolism across heterogeneous sets of annotated genomes using AuCoMe. In : *Genome research*. DOI: 10.1101/gr.277056.122.

Billey, Elodie; Magneschi, Leonardo; Leterme, Sébastien; Bedhomme, Mariette; Andres-Robin, Amélie; Poulet, Laurent et al. (2021) Characterization of the *Bubblegum* acyl-CoA synthetase of *Microchloropsis gaditana*. In : *Plant physiology*, vol. 185, n° 3, p. 815–835. DOI: 10.1093/plphys/kiab110.

Bo, Davide Dal; Magneschi, Leonardo; Bedhomme, Mariette; Billey, Elodie; Deragon, Etienne; Storti, Mattia et al. (2021) Consequences of Mixotrophy on Cell Energetic Metabolism in *Microchloropsis gaditana* Revealed by Genetic Engineering and Metabolic Approaches. In : *Frontiers in plant science*, vol. 12, p. 628684. DOI: 10.3389/fpls.2021.628684.

Boyle, Nanette R.; Morgan, John A. (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. In : *BMC systems biology*, vol. 3, p. 4. DOI: 10.1186/1752-0509-3-4.

Burgard, A. P.; Vaidyaraman, S.; Maranas, C. D. (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. In : *Biotechnology progress*, vol. 17, n° 5, p. 791–797. DOI: 10.1021/bp0100880.

Chu, Wan-Loy (2012) Biotechnological applications of microalgae. In : *leJSME*, vol. 6, n° 1, S24-S37.

Corteggiani Carpinelli, Elisa; Telatin, Andrea; Vitulo, Nicola; Forcato, Claudio; D'Angelo, Michela; Schiavon, Riccardo et al. (2014) Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. In : *Molecular plant*, vol. 7, n° 2, p. 323–335. DOI: 10.1093/mp/sst120.

Dolch, Lina-Juana; Rak, Camille; Perin, Giorgio; Tourcier, Guillaume; Broughton, Richard; Leterrier, Marina et al. (2017) A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in *Nannochloropsis*. In : *Plant physiology*, vol. 173, n° 1, p. 742–759. DOI: 10.1104/pp.16.01420.

Ebrahim, Ali; Lerman, Joshua A.; Palsson, Bernhard O.; Hyduke, Daniel R. (2013) COBRApy: COntstraints-Based Reconstruction and Analysis for Python. In : *BMC Systems Biology*, vol. 7, n° 1, p. 74. DOI: 10.1186/1752-0509-7-74.

Fang, Xu; Wei, Cong; Zhao-Ling, Cai; Fan, Ouyang (2004) Effects of organic carbon sources on cell growth and eicosapentaenoic acid content of *Nannochloropsis* sp. In : *Journal of Applied Phycology*, vol. 16, n° 6, p. 499–503. DOI: 10.1007/s10811-004-5520-1.

Fawley, Marvin W.; Jameson, Ian; Fawley, Karen P. (2015) The phylogeny of the genus *Nannochloropsis* (Monodopsidaceae, Eustigmatophyceae), with descriptions of *N. australis* sp. nov. and *Microchloropsis* gen. nov. In : *Phycologia*, vol. 54, n° 5, p. 545–552. DOI: 10.2216/15-60.1.

Fu, Weiqi; Nelson, David R.; Mystikou, Alexandra; Daakour, Sarah; Salehi-Ashtiani, Kourosh (2019) Advances in microalgal research and engineering development. In : *Current opinion in biotechnology*, vol. 59, p. 157–164. DOI: 10.1016/j.copbio.2019.05.013.

Gentile, M. P.; Blanch, H. W. (2001) Physiology and xanthophyll cycle activity of *Nannochloropsis gaditana*. In : *Biotechnology and bioengineering*, vol. 75, n° 1, p. 1–12. DOI: 10.1002/bit.1158.

Gim, Geun Ho; Ryu, Jaewon; Kim, Moon Jong; Kim, Pyung Il; Kim, Si Wouk (2016) Effects of carbon source and light intensity on the growth and total lipid production of three microalgae under different culture conditions. In : *Journal of industrial microbiology & biotechnology*, vol. 43, n° 5, p. 605–616. DOI: 10.1007/s10295-016-1741-y.

Gong, Yanhai; Kang, Nam K.; Kim, Young U.; Wang, Zengbin; Wei, Li; Xin, Yi et al. (2020) The NanDeSyn database for *Nannochloropsis* systems and synthetic biology. In : *The Plant journal : for cell and molecular biology*, vol. 104, n° 6, p. 1736–1745. DOI: 10.1111/tpj.15025.

Gu, Changdai; Kim, Gi Bae; Kim, Won Jun; Kim, Hyun Uk; Lee, Sang Yup (2019) Current status and applications of genome-scale metabolic models. In : *Genome biology*, vol. 20, n° 1, p. 121. DOI: 10.1186/s13059-019-1730-3.

Gudmundsson, Steinn; Thiele, Ines (2010) Computationally efficient flux variability analysis. In : *BMC bioinformatics*, vol. 11, p. 489. DOI: 10.1186/1471-2105-11-489.

Heirendt, Laurent; Arreckx, Sylvain; Pfau, Thomas; Mendoza, Sebastián N.; Richelle, Anne; Heinken, Almut et al. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. In : *Nature protocols*, vol. 14, n° 3, p. 639–702. DOI: 10.1038/s41596-018-0098-2.

Hoffmann, Danielle Yvonne; Shachar-Hill, Yair (2023) Do betaine lipids replace phosphatidylcholine as fatty acid editing hubs in microalgae? In : *Frontiers in plant science*, vol. 14, p. 1077347. DOI: 10.3389/fpls.2023.1077347.

Jouhet, Juliette; Lupette, Josselin; Clerc, Olivier; Magneschi, Leonardo; Bedhomme, Mariette; Collin, Séverine et al. (2017) LC-MS/MS versus TLC plus GC methods: Consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation. In : *PloS one*, vol. 12, n° 8, e0182423. DOI: 10.1371/journal.pone.0182423.

Kanehisa, Minoru; Goto, Susumu (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. In : *Nucleic acids research*, vol. 28, n° 1, p. 27.

Kareya, Mukul Suresh; Mariam, Iqra; Shaikh, Kashif Mohd; Nesamma, Asha Arumugam; Jutur, Pannaga Pavan (2020) Photosynthetic Carbon Partitioning and Metabolic Regulation in Response to Very-Low and

High CO<sub>2</sub> in *Microchloropsis gaditana* NIES 2587. In : *Frontiers in plant science*, vol. 11, p. 981. DOI: 10.3389/fpls.2020.00981.

Keeling, Patrick J. (2004) Diversity and evolutionary history of plastids and their hosts. In : *American journal of botany*, vol. 91, n° 10, p. 1481–1493. DOI: 10.3732/ajb.91.10.1481.

Kilian, Oliver; Benemann, Christina S. E.; Niyogi, Krishna K.; Vick, Bertrand (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, n° 52, p. 21265–21269. DOI: 10.1073/pnas.1105861108.

Kim, Chul Woong; Sung, Min-Gyu; Nam, Kibok; Moon, Myoungsoon; Kwon, Jong-Hee; Yang, Ji-Won (2014) Effect of monochromatic illumination on lipid accumulation of *Nannochloropsis gaditana* under continuous cultivation. In : *Bioresource technology*, vol. 159, p. 30–35. DOI: 10.1016/j.biortech.2014.02.024.

King, Zachary A.; Dräger, Andreas; Ebrahim, Ali; Sonnenschein, Nikolaus; Lewis, Nathan E.; Palsson, Bernhard O. (2015) Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. In : *PLoS computational biology*, vol. 11, n° 8, e1004321. DOI: 10.1371/journal.pcbi.1004321.

Kong, Fantao; Romero, Ismael Torres; Warakanont, Jaruswan; Li-Beisson, Yonghua (2018) Lipid catabolism in microalgae. In : *The New phytologist*, vol. 218, n° 4, p. 1340–1348. DOI: 10.1111/nph.15047.

Lewis, Nathan E.; Hixson, Kim K.; Conrad, Tom M.; Lerman, Joshua A.; Charusanti, Pep; Polpitiya, Ashoka D. et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. In : *Molecular systems biology*, vol. 6, p. 390. DOI: 10.1038/msb.2010.47.

Li-Beisson, Yonghua; Thelen, Jay J.; Fedosejevs, Eric; Harwood, John L. (2019) The lipid biochemistry of eukaryotic algae. In : *Progress in lipid research*, vol. 74, p. 31–68. DOI: 10.1016/j.plipres.2019.01.003.

Lieven, Christian; Beber, Moritz E.; Olivier, Brett G.; Bergmann, Frank T.; Ataman, Meric; Babaei, Parizad et al. (2020) MEMOTE for standardized genome-scale metabolic model testing. In : *Nature biotechnology*, vol. 38, n° 3, p. 272–276. DOI: 10.1038/s41587-020-0446-y.

Lima, Serena; Brucato, Alberto; Caputo, Giuseppe; Schembri, Luca; Scargiali, Francesca (2022) Modelling *Nannochloropsis gaditana* Growth in Reactors with Different Geometries, Determination of Kinetic Parameters and Biochemical Analysis in Response to Light Intensity. In : *Applied Sciences*, vol. 12, n° 12, p. 5776. DOI: 10.3390/app12125776.

Loira, Nicolás; Mendoza, Sebastian; Paz Cortés, María; Rojas, Natalia; Travisany, Dante; Di Genova, Alex et al. (2017) Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production. In : *BMC systems biology*, vol. 11, n° 1, p. 66. DOI: 10.1186/s12918-017-0441-1.

Ma, Xiangmeng; Mi, Yuwei; Zhao, Chen; Wei, Qun (2022) A comprehensive review on carbon source effect of microalgae lipid accumulation for biofuel production. In : *The Science of the total environment*, vol. 806, n° Pt 3, p. 151387. DOI: 10.1016/j.scitotenv.2021.151387.

Ma, Xiao-Nian; Chen, Tian-Peng; Yang, Bo; Liu, Jin; Chen, Feng (2016) Lipid Production from Nannochloropsis. In : Marine Drugs, vol. 14, n° 4. DOI: 10.3390/md14040061.

Min, K. T.; Benzer, S. (1999) Preventing neurodegeneration in the Drosophila mutant bubblegum. In : Science (New York, N.Y.), vol. 284, n° 5422, p. 1985–1988. DOI: 10.1126/science.284.5422.1985.

Mühlroth, Alice; Winge, Per; El Assimi, Aimen; Jouhet, Juliette; Maréchal, Eric; Hohmann-Marriott, Martin F. et al. (2017) Mechanisms of Phosphorus Acquisition and Lipid Class Remodeling under P Limitation in a Marine Microalga. In : Plant physiology, vol. 175, n° 4, p. 1543–1559. DOI: 10.1104/pp.17.00621.

Muñoz, Camilo F.; Südfeld, Christian; Naduthodi, Mihris I. S.; Weusthuis, Ruud A.; Barbosa, Maria J.; Wijffels, René H.; D'Adamo, Sarah (2021) Genetic engineering of microalgae for enhanced lipid production. In : Biotechnology advances, vol. 52, p. 107836. DOI: 10.1016/j.biotechadv.2021.107836.

Norsigian, Charles J.; Pusarla, Neha; McConn, John Luke; Yurkovich, James T.; Dräger, Andreas; Palsson, Bernhard O.; King, Zachary (2020) BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. In : Nucleic acids research, vol. 48, n° D1, D402–D406. DOI: 10.1093/nar/gkz1054.

Orth, Jeffrey D.; Thiele, Ines; Palsson, Bernhard Ø. (2010) What is flux balance analysis? In : Nature biotechnology, vol. 28, n° 3, p. 245–248. DOI: 10.1038/nbt.1614.

Pereira, Bruno; Miguel, Joana; Vilaça, Paulo; Soares, Simão; Rocha, Isabel; Carneiro, Sónia (2018) Reconstruction of a genome-scale metabolic model for Actinobacillus succinogenes 130Z. In : BMC systems biology, vol. 12, n° 1, p. 61. DOI: 10.1186/s12918-018-0585-7.

Petroutsos, Dimitris; Amiar, Souad; Abida, Heni; Dolch, Lina-Juana; Bastien, Olivier; Rébeillé, Fabrice et al. (2014) Evolution of galactoglycerolipid biosynthetic pathways - From cyanobacteria to primary plastids and from primary to secondary plastids. In : Progress in lipid research, vol. 54, p. 68–85. DOI: 10.1016/j.plipres.2014.02.001.

Pham, Nhung (2016) Genome-scale constraint-based metabolic modeling and analysis of Nannochloropsis Sp.

Poliner, Eric; Panchy, Nicholas; Newton, Linsey; Wu, Guangxi; Lapinsky, Andrew; Bullard, Blair et al. (2015) Transcriptional coordination of physiological responses in Nannochloropsis oceanica CCMP1779 under light/dark cycles. In : The Plant journal : for cell and molecular biology, vol. 83, n° 6, p. 1097–1113. DOI: 10.1111/tpj.12944.

Rafay, Ramis; Uratani, Joao M.; Hernandez, Hector H.; Rodríguez, Jorge (2020) Growth and Nitrate Uptake in Nannochloropsis gaditana and Tetraselmis chuii Cultures Grown in Sequential Batch Reactors. In : Frontiers in Marine Science, vol. 7. DOI: 10.3389/fmars.2020.00077.

Ren, Ming; Ogden, Kimberly (2014) Cultivation of Nannochloropsis gaditana on mixtures of nitrogen sources. In : Environmental Progress & Sustainable Energy, vol. 33, n° 2, p. 551–555. DOI: 10.1002/ep.11818.

Rizwan, Muhammad; Mujtaba, Ghulam; Memon, Sheraz Ahmed; Lee, Kisay; Rashid, Naim (2018) Exploring the potential of microalgae for new biotechnology applications and beyond: A review. In : *Renewable and Sustainable Energy Reviews*, vol. 92, p. 394–404. DOI: 10.1016/j.rser.2018.04.034.

Rock, Alberto; Novoveská, Lucie; Green, David (2021) Synthetic biology is essential to unlock commercial biofuel production through hyper lipid-producing microalgae: a review. In : *Applied Phycology*, vol. 2, n° 1, p. 41–59. DOI: 10.1080/26388081.2021.1886872.

Rowe, Elliot; Palsson, Bernhard O.; King, Zachary A. (2018) Escher-FBA: a web application for interactive flux balance analysis. In : *BMC systems biology*, vol. 12, n° 1, p. 84. DOI: 10.1186/s12918-018-0607-5.

Sajjadi, Baharak; Chen, Wei-Yin; Raman, Abdul. Aziz. Abdul; Ibrahim, Shaliza (2018) Microalgae lipid and biomass for biofuel production: A comprehensive review on lipid enhancement strategies and their effects on fatty acid composition. In : *Renewable and Sustainable Energy Reviews*, vol. 97, p. 200–232. DOI: 10.1016/j.rser.2018.07.050.

Segrè, Daniel; Vitkup, Denis; Church George M. (2002) Analysis of optimality in natural and perturbed metabolic networks. In : *PNAS*, vol. 99, n° 23, p. 15112–15117.

Shah, Ab Rauf; Ahmad, Ahmad; Srivastava, Shireesh; Jaffar Ali, B. M. (2017) Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. In : *Algal Research*, vol. 26, p. 354–364. DOI: 10.1016/j.algal.2017.08.014.

Simensen, Vetle; Schulz, Christian; Karlsen, Emil; Bråtelund, Signe; Burgos, Idun; Thorfinnsdottir, Lilja Brekke et al. (2022) Experimental determination of *Escherichia coli* biomass composition for constraint-based metabolic modeling. In : *PloS one*, vol. 17, n° 1, e0262450. DOI: 10.1371/journal.pone.0262450.

Song, Si-Young; Kato, Chieko; Adachi, Eijiro; Moriya-Sato, Ayako; Inagawa-Ogashiwa, Masayo; Umeda, Rieko; Hashimoto, Naohiro (2007) Expression of an acyl-CoA synthetase, lipidosisin, in astrocytes of the murine brain and its up-regulation during remyelination following cuprizone-induced demyelination. In : *Journal of neuroscience research*, vol. 85, n° 16, p. 3586–3597. DOI: 10.1002/jnr.21456.

Spolaore, Pauline; Joannis-Cassan, Claire; Duran, Elie; Isambert, Arsène (2006) Commercial applications of microalgae. In : *Journal of bioscience and bioengineering*, vol. 101, n° 2, p. 87–96. DOI: 10.1263/jbb.101.87.

Steinberg, S. J.; Morgenthaler, J.; Heinzer, A. K.; Smith, K. D.; Watkins, P. A. (2000) Very long-chain acyl-CoA synthetases. Human "bubblegum" represents a new family of proteins capable of activating very long-chain fatty acids. In : *The Journal of biological chemistry*, vol. 275, n° 45, p. 35162–35169. DOI: 10.1074/jbc.M006403200.

Tibocha-Bonilla, Juan D.; Zuñiga, Cristal; Godoy-Silva, Rubén D.; Zengler, Karsten (2018) Advances in metabolic modeling of oleaginous microalgae. In : *Biotechnology for biofuels*, vol. 11, p. 241.

Vieler, Astrid; Wu, Guangxi; Tsai, Chia-Hong; Bullard, Blair; Cornish, Adam J.; Harvey, Christopher et al. (2012) Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. In : *PLoS genetics*, vol. 8, n° 11, e1003064. DOI: 10.1371/journal.pgen.1003064.

Xu, Fang; Cong, Wei; Cai, Zhao-Ling; Ouyang, Fan (2004) Effects of organic carbon sources on cell growth and eicosapentaenoic acid content of *Nannochloropsis* sp. In : *Journal of applied phycology*, vol. 16, p. 499–503.

Ye, Chao; Wei, Xinyu; Shi, Tianqiong; Sun, Xiaoman; Xu, Nan; Gao, Cong; Zou, Wei (2022) Genome-scale metabolic network models: from first-generation to next-generation. In : *Applied microbiology and biotechnology*, vol. 106, n° 13-16, p. 4907–4920. DOI: 10.1007/s00253-022-12066-y.

Zhang, Dongmei; Yan, Fei; Sun, Zhongliang; Zhang, Qinghua; Xue, Shengzhang; Cong, Wei (2014) On-line modeling intracellular carbon and energy metabolism of *Nannochloropsis* sp. in nitrogen-repletion and nitrogen-limitation cultures. In : *Bioresource technology*, vol. 164, p. 86–92. DOI: 10.1016/j.biortech.2014.04.083.

## 2.4 Discussion

### 2.4.1 Choix de la stratégie de simulation de flux

Comme cela a été exposé en introduction, la maximisation de la fonction de biomasse est l'objectif de simulation le plus couramment employé en FBA (ou dans des méthodes dérivées telles que la pFBA ou la MOMA). Cependant, d'autres objectifs de simulation tels que la maximisation de la production d'ATP, la minimisation de la consommation de la source de carbone, la minimisation de l'utilisation du potentiel redox etc. ont été explorés dans diverses études (Schuetz, Kuepfer et Sauer 2007; Nookaew et al. 2008; Kavšček et al. 2015; Hosseini 2020; Yasemi et Jolicoeur 2021). Par ailleurs, réaliser une simulation par FBA en minimisant le taux d'absorption d'un substrat (par exemple du glucose) tout en fixant la production de biomasse, conduit à des résultats très similaires à ceux obtenus en fixant le taux d'absorption de ce substrat et en maximisant la fonction de biomasse (Nookaew et al. 2008; Schuetz, Kuepfer et Sauer 2007). Une approche similaire a ainsi été appliquée pour l'analyse différentielle des flux entre la souche sauvage de *M. gaditana* et le mutant *acsbg* : la minimisation de la consommation de photons a été utilisée comme objectif de simulation, tout en fixant la production de biomasse. Cette stratégie a été sélectionnée afin de correspondre aux conditions de culture et aux données disponibles. Aucune donnée de consommation d'azote, de CO<sub>2</sub> ou de production d'O<sub>2</sub> n'était disponible. Par ailleurs, même si l'intensité lumineuse reçue par la culture de microalgues était connue (60 μmol<sub>photon</sub>.m<sup>-2</sup>.s<sup>-1</sup>) cette donnée n'est pas suffisante pour contraindre la consommation de photons au sein du modèle. En effet, il convient de convertir cette intensité lumineuse reçue par la culture (en μmol<sub>photon</sub>.m<sup>-2</sup>.s<sup>-1</sup>) en intensité lumineuse consommée par un gramme de matière sèche de microalgues (mmol<sub>photon</sub>.gDW<sup>-1</sup>.h<sup>-1</sup>.) Le calcul de la vitesse moyenne d'absorption des photons (MRPA, Mean Rate of Photon Absorption) peut être utilisé pour cette conversion.

$$\text{MRPA}_{exp} = \frac{a_s}{C_x} (q_0 - q_L) \quad (\text{Eq.2.1})$$

Avec  $a_s$  la surface spécifique du PBR, c'est-à-dire le rapport entre la surface éclairée et le volume du photobioréacteur (PBR) (en m<sup>-1</sup>),  $C_x$  la concentration en biomasse microalgale (en gDW.m<sup>-3</sup>),  $q_0$  et  $q_L$  l'intensité lumineuse en face avant et en face arrière du PBR (en μmol<sub>photon</sub>.m<sup>-2</sup>.s<sup>-1</sup>). Ainsi, le calcul du MRPA nécessite de connaître l'intensité lumineuse en face arrière de la culture (cette donnée n'est pas disponible), ainsi que la surface éclairée du photobioréacteur. Les multi-cultivateurs dans lesquels les microalgues ont été cultivées étant cylindriques, le MRPA est complexe à calculer.

Les réactions d'échanges du modèle ne pouvaient donc pas être contraintes avec des données expérimentales, empêchant ainsi l'utilisation de la maximisation de la fonction de biomasse comme objectif de simulation. Cependant, étant donné que nous disposions de données de croissance, le taux de croissance de chaque souche a pu être calculé et utilisé pour contraindre le flux de la fonction de biomasse. La minimisation de la consommation de photon a ensuite été utilisée comme objectif de simulation. Nous avons choisi de minimiser la consommation de photons (et non celle de CO<sub>2</sub> par exemple) car les microalgues ont été cultivées en lumière continue (éclairage à 60  $\mu\text{mol}_{\text{photon}}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ) dans un milieu riche en azote avec un bullage de gaz contenant 0,5 % de CO<sub>2</sub> (Billey et al. 2021). Dans ces conditions, les photons sont considérés comme le seul facteur limitant à la croissance des microalgues.

## 2.4.2 Standards actuels de développement et de publication de GEM

Le domaine de la modélisation métabolique à l'échelle du génome ne cesse de progresser, avec des améliorations apportées aux modèles existants, la génération de nouveaux modèles, ainsi que le développement de nouvelles méthodes d'analyse et de reconstruction de ces modèles (Gu et al. 2019; Ye et al. 2022; Passi et al. 2021). Cependant, plusieurs études ont révélé une grande disparité et un manque de standardisation en ce qui concerne la qualité des GEM et des outils utilisés pour leur reconstruction (Ravikrishnan et Raman 2015; Mendoza et al. 2019; Lieven et al. 2020).

Ainsi, la communauté de modélisation du métabolisme a entrepris un large effort pour se conformer aux principes FAIR (Findable, Accessible, Interoperable, Reusable). Ceci inclut la mise en place de normes pour garantir la transparence de la reconstruction des modèles métaboliques et de leur exploitation, la gestion des versions, la normalisation des formats de sauvegarde et de partage des GEM, ainsi que l'annotation et les tests des modèles. Dans cette optique, la suite de tests MEMOTE a été développée pour standardiser les tests qualités des GEM (Lieven et al. 2020). De plus, suite à un sondage réalisé en 2018 auprès de la communauté COBRA, des directives de reconstruction et de publication ont récemment été proposées (Carey et al. 2020). Enfin, un répertoire Git "standard-GEM" a été créé dans l'objectif de standardiser la mise à disposition des modèles dans un format reproductible et d'assurer le suivi des modifications apportées aux modèles avant et après leur publication (Anton et al. 2023).

La standardisation des GEM étant essentielle pour favoriser la collaboration scientifique, améliorer la comparabilité des modèles et faciliter leur intégration, ces directives ont été suivies au cours de cette thèse.

De ce fait, le suivi de la qualité du modèle a été assuré de manière constante tout au long de sa révision grâce aux tests MEMOTE. Un répertoire GitHub (<https://github.com/Total-RD/Mgaditana-GEM>) a été créé en respectant les normes proposées par Anton et al. 2023 et a été mis à disposition de la communauté scientifique. Parmi les différents standards suggérés pour la reconstruction d'un GEM, Carey et al. proposent des "check-list" destinées à être utilisées lors des phases de reconstruction et d'évaluation des GEM, dans le but d'améliorer l'accessibilité, le contenu et la qualité des réseaux métaboliques.

Les résultats de iMgadit23 pour ces check-lists sont présentés ci-dessous :

## Reconstruction

---

### Modèle

- Convention de dénomination :
  - Approche historique : i + initiales des auteurs + nombre de gènes dans le modèle
  - Approche recommandée : i + indicateur de l'espèce + indicateur d'itération **OUI**, le modèle reconstruit est intitulé iMgadi23 pour *M. gaditana*, soumis pour publication en 2023.
- Référence "machine lisible" de l'organisme intégrée via une annotation MIRIAM
  - nom complet de l'espèce, y compris les identifiants pertinents s'ils sont disponibles (par exemple, le génome de référence NCBI) **OUI**
  - identifiant de la taxonomie **OUI**
  - identifiant de la souche, si nécessaire **OUI**
  - type de tissu, si nécessaire **Non applicable**
  - URL pour obtenir le génome **OUI**
- Référence
  - DOI **OUI**, la publication du modèle est disponible sur bioRxiv (doi: <https://doi.org/10.1101/2023.12.06.570374>)
  - Nom et coordonnées de l'auteur (ou des auteurs) intégré(s) **OUI**
  - Espace de noms cohérent pour tous les identificateurs de modèles +/- Utilisation des BiGG ID et création d'une nomenclature respectant les spécifications de BiGG pour modéliser les lipides.

---

### Métabolite

- Nom descriptif et lisible par l'homme (ex. D-Glucose) **OUI**
- Charge (ex. 0) **OUI**
- Formule chimique (ex. C6H12O6) **OUI**
- Identifiants structurels **NON**
  - InChI strings (si le pH est connu, InChI ID en fonction du pH pour chaque métabolite)
  - SMILES (optionnel)
- Au moins un identifiant de base de données provenant d'une source fiable, telle que :
  - MetaNetX
  - BiGG
  - KEGG Compound
  - ChEBI
  - ModelSEED
  - HMDB
  - MetaCyc

67% des métabolites possèdent un identifiant dans au moins une base de données.

---

### Réaction biochimique

- Nom descriptif et lisible par l'homme (ex. phosphofructokinase) **OUI**
- Equation de la réaction (ex.  $\text{ATP} + \text{L-glutamate} + \text{ammonium} \rightleftharpoons \text{ADP} + \text{L-glutamine} + \text{H}^+ + \text{phosphate}$ ) **OUI**
- Au moins un identifiant de base de données provenant d'une source fiable, telle que :
  - MetaNetX

- Rhea
- BiGG
- KEGG Reaction
- ModelSEED
- MetaCyc
- EC Number

80% des réactions possèdent un identifiant dans au moins une base de données.

- Gène(s) associée(s) (gene-protein-reaction rule, or GPR): **OUI**
- Recommandé : termes de Systems Biology Ontology (SBO) **OUI**, toutes les réactions sont annotées avec un terme SBO.

### Gène

- Nom ou symbole du gène **NON**
- ID du gène ou de la protéine (ex. un mécanisme de mise en correspondance de ces informations avec une séquence)
  - Ensembl **OUI**, *Nannochloropsis gaditana* str. B-31 (GCA\_000569095.1) (NagaB31\_1.0) ([https://protists.ensembl.org/Nannochloropsis\\_gaditana\\_gca\\_000569095](https://protists.ensembl.org/Nannochloropsis_gaditana_gca_000569095))
  - UniProt **OUI**, 100% des gènes ont un identifiant Uniprot (excepté les pseudo-gènes "NO\_GENE" et "Spontaneous")
  - Autre identifiant de base de données spécifique au domaine **NON**
- Position (incluant le chromosome, si applicable) **NON**

## Publication

---

### Disponibilité +/-

- Le modèle est-il disponible publiquement ? **OUI**, le modèle est disponible sur GitHub (<https://github.com/Total-RD/Mgaditana-GEM>), selon le standard proposé par Anton et al. 2023.
    - Sur BioModels ? **NON** (*incompatibilité de licence*).
    - Ailleurs ? (optionnel) **OUI** sur GitHub
  - Est-ce que le modèle a été partagé sur une base de données ? **NON**, mais il est disponible dans un répertoire Git standardisé (Anton et al. 2023)
- 

### Format ✓

- Est-ce que le modèle est enregistré dans un format indépendant en matière de langage (ex. SBML) ? **OUI** (SBML3FBC)
  - Optionnel : formats additionnels (ex. XLS, JSON) **OUI** (XLSX, JSON et MAT)
- 

### Nomenclature +/-

- Est-ce que le nom du modèle indique une version ? **OUI**
  - Est-ce que le nom du modèle indique un organisme ? **OUI**
  - Les identifiants (gènes, métabolites et réactions) proviennent ils d'un unique espace de noms ?
    - Pour les gènes : **OUI**
    - Pour les réactions et les métabolites : +/- Utilisation des BiGG ID et création d'une nomenclature respectant les spécifications de BiGG pour modéliser le métabolisme lipidique.
- 

### Optimisation ✓

- Les tests MEMOTE ont-ils été réalisés ? **OUI**
  - La fonction objective est elle indiquée ? **OUI**
  - Des références sont-elles associées à cette fonction objective ? **OUI**
  - Les contraintes, réactions d'échange, de demande et de puit (sink) sont incluses par défaut ou dans le code ? **OUI** (le modèle contient les contraintes de culture en photo-autotrophie avec du nitrate, d'autres contraintes sont disponibles dans les différents codes de simulations).
  - La reconstruction peut-elle être instanciée sans erreur avec les outils COBRA ? **OUI**
- 

### Simulations ✓

- Les paramètres de simulation (objective reaction, constraints etc) sont-ils indiqués dans au moins un des formats suivants ?
  - README.md file ? **OUI**
  - COMBINE repository ? **NON**
  - Code complet ? ex. iPython notebook ou équivalent **OUI** (les scripts de simulation sont disponibles dans le répertoire GitHub)
- La version de COBRA utilisée est-elle indiquée ? **OUI**

- Les solveurs sont-ils documentés ? **OUI**
- 

#### Manuscrit ✓

- Les efforts déployés par COBRA en matière de logiciels sont-ils reconnus à leur juste valeur ? **OUI**
  - Les itérations précédentes et/ou les autres versions de la reconstruction sont-elles dûment mentionnées ? **Non applicable**
  - Le modèle est-il clairement référencé (c'est-à-dire avec un lien résoluble et un identifiant) dans la section "Matériels et Méthodes" ou dans la "Déclaration de disponibilité des données" ? **OUI**
- 

Les résultats en termes de nomenclature et d'annotation sont discutés plus en détails ci-dessous.

#### 2.4.2.1 Nomenclature

Lors de la reconstruction d'un GEM, il est conseillé d'utiliser des identifiants provenant d'un unique ensemble des noms, par exemple les identifiants d'une base de données telle que BIGG (Carey et al. 2020). Dans iMgadiT\_SilicoLife, la majorité des identifiants des métabolites et des réactions provenaient de la base de données BIGG. Certains ID avaient néanmoins dû être créés du fait de leur absence dans cette base de données. Cependant la nomenclature de BIGG n'est pas adaptée à la modélisation des lipides. En effet, les identifiants BiGG des lipides présentent plusieurs limites : (1) ils ne sont pas exhaustifs, de nombreux lipides ne possèdent pas d'identifiant BiGG, (2) ils ne sont pas homogènes et (3) ils sont difficilement compréhensibles même par les experts du métabolisme lipidique. Les limites des identifiants BiGG pour les lipides, notamment les acides gras et leurs dérivés ont récemment été discutés (Witting 2020). Ces limitations peuvent conduire à des erreurs de modélisation, telle que la duplication du même métabolite sous deux identifiants différents et donc considéré à tort comme deux entités chimiques différentes. C'était le cas par exemple pour le 3-hydroxyoctadecanoyl-CoA dans iMgadiT\_SilicoLife qui était dupliqué sous les identifiants 3hC18\_0\_CoA et 3hodcoa.

C'est la raison pour laquelle il a été décidé de mettre en place une nouvelle nomenclature pour la modélisation du métabolisme lipidique dans iMgadiT. Cette nomenclature se base sur la nomenclature abrégée chimique qui est largement répandue dans la communauté Lipide (IUPAC 1978; Liebisch et al. 2020). Afin d'assurer la cohérence avec le reste des identifiants du modèle, les standards et directives de BiGG quant aux nomenclatures ont été respectés. Cette nomenclature, spécifiant la classe du lipide ainsi que sa structure en termes de nombre d'atomes de carbone et d'insaturations, facilite grandement la lisibilité du modèle (Figure 2.4). Cette nomenclature a été appliquée aux 917 lipides du modèle ainsi qu'au

1017 réactions impliquées dans le métabolisme lipidique. Les noms de ces métabolites et de ces réactions ont également été modifiés afin de les rendre plus explicites.

Une initiative similaire a déjà été utilisée dans un modèle de *Caenorhabditis elegans* (Witting 2020). Cependant jusqu'à présent, aucune nomenclature n'a été largement adoptée par la communauté COBRA pour la modélisation des lipides. Cette absence de standardisation découle en partie de la diversité des nomenclatures pour les lipides, ainsi que de la disparité de détails et de qualité de représentation du métabolisme lipidique au sein des GEM.

#### 2.4.2.2 Annotations

Le métabolisme lipidique est peu annoté dans iMgadi23 faisant ainsi baisser ses scores d'annotations dans MEMOTE. La position des insaturations n'étant pas prise en compte dans le modèle, un métabolite lipidique dans iMgadi23 correspond en général à plusieurs lipides des bases de données. Les identifiants LIPID MAPS (Fahy et al. 2007; Liebisch et al. 2020) sont bien développés pour décrire les lipides, mais chaque identifiant est associé à un lipide avec des insaturations aux positions données, ce qui complexifie l'annotation des lipides dans iMgadi23. Plusieurs identifiants LIPID MAPS pourraient être ajoutés à un métabolite du modèle. Une perspective d'amélioration consisterait à ajouter les annotations ChEBI (Degtyarenko et al. 2008) aux lipides, car cette base de données propose de nombreux identifiants sans la position des insaturations. Cela permettrait également d'élargir les perspectives d'utilisation du modèle. En effet, récemment un outil spécifiquement conçu pour les lipides a été développé afin de faciliter la comparaison de modèles métaboliques entre eux, et l'exploitation des GEM pour l'analyse de données omiques (notamment de données lipidomiques) (Poupin et al. 2020). L'utilisation de cet outil nécessite la présence des identifiants ChEBI dans les annotations du modèle.

### 2.4.3 Perspectives d'amélioration d'iMgadi23

#### 2.4.3.1 Prise en compte des positions des insaturations des acides gras

D'un point de vue chronologique, la correction des voies de synthèse des lipides a été effectuée avant la correction de la  $\beta$ -oxydation. Initialement dans un souci de simplification, il avait été décidé de ne pas prendre en compte les positions des insaturations dans le modèle. Cependant l'ajout des réactions de la  $\beta$ -oxydation des acides gras polyinsaturés a nécessité la spécification de la position de ces insaturations. La position de ces insaturations est en effet nécessaire pour comprendre et vérifier les étapes de la dégradation de ces acides gras, pour permettre l'ajout précis des enzymes etc. La position des

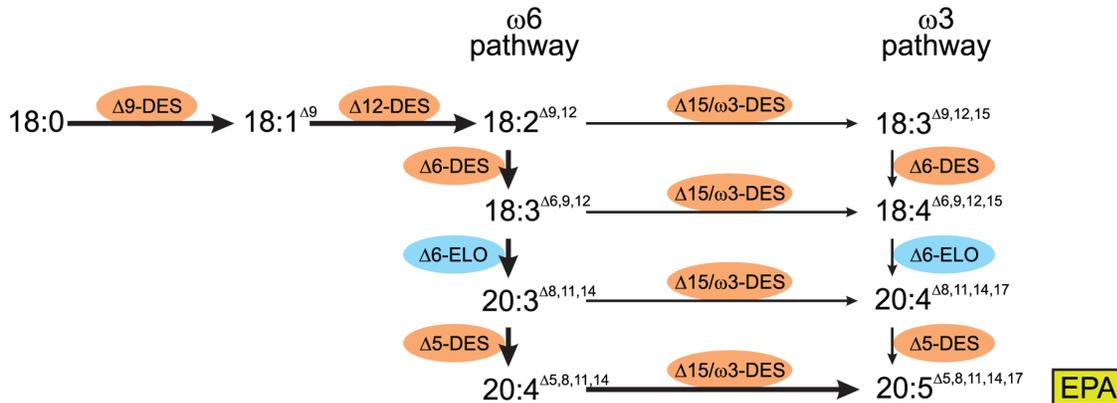
insaturations a donc été précisée dans les identifiants des réactions et des métabolites de cette voie. Ceci crée une disparité de détails et de nomenclature au sein du modèle. Pour rétablir la cohérence au sein du modèle, deux solutions sont envisageables :

- (1) Le contenu réactionnel du modèle peut être maintenu tel qu'il est, toutefois, la position des insaturations pourrait être retirée des identifiants des métabolites et des réactions de la  $\beta$ -oxydation. Afin de ne pas perdre d'information les détails concernant les insaturations pourraient être intégrés dans les annotations et/ou les notes des réactions et des métabolites.
- (2) La position des insaturations peut également être entièrement intégrée dans le métabolisme lipidique. Cette solution constitue une piste d'amélioration du modèle qui est discutée plus en détails ci-dessous.

L'une des perspectives d'amélioration du modèle consisterait à prendre en considération la position des insaturations dans l'ensemble du métabolisme lipidique. Cette amélioration serait particulièrement pertinente pour les acides gras polyinsaturés à très longue chaîne (VLC-PUFA). En effet, chez les microalgues, les VLC-PUFA sont synthétisés par la "voie oméga", conduisant à la production de VLC-PUFA  $\omega 3$  et  $\omega 6$ . Il serait intéressant de subdiviser la voie de synthèse de ces acides gras, comme elle est actuellement décrite dans iMgadi23, afin de distinguer les voies  $\omega 3$  et  $\omega 6$  (Figure 2.5). Cependant cette distinction nécessiterait de nombreuses modifications au sein du modèle. En effet, les divers acides gras  $\omega 3$  et  $\omega 6$  devraient ensuite être intégrés à la synthèse des glycérolipides. Ceci entraînerait une augmentation significative du nombre de lipides spécifiques pouvant être produits, ainsi qu'une augmentation concomitante du nombre de réactions nécessaires à leur synthèse. Par exemple, dans le modèle la synthèse de DAG à partir d'acide phosphatidique par la Phosphatidate Phosphatase (réaction PAP) est actuellement décrite par 32 réactions dans le RE. La prise en compte de la distinction entre  $\omega 3$  et  $\omega 6$  nécessiterait l'ajout de 23 réactions supplémentaires, portant ainsi le total à 55 réactions PAP dans le RE. Il serait également nécessaire d'ajouter la dégradation complète des acides gras  $\omega 6$  dans le peroxysoxe et la mitochondrie, étant donné que seule la dégradation des acides gras  $\omega 3$  est actuellement décrite dans le modèle.

De plus, de nombreuses données lipidomiques ne fournissent pas d'informations sur la position des insaturations, y compris les données générées par la plateforme lipidomique du LPCV. L'exploitation de ces données nécessiterait donc l'ajout d'un grand nombre de réactions d'équivalence au sein du modèle. Ainsi, l'intégration de ces changements augmenterait considérablement non seulement le nombre de réactions décrivant le métabolisme lipidique, mais également la variabilité de ce mécanisme. Cependant,

cette extension serait réellement pertinente pour améliorer notre compréhension du métabolisme lipidique, en particulier pour explorer les voies  $\omega 3$  et  $\omega 6$ , offrant ainsi de précieuses perspectives de recherche.



**Figure 2.5. Voies proposées pour la synthèse des VLC-PUFAs chez les microalgues des genres *Nannochloropsis* et *Microchloropsis* (source : Vieler et al. 2012).**

#### 2.4.3.2 Ajout de nouvelles fonctions de biomasse

iMgadi23 possède deux fonctions de biomasse (souche sauvage et souche mutante *acsbg*) dont la composition lipidique repose sur des glycérolipidomes générés au LPCV. Des glycérolipidomes de différentes souches sauvages et mutantes de *M. gaditana* cultivées dans différentes conditions sont également disponibles. De plus, de nouvelles données sont générées régulièrement.

Une des perspectives d'application d'iMgadi consiste à intégrer ces profils lipidomiques au sein de nouvelles fonctions de biomasse afin de pouvoir réaliser des analyses spécifiques de différents contextes. BioModTool, un outil bioinformatique permettant d'automatiser les calculs des coefficients stœchiométriques et la génération de nouvelles fonctions de biomasse a ainsi été développé au cours de la thèse pour répondre à cette problématique.

## 3 Chapitre 3

---

# BioModTool

## 3.1 Introduction

### 3.1.1 Préambule

Dans les GEM, la fonction objective de biomasse (ou fonction de biomasse) consomme l'ensemble des métabolites nécessaires pour produire un gramme de matière sèche (gDW), permettant ainsi de simuler la croissance de l'organisme modélisé. iMgadi23 comporte deux fonctions de biomasse représentant la composition de la biomasse d'une souche sauvage et d'une souche mutante (*acsbg*) de *M. gaditana* cultivées en conditions photoautotrophes. Cependant, la composition de la biomasse de *M. gaditana*, tout comme celle des microorganismes en général, est fortement influencée par les conditions environnementales. Ainsi, plusieurs paramètres de culture tels que (1) la lumière (intensités lumineuses et spectres lumineux) (Kim et al. 2014; Janssen, Driessen et al. 2018b), (2) la température (Ferrer-Ledo et al. 2023), (3) les carences en nutriments (azote, phosphate) (Simionato et al. 2013; Janssen, Driessen et al. 2018a; Matsui et al. 2020; Savvidou et al. 2020), (4) les modes de cultures en hétérotrophie (Marudhupandi, Sathishkumar et Kumar 2016) et en mixotrophie (Fang et al. 2004), ont une influence significative sur la composition de la biomasse des espèces *Nannochloropsis* et *Microchloropsis*. Inclure la variabilité de la composition de la biomasse dans le modèle permettrait d'analyser l'impact de conditions environnementales ou génétiques sur le métabolisme de *M. gaditana*.

Par ailleurs, le LPCV possède une plateforme de lipidomique spécialisée dans l'analyse des glycérolipides. Des glycérolipidomes de *M. gaditana* WT ainsi que de différentes souches mutantes cultivées dans diverses conditions sont disponibles. De nouvelles données sont également fréquemment générées. L'interprétation de ces données lipidomiques peut s'avérer complexe et requiert une expertise du métabolisme des lipides de *M. gaditana*. L'intégration de ces données lipidomiques sous forme de fonctions de biomasse dans le modèle iMgadi23 constituerait un atout significatif pour leur analyse.

L'intégration de ces données requiert la définition de nouvelles fonctions de biomasse, ce qui implique plusieurs étapes de calculs et de modélisation. En effet, une contrainte de la fonction de biomasse est

qu'elle doit représenter la consommation des métabolites nécessaire à la production d'un gramme de matière sèche. Les données de composition de biomasse doivent donc être converties et normalisées afin de pouvoir respecter cette contrainte. La conversion de données de composition de biomasse peut s'avérer complexe et nécessite souvent des approximations (Tollete et al. 2023). L'automatisation de ce processus à l'aide d'un outil informatique serait particulièrement bénéfique.

L'utilisation de logiciels dédiés à la création des fonctions de biomasse est pertinente non seulement pour iMgadi23, mais aussi pour l'ensemble des GEM. En effet, une étape cruciale dans la reconstruction des GEM est la définition de la fonction de biomasse (Feist and Palsson 2010). Plusieurs études ont démontré l'impact significatif de la composition de la biomasse sur les prédictions des modèles (Pramanik et Keasling 1998; Nookaew et al. 2008; Lavoie et al. 2020; Schulz et al. 2021). La définition de réactions de biomasse de qualité, adaptées aux conditions environnementales ou génétiques, est donc nécessaire afin d'assurer la fiabilité des prédictions des GEM. Malgré leur importance, il existe un manque d'outils informatiques permettant d'intégrer des réactions de biomasse dans les modèles métaboliques. BOFdat, un package Python automatisant la création de fonctions de biomasse à partir de différents ensembles de données, notamment génomiques, transcriptomiques, protéomiques et lipidomiques, a récemment été développé (Lachance et al. 2019). Cependant, ces données omiques ne sont pas toujours disponibles, ce qui limite le nombre d'organismes et de conditions dans lesquels BOFdat peut être appliqué.

Ce chapitre présente BioModTool (Biomass Modeling Tool), un nouveau package Python dédié à la création de fonctions de biomasse pour les modèles métaboliques. BioModTool est accompagné de fichiers Excel types que l'utilisateur complète avec des données de composition de biomasse. Dans un premier temps, BioModTool charge les données et le GEM à modifier. Il calcule ensuite les coefficients stœchiométriques, ce qui implique des étapes de normalisation et de conversion d'unités. Ensuite, des pseudo-réactions et des pseudo-métabolites nécessaires à la définition de la fonction de biomasse sont instanciés et intégrés au modèle métabolique d'intérêt. Un module spécifique, nommé `lipidomic_data_lpcv.py` a également été conçu pour automatiser l'intégration des données lipidomiques provenant du LPCV.

### 3.1.2 Cycle de développement de BioModTool

Le développement de BioModTool a été réalisé en suivant les différentes phases du cycle en V.

1. Analyse des besoins des utilisateurs et définition des spécifications fonctionnelles.
2. Conception : proposition de solutions techniques.
3. Réalisation : écriture des programmes.
4. Validation : réalisation des tests unitaires et fonctionnels.

Ce chapitre détaille les différentes étapes de ce développement.

## 3.2 Spécifications des besoins

La création d'une fonction de biomasse nécessite plusieurs étapes.

**Etape 1 :** La recherche de données de composition de biomasse, ou la génération de données en suivant des protocoles adaptés.

**Etape 2 :** La normalisation des données et leur conversion en unité compatible avec les modèles métaboliques.

**Etape 3 :** La création de la fonction de biomasse : instanciation (c'est-à-dire la création d'une nouvelle instance de la classe, ici "Reaction") et ajout au modèle de la ou des réactions composant la fonction de biomasse.

**Etape 4 :** La validation de la fonction de biomasse.

Les étapes 2 et 3 sont itératives et susceptibles de générer des erreurs. L'objectif de BioModTool est d'automatiser ces deux étapes afin de faciliter la génération de fonctions de biomasse.

La première phase du développement de BioModTool a consisté à effectuer une spécification des besoins.

Cette étape avait pour objectif principal:

1. D'identifier les cas d'utilisation et les utilisateurs visés par BioModTool (y compris par le module `lipidomic_data_lpcv.py`).
2. De définir la nature des données manipulées par BioModTool.
3. D'établir la structure de la fonction de biomasse.

### 3.2.1 Identification des cas d'utilisation et des utilisateurs ciblés

L'objectif de BioModTool est de permettre à un utilisateur disposant des données de composition de biomasse de générer et d'ajouter une fonction de biomasse dans un modèle GEM préexistant. BioModTool doit réaliser les fonctions suivantes: normaliser les données de l'utilisateur, les convertir dans des unités appropriées, instancier les pseudo-réactions et les éventuels pseudo-métabolites constituant la fonction de biomasse, puis les intégrer au modèle métabolique fourni par l'utilisateur. De plus, BioModTool doit avoir la capacité de calculer la formule chimique et la charge des pseudo-métabolites créés, de manière à garantir le respect des contraintes de conservation des charges et des masses des GEM. BioModTool doit aussi permettre de supprimer une fonction de biomasse au format BioModTool d'un GEM. Toutes ces fonctionnalités doivent pouvoir se faire de façon **automatique**.

BioModTool doit être suffisamment **facile d'utilisation** pour être utilisé par deux catégories d'utilisateurs : ceux qui sont à l'aise avec la manipulation de lignes de code et ceux qui ne le sont pas.

### 3.2.2 Applicabilité à différents modèles métaboliques

#### 3.2.2.1 Identifiants de métabolites

BioModTool doit être en mesure de modifier le modèle iMgadi23 ainsi que de nombreux GEM disponibles dans la littérature. Il n'existe pas de nomenclature universellement adoptée pour les identifiants des métabolites au sein des GEM. Par conséquent, les **identifiants des métabolites consommés** dans la fonction de biomasse doivent être facilement **modifiables**.

#### 3.2.2.2 Compartiments

Enfin, bien que la fonction de biomasse soit couramment ajoutée dans le cytosol des GEM (souvent identifié par l'ajout du suffixe "\_c" aux identifiants des métabolites et des réactions), cette convention n'est pas universellement suivie. L'utilisateur peut souhaiter intégrer la fonction de biomasse dans un autre compartiment ou le GEM sélectionné peut ne pas disposer de cytosol. C'est par exemple le cas du modèle d'*Alicyclobacillus acidocaldarius* qui se compose uniquement de deux compartiments : un compartiment interne (Internal\_Species) et un compartiment externe (External\_Species) (Beck, Hunt et Carlson 2018). Le **choix du compartiment** dans lequel la fonction de biomasse sera ajoutée ainsi que sa dénomination doivent donc être laissés à l'utilisateur.

### 3.2.3 Détermination des données de composition de biomasse exploitables

Les données de composition de biomasse présentent une grande diversité en termes de provenance et d'unités utilisées pour les décrire.

#### 3.2.3.1 Provenance et format des données

De nombreuses données de composition de biomasse sont disponibles dans la littérature. De plus, plusieurs protocoles d'analyse de la composition de la biomasse ont été développés pour la reconstruction des GEM (Thiele et Palsson 2010; Beck, Hunt et Carlson 2018; Simensen et al. 2022). Cependant, la disponibilité des données peut considérablement varier en fonction de l'organisme étudié ou de ses conditions de culture.

BioModTool doit offrir une **flexibilité** quant à l'**origine et au format des données** exploitables.

#### 3.2.3.2 Unités

Il existe une réelle diversité d'unités pour exprimer l'abondance d'un métabolite dans un organisme (Tollete et al. 2023). Ainsi, différentes unités de concentration massique et molaire sont fréquemment utilisées pour caractériser la biomasse (Table 3.1).

**Table 3.1. Exemples d'unités de concentration massique et molaire utilisées pour décrire la composition de la biomasse.**

Concentration massique	Concentration molaire
- Grammes par litre	- Moles par litre
- Grammes par cellule	- Moles par cellule
- Grammes par gramme de matière sèche	- Moles par gramme de matière sèche
- Pourcentage ou fraction massique	- Pourcentage ou fraction molaire
- Grammes par grammes de protéine (pour les acides aminés)	- Moles par moles d'ADN (pour les dNTPs)
- Grammes par grammes de lipides totaux ou grammes par grammes de TAG (pour les TAG)	- Moles par moles de lipides totaux ou moles par moles de TAG (pour les TAG)
- Etc.	- Etc.

Afin d'élargir les possibilités d'utilisation de BioModTool, il est nécessaire que les **unités** acceptées soient **flexibles**.

### 3.2.4 Structure et composition de la fonction de biomasse

Comme présenté en introduction, il est possible de modéliser une fonction de biomasse en utilisant une réaction unique consommant l'ensemble des métabolites dans les proportions appropriées. Cependant, il est courant de représenter la fonction de biomasse par plusieurs pseudo-réactions organisées en deux ou trois niveaux. Ainsi, le premier niveau consomme les macromolécules principales de la cellule, telles que l'ADN, l'ARN, les protéines, les glucides, les lipides etc. Le second niveau correspond à la consommation des monomères (les dNTPs, les acides aminés etc.), symbolisant la polymérisation des macromolécules. Un troisième niveau peut également être ajouté pour détailler le profil de différentes classes lipidiques. La **structure de la fonction de biomasse** générée par BioModTool doit être **flexible** pour s'adapter aux données disponibles, à l'organisme ciblé, ainsi qu'aux préférences de l'utilisateur. De plus, BioModTool doit pouvoir s'adapter à des GEM de différents organismes, la **composition** de la biomasse en termes de macromolécules et de monomères consommés doit être **ajustable** en fonction des besoins.

La composition de la fonction de biomasse dépend également de la connaissance de la composition cellulaire et des besoins énergétiques. La composition des fonctions de biomasse peut être séparée en trois niveaux de détails: (1) niveau basique, (2) intermédiaire et (3) avancé (Feist et Palsson 2010). Initialement, les GEM ont tendance à avoir des fonctions de niveau basique dans lesquelles seules les principales macromolécules (ARN, ADN, protéines et lipides) sont consommées. Ces formulations sont ensuite affinées et élargies en incluant la consommation en énergie de la polymérisation des acides nucléiques et des protéines (niveau intermédiaire). Enfin, des métabolites spécifiques à l'organisme d'intérêt tels que des vitamines, des cofacteurs, des ions inorganiques et des composants membranaires peuvent également être ajoutés (niveau avancé). BioModTool doit pouvoir être **compatible** avec ces **différents niveaux de détails**.

### 3.2.5 Besoins spécifiques des utilisateurs du LPCV : exploitation des données lipidomiques de *M. gaditana*.

La plateforme lipidomique du LPCV produit des données au format standardisé pour divers organismes, y compris *M. gaditana*. L'outil développé doit être compatible avec le format des données lipidomiques du LPCV, permettant ainsi de les intégrer de manière automatique au modèle iMgadi23. Ce processus implique **la conversion des données lipidomiques du format LPCV au format requis par le modèle** et leur incorporation aux autres données composition de biomasse. Ces données doivent être par la suite utilisables par BioModTool afin d'intégrer de nouvelles fonctions de biomasse à iMgadi23.

## 3.3 Solutions techniques proposées

### 3.3.1 Implémentation de BioModTool.

BioModTool est disponible sur GitHub : <https://github.com/Total-RD/BioModTool>

#### 3.3.1.1 Choix du langage : Python

BioModTool implique la manipulation de GEM. Les bibliothèques COBRA, disponible en MATLAB (COBRA Toolbox) et en Python (COBRAPy), sont largement utilisées par la communauté de la modélisation métabolique (Ebrahim et al. 2013; Heirendt et al. 2019). Le choix du langage Python pour le développement de BioModTool s'est basé sur la disponibilité gratuite et open source de nombreuses bibliothèques.

#### 3.3.1.2 Développement de Jupyter Notebooks et d'une interface graphique

Des Jupyter Notebooks (Kluyver et al. 2016) ont été développés afin de faciliter l'utilisation de BioModTool. Une interface graphique, codée en Tkinter et accessible via un exécutable, est mise à disposition pour permettre l'utilisation de BioModTool sans aucune ligne de code.

### 3.3.1.3 Organisation du code

L'outil BioModTool est organisé en différents modules, regroupant les fonctions par fonctionnalités (Table 3.2). Un module spécifique, `lipidomic_data_LPCV.py`, compatible avec BioModTool, a été développé afin d'élargir ses fonctionnalités pour répondre aux besoins particuliers des utilisateurs du LPCV. Ce module n'est pas présent dans la version "tout public" de BioModTool. Une version interne (BioModTool-LPCV) est disponible.

**Table 3.2. Organisation du package BioModTool.**

Modules	Fonctionnalités
<code>load.py</code>	Charger les entrées utilisateur.
<code>calculate.py</code>	Effectuer les différentes étapes de calculs des coefficients stœchiométriques.
<code>update.py</code>	Instancier la fonction de biomasse (différents pseudo-métabolites et pseudo-réactions) et l'ajouter au GEM.
<code>test.py</code>	Effectuer les différents tests nécessaires à l'utilisation de BioModTool (tests des entrées utilisateurs, des réactions créées etc.). Effectuer également le test de validation de la fonction de biomasse : test du respect de la loi de conservation des masses et des charges par l'ensemble des réactions de la fonction de biomasse.
<code>save.py</code>	Sauvegarder les calculs réalisés et le GEM mis à jour.
<code>main_add_biomass_objective_function.py</code>	Organiser et structurer le package. Contient la fonction principale à partir de laquelle BioModTool démarre son exécution.
<code>remove.py</code>	Supprimer une fonction de biomasse au format BioModTool d'un GEM.
<code>interface_BioModTool.py</code>	Interface de BioModTool.

### 3.3.2 Format des données de composition de biomasse et structure de la fonction de biomasse

Pour satisfaire les exigences de flexibilité des données, des GEM et de la structure de la fonction de biomasse, il a été décidé de créer un fichier Excel structuré dans lequel l'utilisateur saisit manuellement ses données.

#### 3.3.2.1 Flexibilité de la structure de la fonction de biomasse

Le fichier Excel est organisé en différentes feuilles de calcul. Chaque feuille de calcul (excepté le Read\_Me) correspond à une macromolécule (ou à un pool de molécules) et donne lieu à la création d'une pseudo-réaction. La composition du fichier Excel peut être adaptée pour correspondre à la structure de la fonction de biomasse. Une feuille de calcul type est disponible pour chacun des niveaux d'organisation de la fonction de biomasse (niveau 1, niveau 2, niveau 2 – lipides, niveau 3). Pour chaque niveau, deux tables sont disponibles afin de respectivement inclure les coefficients des métabolites "basiques" (Table 1) et ceux des métabolites "intermédiaires" et "avancés" (Table 2).

#### 3.3.2.2 Flexibilité des unités

Le choix de l'unité des métabolites basiques se fait dans une liste déroulante. Seules deux unités sont proposées à l'utilisateur "g per ..." et "mol per ...", les données étant par la suite normalisées par BioModTool, ces deux groupes d'unités permettent de regrouper l'ensemble des unités de concentrations massiques et molaires. Au sein d'une même feuille de calcul les unités sont contraintes à être identiques mais des unités différentes peuvent être utilisées pour les différentes feuilles de calculs.

#### 3.3.2.3 Flexibilité des GEM modifiables

Afin de faciliter son utilisation, BioModTool comporte une fonction permettant de charger un GEM au format SBML, format de référence des GEM (Carey et al. 2020; Lieven et al. 2020) ou au format JSON. Le format JSON est un format également largement utilisé par la communauté scientifique qui est notamment proposé (lecture et écriture de modèles) par les bibliothèques COBRA (Ebrahim et al. 2013; Heirendt et al. 2019). De même, une fonction permettant de sauver le modèle mis à jour, dans ces deux formats, est disponible. Cependant, la fonction principale de BioModTool, permettant la création et l'intégration de la nouvelle fonction de biomasse, requiert en entrée un modèle métabolique (objet de

type cobra\_model) plutôt qu'un simple chemin d'accès au modèle. Ceci permet de maximiser la compatibilité de BioModTool avec COBRAPy. Un utilisateur maîtrisant COBRAPy peut ainsi directement utiliser cette librairie pour charger et sauver un modèle. Cela offre la flexibilité de travailler avec d'autres formats de modèles, tel que le format MAT par exemple.

Les fonctions de manipulation des GEM reposent sur les fonctions de la librairie COBRAPy qui sont compatibles avec la grande majorité des GEM. Les identifiants des métabolites doivent être renseignés par l'utilisateur en fonction du modèle choisi.

#### 3.3.2.4 Limitation des sources d'erreurs pour l'utilisateur

Les interactions des utilisateurs avec les outils informatiques ne sont pas simples à développer et sont très souvent source d'erreurs. Afin d'essayer de minimiser les erreurs pouvant être commises par les utilisateurs, la structure du fichier Excel a été simplifiée au maximum. Les cellules destinées à recevoir les coefficients sont soumises à des critères de validation. Ainsi, si l'utilisateur ne renseigne pas une valeur numérique, un message d'erreur est affiché. De plus, seules les cellules qu'il peut ou doit modifier (identifiables par un code couleur) ne sont pas verrouillées.

Des instructions pour l'utilisateur sont disponibles dans le Read\_Me et pour chaque feuille de calcul à compléter (lignes 1 à 7 de chaque feuille).

Une fois les spécifications établies et les solutions techniques définies, l'implémentation de BioModTool a été réalisée.

## 3.4 BioModTool – version tout public

### 3.4.1 Présentation de BioModTool

**Article 2** : BioModTool: from biomass composition data to structured biomass objective functions for genome scale metabolic models.

Clémence Dupont-Thibert<sup>1,2</sup>, Sylvaine Roy<sup>1</sup>, Gilles Curien<sup>1</sup>, Maxime Durot<sup>2</sup>.

1 - Université Grenoble Alpes, CNRS, CEA, INRAE, Interdisciplinary Research Institute of Grenoble, IRIG-Laboratoire de Physiologie Cellulaire et Végétale, 17 Avenue des Martyrs, 38000 Grenoble, France

2 -TotalEnergies, OneTech, Centre de Recherche de Solaize CRES, Chemin du Canal, 69360 Solaize, France

**N.B.1** : *Cet article est en finalisation de rédaction.*

**N.B.2** : *La numérotation des figures et des tables est indépendante du manuscrit de thèse.*

**N.B.3** : *Les textes supplémentaires présentant notamment l'implémentation, l'utilisation et la validation de BioModTool ont été intégrés dans ce chapitre (Sections [3.4.2](#) et [3.4.3](#)).*

**N.B.4** : *BioModTool est disponible sur GitHub (<https://github.com/Total-RD/BioModTool>)*

# BioModTool: from biomass composition data to structured biomass objective functions for genome scale metabolic models.

DUPONT THIBERT Clémence<sup>1,2</sup>, ROY Sylvaine<sup>1</sup>, CURIEN Gilles<sup>1</sup>, DUROT Maxime<sup>2</sup>.

<sup>1</sup> Université Grenoble Alpes, CNRS, CEA, INRAE, Interdisciplinary Research Institute of Grenoble, IRIG-Laboratoire de Physiologie Cellulaire et Végétale, 17 Avenue des Martyrs, 38000 Grenoble, France

<sup>2</sup> TotalEnergies, OneTech, Centre de Recherche de Solaize CRES, Chemin du Canal, 69360 Solaize, France

## Abstract

**Summary:** BioModTool is a Python program allowing easy generation of biomass objective functions for genome scale metabolic models from user data. BioModTool loads biomass composition data in the form of a structured Excel file previously completed by the user, normalizes, converts these data into model compatible unit, and creates a structured biomass objective to update a metabolic model. BioModTool can be run as Python command-lines but it also comes with an interface allowing its use by non-modelers. By providing easy definition of new biomass objective functions, BioModTool has the potential to support new genome scale metabolic reconstructions, improve existing ones, and to facilitate biomass-specific experimental datasets analyses with genome-scale models.

**Availability and implementation:** BioModTool is publicly available on GitHub (<https://github.com/Total-RD/BioModTool>) under a GNU Lesser General Public License (LGPL) license. BioModTool is compatible with Windows, Linux and MacOS operating systems.

**Contact:** [maxime.durot@totalenergies.com](mailto:maxime.durot@totalenergies.com)

**Supplementary information:** Supplementary texts and data are available at Bioinformatics online.

## 1. Introduction

A genome-scale metabolic model (GEM) is a computational network representing all known enzymatic and spontaneous metabolic reactions and metabolic genes of a given organism. A crucial step in GEM reconstruction is the definition of biomass reaction(s), commonly referred to as Biomass Objective Function (BOF) (Feist et Palsson 2010). A BOF consumes all metabolites required to produce one gram of dry-weight (gDW) biomass and is widely used as simulation objective using Constraint-Based Reconstruction and Analysis (COBRA) methods such as Flux Balance Analysis (FBA) (Ebrahim et al. 2013; O'Brien, Monk et Palsson 2015; Gu et al. 2019; Heirendt et al. 2019). GEMs usually assume a constant biomass composition, while *in vivo*, biomass composition is adjusted in response to changing environmental conditions: nutrient availability, temperature, stress etc. (Schaechter, Maaløe et Kjeldgaard 1958; Cotner, Makino et Biddanda 2006; Zuñiga et al. 2018). Moreover, numerous studies have demonstrated the high impact of biomass composition on models behaviors and predictions (Pramanik et Keasling 1998; Nookaew et al. 2008; Lavoie et al. 2020; Schulz et al. 2021). High-quality, condition dependent BOF are therefore necessary to enable reliable predictions in GEM.

The need of accurate characterization of the biomass composition has been addressed over the years by experimental protocols (Feist et Palsson 2010; Beck, Hunt et Carlson 2018; Simensen et al. 2022). Converting experimental data into BOF stoichiometric coefficients is fastidious and requires several normalization and unit conversion steps which can be source of errors. Despite importance of BOF in GEMs, there is a lack of tools to convert the experimental data into functional BOFs. Recently, BOFdat, a Python workflow computing BOF from different datasets including genomic, transcriptomic, proteomic and lipidomic was developed (Lachance et al. 2019). However, omics data are not always available, therefore limiting the number of organisms and conditions in which BOFdat can be applied.

Here we present BioModTool ([Biomass Modeling Tool](#)), a new Python package to generate BOFs in GEMs from a single editable structured dataset. BioModTool is provided with template Excel files (Supplemental Files S1-S3) to be filled by the user with biomass composition data that can be found in the literature, determined using appropriate protocols or even determined using appropriate BOFdat function when a specific omic dataset is available. First, BioModTool calculates stoichiometric coefficients of all metabolites to be consumed in BOF, involving normalization and unit conversion steps. Pseudo-reactions and pseudo-metabolites required for BOF definition are then created and added to the GEM to update.

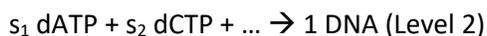
## 2. BioModTool usage

### 2.1 BOF structure

Before running BioModTool, user must define BOF structure. BOF are commonly organized into different levels, each level being described by appropriate pseudo-reaction(s). Up to three levels BOFs can be defined by BioModTool (Figure 1). The first level consumes the major macromolecules of the cell such as DNA, RNA, proteins, carbohydrates, lipids etc., and is typically defined as follows:



where  $s_i$  are the calculated macromolecules abundance in cell biomass (in  $\text{mmol.gDW}^{-1}$ ) and DNA, RNA, PROTEINS etc. are pseudo-metabolites resulting from the consumption of appropriate monomers (such as deoxynucleotide triphosphates (dNTP), amino acids etc.). For example, DNA is a pseudo-metabolite produced by a pseudo-reaction consuming the four dNTPs:

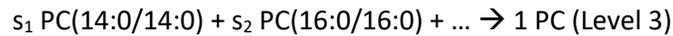


where  $s_i$  is the molar fraction of each metabolite in the macromolecule (here in  $\text{mol.mol}_{\text{DNA}}^{-1}$ ).

Due to diversity of existing lipid entities, lipid pseudo-reactions are usually more challenging to formulate. One solution consists in introducing additional pseudo-reactions and pseudo-metabolites to gather lipid metabolites by class. This is possible with BioModTool by introducing a third level in the BOF. Lipid class consideration requires definition of lipid pseudo-reaction as follows:



where  $s_i$  is the molar fraction of the lipid class in total lipids ( $\text{mol}_{\text{class}} \cdot \text{mol}_{\text{LIPIDS}}^{-1}$ ) and PC (phosphatidylcholine), PE (phosphatidylethanolamine) etc. are pseudo-metabolites representing each lipid class described to be consumed in the BOF. Lipid class consideration can consequently result in the addition of a third level of pseudo-reactions, consuming specific lipids. For example, for PC:



where  $s_i$  is the molar fraction of each specific lipid in its given class (here in  $\text{mol} \cdot \text{mol}_{\text{PC}}^{-1}$ ).

Levels 2 and 3 are optional, a single level BOF can be generated by BioModTool. The aforementioned macromolecules are just examples. Additional macromolecules, such as pigments for photosynthetic organisms, can be added at the user convenience. The same applies to lipid classes. For each pseudo-reaction, maintenance costs of growth or pseudo-metabolites synthesis, requirement of necessary vitamins, elements, and cofactors etc. can also be included in BOF generated by BioModTool.

## 2.2 Biomass composition

User must update and complete one of the provided Excel files (single level BOF, two levels BOF, three levels BOF [Supplemental Files S1-S3](#)) with biomass composition data and metabolite identifiers of the chosen GEM according to the desired BOF structure. Excel file is divided into sheets, each sheet corresponding to a pseudo-reaction to be created (e.g., BIOMASS, DNA, LIPIDS, PC etc.). Each sheet contains two distinct tables. First table (columns A to E) contains metabolites of basic level defining the macromolecular or monomer content in the cell. The units used can vary between the pseudo-reactions to be created (one Excel sheet corresponding to one pseudo-reaction), but the metabolites consumed in a given pseudo-reaction must be given in the same unit. For example, DNA and PROTEIN composition can be given in different units, but all amino-acids abundance must be in the same unit (idem for dNTPs). These data will be normalized so that the total mass recovered from all data corresponds to one gram of dry weight before being converted into GEM compatible units.

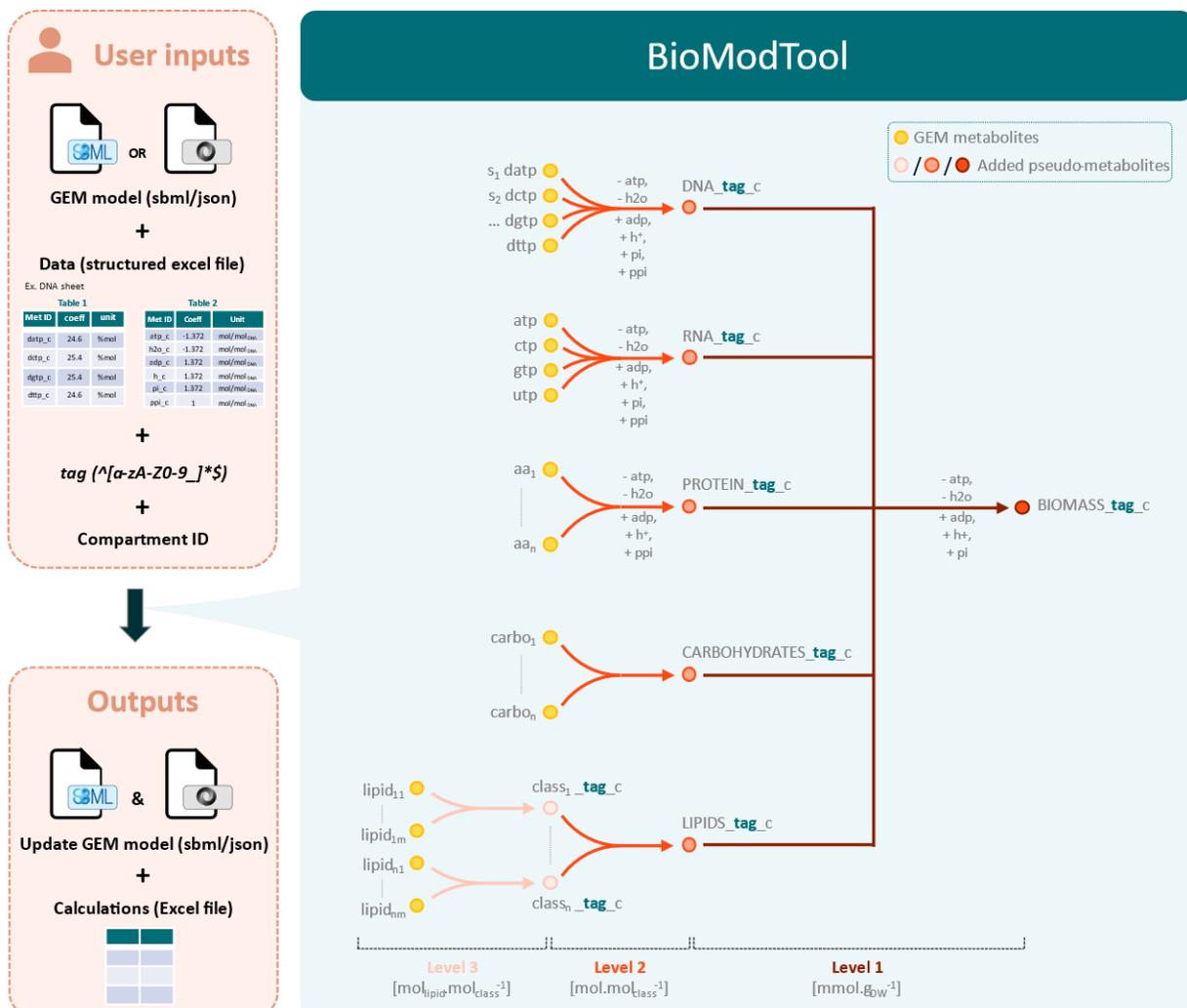
BOF can be formulated at a different level of details (basic, intermediate and advanced) (Feist et Palsson 2010). A second table (columns H to L) can be completed by the user to improve BOF level of details by adding energy requirement for the macromolecule synthesis (intermediate level) and vitamins, elements, and cofactors required for growth (advanced level). Data table 2 will not be converted and will directly be used as stoichiometric coefficients, therefore they must be given in  $\text{mmol} \cdot \text{gDW}^{-1}$  or  $\text{mol} \cdot \text{mol}_{\text{pool}}^{-1}$ . In BioModTool all polymers are considered to be one monomer long, stoichiometric coefficients of energy and byproducts requirements must be given on the basis of these polymers lengths.

## 2.3 Run BioModTool

BioModTool can be run either with Python command-lines or using the user interface (implemented with Tkinter). In both cases, user must provide several inputs (Figure 1).

- (1) A GEM in SBML or JSON format. According to the chosen model, user must indicate if charge and/or formula are available for all metabolites consumed in the BOF and must define in which compartment the BOF will be added (among compartments available in model).
- (2) The Excel file as described above, filled with biomass composition data and metabolites identifiers. Metabolites identifiers used must comply with the chosen GEM identifiers.
- (3) A string defining a "tag" (unique identifier) that will be added as a suffix to all pseudo-metabolites and pseudo-reactions of the newly generated BOF. This "tag" facilitates identification of all pseudo-reactions and pseudo-metabolite constituting a BOF and provides the possibility to introduce several BOFs in a single GEM.

A function to remove a BOF given its tag is also available in BioModTool (only in command-line version).



**Figure 1. BioModTool usage to generate and add a new BOF to GEM.**

First, BioModTool loads and tests user inputs (GEM, data, tag and compartment identifier). Then, biomass composition data are normalized and converted in appropriate unit. Pseudo-metabolites and pseudo

reactions are instantiated and added to the GEM. When possible, formula and charge are calculated and added to pseudo-metabolite attributes. Stoichiometric coefficient calculations results and updated GEM are respectively saved as XLSX and SBML/JSON files. Circles indicate metabolites existing in the chosen GEM (yellow) and pseudo-metabolites that will be added to GEM by BioModTool (reddish). Arrows represent pseudo-reactions created and added by BioModTool. Final stoichiometric coefficients units of the three levels are indicated between brackets.

Installation and usage procedures are fully detailed in Supplemental\_Text1 (see section 3.4.2 of the manuscript).

### 3. Applications

To demonstrate its applicability, BioModTool was applied to GEM of two bacteria species *Escherichia coli* (iML1515, (Monk et al. 2017)) and *Alicyclobacillus acidocaldarius* (CNA\_Alicyclo, (Beck, Hunt et Carlson 2018)), and one microalga *Chlamydomonas reinhardtii* (iRC1080, (Chang et al. 2011)), using data provided in (Beck, Hunt et Carlson 2018) for iML1515 and CNA\_Alicyclo, and in (Chang et al. 2011) for iRC1080. Generated BOFs were validated in terms of charge and mass balance. Stoichiometric coefficients and calculated formula and charges were also validated by comparison with expected values (Supplemental\_Text2: see section 3.4.3 of the manuscript). For each updated GEM, its ability to predict growth by parsimonious FBA using the newly added BOF as objective was tested and validated. Predicted growth rates were consistent with results obtained using original BOF. The scripts and required input files for these examples are available on GitHub repository ([https://github.com/Total-RD/BioModTool/tree/main/Application\\_examples](https://github.com/Total-RD/BioModTool/tree/main/Application_examples)).

### 4. Conclusion

BioModTool bridges the gap between biomass composition data and generation of structured BOFs. By allowing easy creation of new BOFs, it provides a valuable tool for GEM reconstruction and refinement, and will be useful to account for the dependence of biomass composition on environmental conditions in GEM. BioModTool offers flexibility not only in usable sources to obtain biomass composition data (literature, experimental measurements, utilization of other computational tools such as BOFdat functions with a given omic data set etc.) but also in level of details that can be included in BOF, ranging from basic level to advanced level depending on data availability. This flexibility makes BioModTool usable in many situations. Finally, availability of a graphical interface extends BioModTool usage to non-modelers. When combined to existing interactive simulation tools, such as Escher-FBA (Rowe, Palsson et King 2018), BioModTool has the potential to further spread GEM applications.

### 5. Data availability

BioModTool is implemented in Python. BioModTool is publicly available on GitHub (<https://github.com/Total-RD/BioModTool>) under a GNU Lesser General Public License (LGPL). Scripts and data used for BioModTool validation are available in GitHub repository.

## References

- Beck, Ashley; Hunt, Kristopher; Carlson, Ross (2018): Measuring Cellular Biomass Composition for Computational Biology Applications. In *Processes* 6 (5), p. 38. DOI: 10.3390/pr6050038.
- Chang, Roger L.; Ghamsari, Lila; Manichaikul, Ani; Hom, Erik F. Y.; Balaji, Santhanam; Fu, Weiqi et al. (2011): Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. In *Molecular systems biology* 7, p. 518. DOI: 10.1038/msb.2011.52.
- Cotner, James B.; Makino, Wataru; Biddanda, Bopaiah A. (2006): Temperature Affects Stoichiometry and Biochemical Composition of *Escherichia coli*. In *Microbial Ecology* 52 (1), pp. 26–33. DOI: 10.1007/s00248-006-9040-1.
- Ebrahim, Ali; Lerman, Joshua A.; Palsson, Bernhard O.; Hyduke, Daniel R. (2013): COBRApy: COntstraints-Based Reconstruction and Analysis for Python. In *BMC Syst Biol* 7 (1), p. 74. DOI: 10.1186/1752-0509-7-74.
- Feist, Adam M.; Palsson, Bernhard O. (2010): The biomass objective function. In *Current opinion in microbiology* 13 (3), pp. 344–349. DOI: 10.1016/j.mib.2010.03.003.
- Gu, Changdai; Kim, Gi Bae; Kim, Won Jun; Kim, Hyun Uk; Lee, Sang Yup (2019): Current status and applications of genome-scale metabolic models. In *Genome biology* 20 (1), p. 121. DOI: 10.1186/s13059-019-1730-3.
- Heirendt, Laurent; Arreckx, Sylvain; Pfau, Thomas; Mendoza, Sebastián N.; Richelle, Anne; Heinken, Almut et al. (2019): Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. In *Nature protocols* 14 (3), pp. 639–702. DOI: 10.1038/s41596-018-0098-2.
- Lachance, Jean-Christophe; Lloyd, Colton J.; Monk, Jonathan M.; Yang, Laurence; Sastry, Anand V.; Seif, Yara et al. (2019): BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. In *PLoS computational biology* 15 (4), e1006971. DOI: 10.1371/journal.pcbi.1006971.
- Lavoie, Michel; Saint-Béat, Blanche; Strauss, Jan; Guérin, Sébastien; Allard, Antoine; V. Hardy, Simon et al. (2020): Genome-Scale Metabolic Reconstruction and in Silico Perturbation Analysis of the Polar Diatom *Fragilariopsis cylindrus* Predicts High Metabolic Robustness. In *Biology* 9 (2), p. 30. DOI: 10.3390/biology9020030.
- Monk, Jonathan M.; Lloyd, Colton J.; Brunk, Elizabeth; Mih, Nathan; Sastry, Anand; King, Zachary et al. (2017): iML1515, a knowledgebase that computes *Escherichia coli* traits. In *Nature biotechnology* 35 (10), pp. 904–908. DOI: 10.1038/nbt.3956.
- Nookaew, Intawat; Jewett, Michael C.; Meechai, Asawin; Thammarongtham, Chinae; Laoteng, Kobkul; Cheevadhanarak, Supapon et al. (2008): The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. In *BMC systems biology* 2, p. 71. DOI: 10.1186/1752-0509-2-71.
- O'Brien, Edward J.; Monk, Jonathan M.; Palsson, Bernhard O. (2015): Using Genome-scale Models to Predict Biological Capabilities. In *Cell* 161 (5), pp. 971–987. DOI: 10.1016/j.cell.2015.05.019.

Pramanik, J.; Keasling, J. D. (1998): Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. In *Biotechnol. Bioeng.* 60 (2), pp. 230–238. DOI: 10.1002/(SICI)1097-0290(19981020)60:2<230::AID-BIT10>3.0.CO;2-Q.

Rowe, Elliot; Palsson, Bernhard O.; King, Zachary A. (2018): Escher-FBA: a web application for interactive flux balance analysis. In *BMC systems biology* 12 (1), p. 84. DOI: 10.1186/s12918-018-0607-5.

Schaechter, M.; Maaløe, O.; Kjeldgaard, N. O. (1958) Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella typhimurium*. In : *Microbiology*, vol. 19, n° 3, p. 592–606. DOI: 10.1099/00221287-19-3-592.

Schulz, Christian; Kumelj Tjasa; Karlsen, Emil; Almaas, Eivind (2021): Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. In *PLoS computational biology*. DOI: 10.1371/journal.pcbi.1008528.

Simensen, Vetle; Schulz, Christian; Karlsen, Emil; Bråtelund, Signe; Burgos, Idun; Thorfinnsdottir, Lilja Brekke et al. (2022): Experimental determination of *Escherichia coli* biomass composition for constraint-based metabolic modeling. In *PloS one* 17 (1), e0262450. DOI: 10.1371/journal.pone.0262450.

Zuñiga, Cristal; Levering, Jennifer; Antoniewicz, Maciek R.; Guarnieri, Michael T.; Betenbaugh, Michael J.; Zengler, Karsten (2018): Predicting Dynamic Metabolic Demands in the Photosynthetic Eukaryote *Chlorella vulgaris*. In *Plant physiology* 176 (1), pp. 450–462. DOI: 10.1104/pp.17.00605.

## 3.4.2 Implémentation et utilisation de BioModTool

Cette section correspond au fichier (Supplemental\_Text1) disponible en complément de l'Application Note de BioModTool.

*N.B. : La numérotation des figures et des tables est indépendante du manuscrit de thèse.*

### BioModTool – Install and Usage

#### 1. BioModTool installation

In a prompt, run the commands as shown.

a - clone git repository: `git clone https://github.com/Total-RD/BioModTool.git`

b - change directory: `cd my_path/BioModTool`

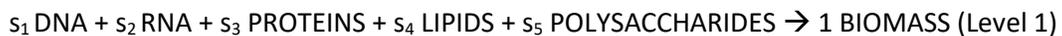
c - install package: `pip install .`

#### 2. BioModTool usage

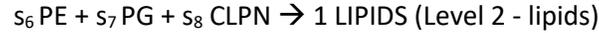
Usage of BioModTool is illustrated using the example of *E. Coli* iML1515 GEM (Monk et al. 2017). Data used to determine *E. Coli* biomass composition come from S4\_Biomass\_Composition supplementary file from Beck, Hunt et Carlson 2018.

##### 2.1 BOF structure and Biomass composition

(Beck, Hunt et Carlson 2018) provide monomer and macromolecular composition of the biomass for: DNA, RNA, proteins, lipids and polysaccharides. Regarding lipids, three classes are defined: PE (Phosphatidylethanolamine), PG (Phosphatidylglycerol) and CLPN (Cardiolipin). A single fatty acid profile is available. Same fatty acid profile for the three lipid classes was therefore considered. Given available data, BOF is structured as follows (with  $s_i$  representing stoichiometric coefficients):



Level 2 is composed of five pseudo-reactions, one for each aforementioned macromolecules, with lipid pseudo-reaction defined as:



Finally, level 3 is defined by three pseudo-reactions: one for each lipid class.

(Beck, Hunt et Carlson 2018) data were used to complete BioModTool Excel file. For example, "BIOMASS" sheet was filled based on macromolecular/polymer composition of biomass (Table 1).

**Table 1. Macromolecular composition of E. Coli (Beck, Hunt et Carlson 2018)**

Macromolecule	$g_{\text{polymer}} \cdot g_{\text{DW}}^{-1}$
DNA	1.0
RNA	17.2
proteins	35.2
lipids	6.7
polysaccharides	4.2

Raw data are given in  $g_{\text{polymer}} \cdot g_{\text{DW}}^{-1}$ , so unit "g per ..." was selected in "D12 cell" drop-down list. No energy or other cofactors are required in BOF level 1, so Table 2 was left empty. BIOMASS sheet fully completed is illustrated in (Figure 1). A level\_2 sheet was created and named "DNA" (Figure 2). *E. Coli* genome GC-content was used to detail DNA composition in Table 1 (24.6% A, 24.6% T, 25.4% C and 25.4% G) using metabolites identifiers (datp\_c, dttp\_c, dctp\_c and dgtt\_c) from in iML1515 GEM (Monk et al. 2017). DNA polymerization produces one diphosphate per dNTP addition, therefore production of one mole of ppi\_c metabolite per mole of DNA was added in DNA sheet Table 2. Same methodology was applied to create and fill the nine sheets required to generate the BOF matching the structure described above (BIOMASS, DNA, RNA, PROTEINS, LIPIDS, POLYSACCHARIDES, PE, PG and CLPN). In Beck, Hunt et Carlson 2018 polymer lengths for the macromolecular synthesis reactions was set to 10 NTPs for RNA, 10 glucose-1-phosphates for polysaccharides and 100 amino acids for proteins. Polymerization energy requirements and byproducts in Beck, Hunt et Carlson 2018 data are given on the basis of these polymers lengths. On the other hand in BioModTool all polymers are considered to be one monomer long, stoichiometric coefficients of energy and byproducts requirement reported by Beck, Hunt et Carlson 2018 were therefore adjusted.

Fully completed Excel file is available as (Supplemental File S4).

	A	B	C	D	E
1	<b>BIOMASS (level 1) - Instructions</b>				
2	Please fill the sheet according to the following instructions:				
3	1 - Fill metabolite IDS	<u>Table1</u> : Pseudo-metabolite ID (also used for pseudo-reaction ID). /\ Add a sheet for each pseudo metabolite. (Sheet name = pseudo-metabolite ID)			
4		Additional informations can be added. Names will be added to pseudo-metabolites and pseudo-reactions (optional). Comments/Source cells are not taken into account by BioModTool.			
5	2 - Indicate data unit	<u>Table 1</u> : Choose appropriate unit (in D12 cell).	Possible units:	g per ...	for example g/gDCW, g/cells, mass% ...
6				mol per ...	for example mol/gDCW, mol/cells, mol% ...
7	3 - Fill coefficients	<u>Table 1</u> : coefficient value for each pseudo-metabolite (macromolecule/polymer). /\ Coefficient >= 0			
8					
9	<b>BIOMASS (level 1) - Data read by BioModTool</b>				
10	<b>Table 1: Basic level metabolites</b> <i>coefficient &gt;= 0</i> <i>Coefficients will be converted and normalized by BioModTool.</i>				
11	<b>Metabolite name</b>	<b>Metabolite ID in model</b>	<b>Coefficient</b>	<b>Unit</b>	<b>Comments/Source</b>
12	DNA pseudo-metabolite	DNA	1	g per ...	data in gpolymer/gDW Beck et al., 2018
13	RNA pseudo-metabolite	RNA	17.2	g per ...	Beck et al., 2018
14	PROTEINS pseudo-metabolite	PROTEINS	35.2	g per ...	Beck et al., 2018
15	LIPIDS pseudo-metabolite	LIPIDS	6.7	g per ...	Beck et al., 2018
16	POLYSACCHARIDES pseudo-metabolite	POLYSACCHARIDES	4.2	g per ...	Beck et al., 2018
17				g per ...	
18				g per ...	

Figure 1. Example of a "BIOMASS" sheet of BioModTool Excel filled with *E. coli* data from Beck et al., 2018.

	A	B	C	D	E	F	
	<b>DNA (level 2) - Instructions</b>						
1	<b>Please fill the sheet according to the following instructions:</b>						
2	<b>1 - Fill metabolite IDS</b>	<b>Tables 1 &amp; 2: Metabolite IDs.</b>					
3		<b>!\ Must correspond to a metabolite of the chosen model.</b>					
		<i>Additional information can be added. Names will be added to pseudo-metabolites and pseudo-reactions (optional). Comments/Source cells are not taken into account by BioModTool.</i>					
4	<b>2 - Indicate data unit</b>	<b>Table 1:</b> Choose appropriate unit	Possible units:	<b>g per ...</b>	for example g/gDCW, g/cells, mass% ...		
5		(in D12 cell).		<b>mol per ...</b>	for example mol/gDCW, mol/cells, mol% ...		
6	<b>3 - Fill coefficients</b>	<b>Table 1:</b> coefficient value for each pseudo-metabolite (macromolecule/polymer).					
		<b>!\ Coefficient &gt;= 0</b>					
7							
8	<b>DNA (level 2) - Data read by BioModTool</b>						
9	<b>Table 1: Basic level metabolites</b> <i>coefficient &gt;= 0</i> <i>Coefficients will be converted and normalized by BioModTool.</i>						
10	<b>Metabolite name</b>	<b>Metabolite ID in model</b>	<b>Coefficient</b>	<b>Unit</b>	<b>Comments/Source</b>		
11	dATP	datp_c	24.6000	mol per ...	data in gpolymer/gDW Beck et al., 2018		
12	dCTP	dctp_c	25.4000	mol per ...	Beck et al., 2018		
13	dGTP	dgtp_c	25.4000	mol per ...	Beck et al., 2018		
14	dTTP	dttp_c	24.6000	mol per ...	Beck et al., 2018		
15				mol per ...			

Figure 2. Example of a "DNA" sheet of BioModTool Excel filled with *E. coli* data from Beck et al., 2018. (Continue next page)

## DNA (level 2) - Instructions

**Tables 1 & 2: Metabolite IDs.**

**!\ Must correspond to a metabolite of the chosen model.**

*Additional information can be added. Names will be added to pseudo-metabolites and pseudo-reactions (optional). Comments/Source cells are not taken into account by BioModTool.*

**Table 2:** Unit can not be changed (mmol/gDW)

**Table 2:** coefficient value for each metabolite.

**!\ Coefficient < 0 if the metabolite is consumed and >0 if metabolite is produced by the reaction.**

## DNA (level 2) - Data read by BioModTool

**Table 2: Intermediate and advanced levels metabolites (optional)**

*if consumed: coeff < 0,*

*if produced coeff >0*

**!\ Coefficients will be directly used in pseudo-reaction (no conversion).**

Metabolite name	Constant metabolite ID in model	Constant metabolite Coefficient	Constant metabolite Unit	Comments/Source
Diphosphate	ppi_c	1.0000	mol/mol macromolecule	Beck et al., 2018
			mol/mol macromolecule	

End of Figure 2.

## 2.2 Run BioModTool

BioModTool can be run either in Python command-lines or using the user interface (implemented with Tkinter).

### 2.2.1 Line-command version

The following section presents an example of a Jupyter Notebook to run BioModTool. This Jupyter Notebook is available: `Ecoli_iML1515_script_BioModTool_add_biomass_reaction.ipynb` at <https://github.com/Total->

[RD/BioModTool/tree/main/Application\\_examples/1\\_iML1515\\_ecoli/test\\_with\\_jupyternotebook](https://github.com/Total-RD/BioModTool/tree/main/Application_examples/1_iML1515_ecoli/test_with_jupyternotebook).

## Addition of a new BOF to iML1515 using BioModTool:

- Cobra model: iML1515 (*E. coli*) (Monk et al. 2017)
- Biomass composition data from (Beck, Hunt et Carlson 2018)

### Imports

In [1]:

```
import BioModTool.load
import BioModTool.main_add_biomass_objective_function
import BioModTool.save
```

## 1 - Load Genome Scale Metabolic Model

### 1.a) Model directory

In [2]:

```
path_to_model = "path_to_model_repository\\iML1515.xml"
```

In [3]:

```
original_model = BioModTool.load.load_cobra_model(path_to_model)
```

### 1.b) Calculate formula and/or charge ?

**!\\ Note that if you select formula = False, the molecular weight cannot be calculated from the metabolite formula preventing unit conversion.**

- Level 1 data must be given in mmol.gDW-1
- Levels 2 and 3 data must be given in mol per ...

In [4]:

```
calculate_formula = True
calculate_charge = True
```

### 1.c) Choose a compartment to add biomass reactions

- Reactions and pseudo metabolites will be added in compartment given by the user.
- Compartment must be chosen among model's compartments (key of cobra\_model.compartments dictionary)
- Biomass reactions are commonly added in cytosol ("c").

In [5]:

```
# Compartments in model:
original_model.compartments
```

Out[5]:

```
{'c': 'cytosol', 'e': 'extracellular space', 'p': 'periplasm'}
```

In [6]:

```
BOF_compartment = "c"
```

## 2 – Biomass composition data

### 2.a) Define data file

Data must be given in an Excel file with a specific format. See Supplemental Files S1-S3.

In [7]:

```
path_to_data = "Biomass_composition_ecoli_Beck2018.xlsx"
```

## 3 – Structure of biomass objective function

### 3.a) Biomass reaction structure

Dictionary defining BOF structure:

Must be defined in accordance with Excel data file (sheet names).

- Mandatory:
  - one and only one key with value = "level\_1"
- Optional:
  - if desired structure contains two or more levels:
    - one or several key(s) with value = "level\_2"
  - if desired structure contains three levels:
    - one and only one key with value = "level\_2\_lipid"
    - one or several key(s) with value = "level\_2"

In [8]:

```
dict_pool_id_Ecoli = {'BIOMASS': 'level_1',
  'POLYSACCHARIDES': 'level_2',
  'DNA': 'level_2',
  'RNA': 'level_2',
  'PROTEINS': 'level_2',
  'LIPIDS': 'level_2_lipid',
  'PG': "level_3",
  'PE': "level_3",
  'CLPN': "level_3"}
```

### 3.b) Choose a suffix

- Expected suffix: string containing only alphanumeric characters or \_
- Test performed in BioModTool: `re.match("^[a-zA-Z0-9_]*$", suffix)`
- All added pseudo-reactions and pseudo-metabolites will contain the given suffix.

In [9]:

```
suffix = "BioModTool_beck2018"
```

## 4 – Create and add the new biomass objective function

In [10]:

```
updated_model =
BioModTool.main_add_biomass_objective_function.add_biomass_objective_function(
  cobra_model = original_model,
  path_to_data = path_to_data,
  suffix = suffix,
  dict_structure = dict_structure_BOF,
  user_compartment = BOF_compartment,
  calculate_charge = calculate_charge,
  calculate_formula = calculate_charge,
  saving_final_data = True)
```

Out[10]:

Metabolite POLYSACCHARIDES\_BioModTool\_beck2018\_c (formula: C6.0H10.0O5.0, charge: 0) added to model.

Reaction POLYSACCHARIDES\_BioModTool\_beck2018\_c added to model.

```

atp_c + g1p_c --> POLYSACCHARIDES_BioModTool_beck2018_c + adp_c + ppi_c
-----
Metabolite DNA_BioModTool_beck2018_c (formula: C9.746H11.246O6.0N3.754P1.0,
charge: -1) added to model.
Reaction DNA_BioModTool_beck2018_c added to model.
0.246 datp_c + 0.254 dctp_c + 0.254 dgtp_c + 0.246 dttp_c -->
DNA_BioModTool_beck2018_c + ppi_c
-----
[...]
-----
Metabolite BIOMASS_BioModTool_beck2018_c (formula:
C40.59259727004123H62.339155478773726O16.850976170654622N10.09717316334253P1.0
37995606777102S0.2007301352, charge: -1) added to model.
Reaction BIOMASS_BioModTool_beck2018_c added to model.
0.0505 DNA_BioModTool_beck2018_c + 0.13999 LIPIDS_BioModTool_beck2018_c +
0.40285 POLYSACCHARIDES_BioModTool_beck2018_c + 5.02831
PROTEINS_BioModTool_beck2018_c + 0.84039 RNA_BioModTool_beck2018_c -->
BIOMASS_BioModTool_beck2018_c
-----

```

## 5 – Save updated model

BioModTool comes with a function to save model both in JSON and SBML formats.

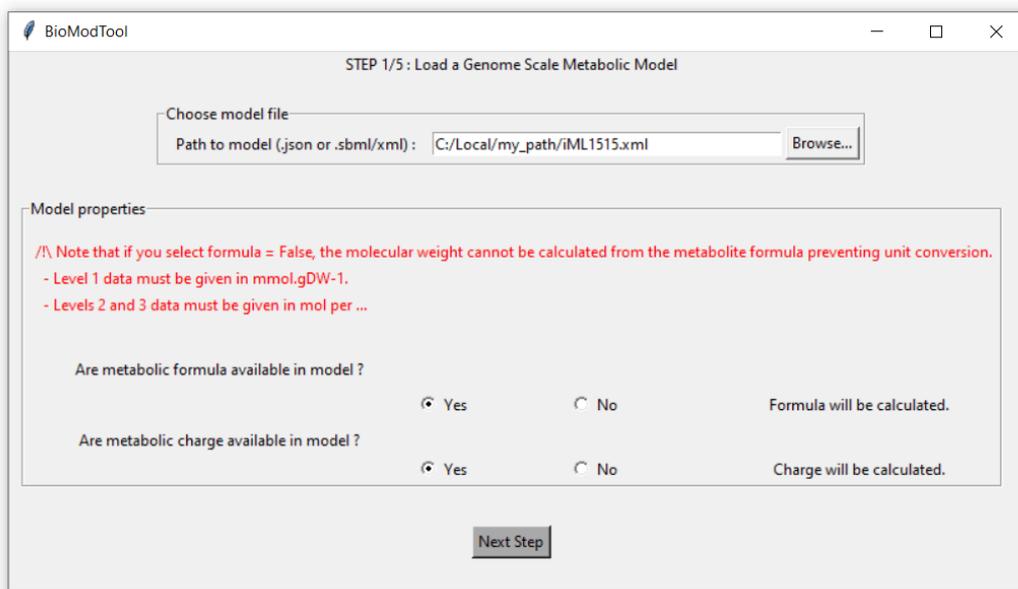
In [11]:

```
BioModTool.save.save_model(updated_model, "iML1515_updated" )
```

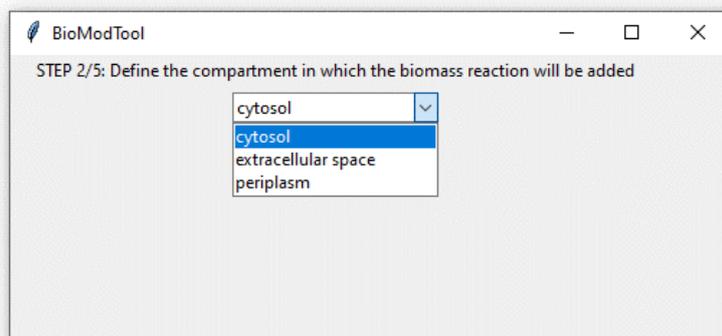
### 2.2.2 Using graphical interface

BioModTool comes with a graphical user interface: `interface_BioModTool.py`. The interface can be executed using Python and is compatible with Windows, Linux and MacOS operating systems. On Windows, the interface can also be launched directly from an executable: `interface_BioModTool.exe`. N.B. : `interface_BioModTool.exe` is in the following repository **BioModTool \dist**

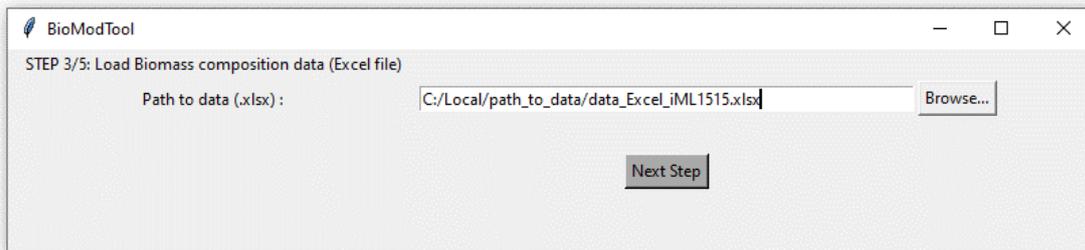
**Step 1:** Browse GEM model file (JSON or SBML format) and select via radio buttons if formula and/or charge are available for all metabolites consumed in the BOF.



**Step 2:** Select a compartment in the drop-down list. Biomass objective function will be added to the selected compartment.

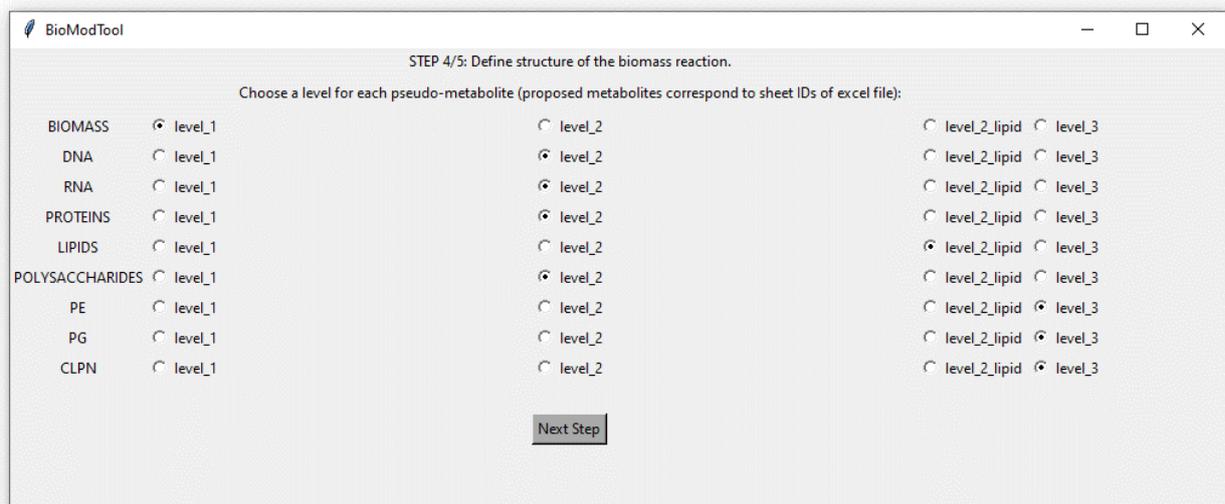


**Step 3:** Browse biomass composition data (Excel file, \*.xlsx).

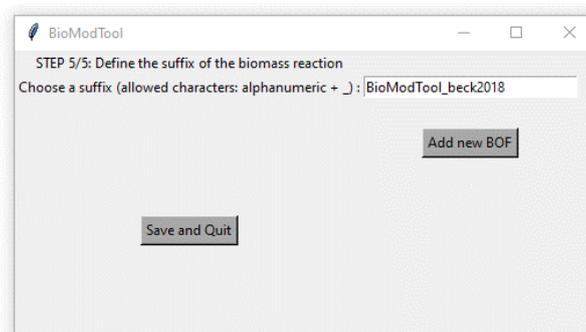


**Step 4:** Based on Excel data file structure (available sheets), define BOF structure. To do so, indicate desired level for each pseudo-metabolite by selecting desired level:

- level\_1,
- level\_2,
- level\_2\_lipide,
- level\_3.



**Step 5:** Define a suffix, only alphanumeric characters (and "\_") are accepted. Add the BOF, save and quit BioModTool.



Updated model is saved in SBML and JSON formats in a directory chosen by the user. Excel file with calculations details is saved in the working directory.

### 2.3 Remove a BOF with BioModTool

A function removing a BioModTool formatted BOF from GEM is available (`remove_biomass_objective_function`). The function takes a model and a suffix as parameters, remove all reactions harboring the given suffix ("`_suffix_`") in their ID and returns updated model. This functionality is available in line command version.

In [12]:

```
model_wo_BOF = remove_biomass_objective_function(updated_model, suffix)
```

## 3. Implementation

First, BioModTool loads and tests user inputs. Then, stoichiometric coefficients are calculated from data.

For levels 2 and 3 pseudo-reactions, final stoichiometric coefficients are in molar fraction ( $\text{mol}\cdot\text{mol}^{-1}$ ). For each pseudo-reaction if user data are in "g per ...", data are converted in "mol per ..." as follows:

$$coeff_{mol} = \frac{coeff_g}{MW}$$

With  $coeff_{mol}$  molar coefficient in "mol per ...",  $coeff_g$  initial mass coefficient in "g per ..." and  $MW$  molecular weight of the metabolite calculated from metabolite formula in the GEM.

In BioModTool,  $MW$  are calculated using COBRApy function `formula_weight` from metabolite instances (Ebrahim et al. 2013). Molar coefficients are then normalized so that the sum of all molar fractions is one, and rounded to the fifth decimal place.

By definition, produced biomass must represent one gram of dry weight. Therefore, for level 1 pseudo-reaction (typically biomass), final stoichiometric coefficients are in  $\text{mmol}\cdot\text{gDW}^{-1}$ . Before being converted in  $\text{mmol}\cdot\text{gDW}^{-1}$ , coefficients given by the user are first converted in  $\text{g}\cdot\text{gDW}^{-1}$  to allow their normalization. Data in "mol per ..." are converted in "g per ..." as follows:

$$coeff_g = coeff_{mol} * MW$$

With  $coeff_g$  mass coefficient in "g per ...",  $coeff_{mol}$  initial molar coefficient in "mol per ..." and  $MW$  molecular weight of the metabolite obtained from metabolite formula. Mass coefficients are then

normalized such that the sum of metabolite mass fraction is one, before being converted in final unit mmol.gDW<sup>-1</sup> as follows:

$$coeff_{mmol/gDW} = 1000 * \frac{coeff_{g/gDW}}{MW}$$

With  $coeff_{mmol/gDW}$  final stoichiometric coefficient in mmol.gDW<sup>-1</sup>,  $coeff_{g/gDW}$  mass fraction in g.gDW<sup>-1</sup> and  $MW$  molecular weight of the metabolite. Final coefficients are rounded to fifth decimal place. BioModTool calculates pseudo-metabolites charge and formula (when possible and desired by the user), and instantiates pseudo-metabolites and pseudo-reactions. Annotation being an important feature in GEM models, SBO terms "SBO:0000247" and "SBO:0000629" are respectively added to metabolites and reactions annotations (Lieven et al. 2020).

In accordance with the FAIR (Findable, Accessible, Interoperable, Reusable) principles and to offer users a comprehensible overview of the calculations executed by BioModTool, the results of the various calculations steps are recorded in an Excel file. The pseudo-metabolites created, along with their charge and chemical formula, are also summarized in this Excel file (worksheet: added\_metabolites). In addition, BioModTool performs a charge and mass balance test of all the reactions making up the biomass function. Results of this test are saved in the Excel file (worksheet: mass\_charge\_balance).

Steps that require loading, updating, saving GEM and manipulating metabolites and reactions rely on COBRAPy packages (Ebrahim et al. 2013). Full codes are available on GitHub (<https://github.com/Total-RD/BioModTool>).

## References

Beck, Ashley; Hunt, Kristopher; Carlson, Ross (2018): Measuring Cellular Biomass Composition for Computational Biology Applications. In Processes 6 (5), p. 38. DOI: 10.3390/pr6050038.

Ebrahim, Ali; Lerman, Joshua A.; Palsson, Bernhard O.; Hyduke, Daniel R. (2013): COBRAPy: COConstraints-Based Reconstruction and Analysis for Python. In BMC Syst Biol 7 (1), p. 74. DOI: 10.1186/1752-0509-7-74.

Lieven, Christian; Beber, Moritz E.; Olivier, Brett G.; Bergmann, Frank T.; Ataman, Meric; Babaei, Parizad et al. (2020): MEMOTE for standardized genome-scale metabolic model testing. In Nature biotechnology 38 (3), pp. 272–276. DOI: 10.1038/s41587-020-0446-y.

Monk, Jonathan M.; Lloyd, Colton J.; Brunk, Elizabeth; Mih, Nathan; Sastry, Anand; King, Zachary et al. (2017): iML1515, a knowledgebase that computes Escherichia coli traits. In Nature biotechnology 35 (10), pp. 904–908. DOI: 10.1038/nbt.3956.

### 3.4.3 Validation de BioModTool

Afin de valider BioModTool, ce dernier a été utilisé afin d'ajouter des fonctions de biomasses à trois GEM de la littérature. Cette section correspond au fichier (Supplemental\_Text2) disponible en complément de l'Application Note de BioModTool.

*N.B.1 : La numérotation des figures et des tables est indépendante du manuscrit de thèse.*

## BioModTool - Validation

BioModTool was applied to GEM of two bacteria species *Escherichia coli* (iML1515, (Monk et al. 2017)) and *Alicyclobacillus acidocaldarius* (CNA\_Alicyclo, (Beck, Hunt et Carlson 2018)), and one microalga *Chlamydomonas reinhardtii* (iRC1080, (Chang et al. 2011)). These examples illustrate the three different contexts in which BioModTool can be used:

- Add a three level BOF in a model in which metabolic formula are available: iML1515
- Add a two level BOF in a model in which metabolic formula are not available: CNA\_Alicyclo
- Add a one level BOF in a model in which metabolic formula are not available: iRC1080

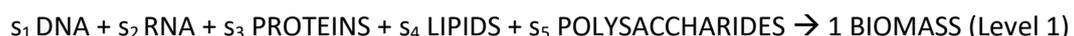
All scripts and results of these three application examples are fully available: [https://github.com/Total-RD/BioModTool/tree/main/Application\\_examples](https://github.com/Total-RD/BioModTool/tree/main/Application_examples).

## 1. Addition of a 3 levels BOF to iML1515 GEM

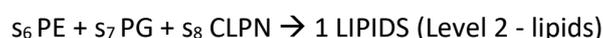
### 1.1 Methods

#### 1.1.1 BOF addition using BioModTool

BioModTool was applied to iML1515, the latest version of the *E. Coli* model (Monk et al. 2017). Biomass composition data were obtained from (Beck, Hunt et Carlson 2018) (File S4\_Biomass\_Composition). Data available allow definition of a three levels BOF. Level 1 pseudo-reaction is defined as:



With  $s_i$  stoichiometric coefficients. Level 2 is composed of five pseudo-reactions, one for each aforementioned macromolecules, with lipid pseudo-reaction defined as:



Finally, level 3 is defined by three pseudo-reactions, one for each lipid class.

Raw data were used to fill BioModTool Excel (Supplemental File S4). BioModTool was used to generate and add a new BOF to iML1515 with following parameters: dict\_structure = {'POLYSACCHARIDES': 'level\_2', 'DNA': 'level\_2', 'RNA': 'level\_2', 'PROTEINS': 'level\_2', 'LIPIDS': 'level\_2\_lipid', 'PG': "level\_3", 'PE': "level\_3", 'CLPN': "level\_3", 'BIOMASS': 'level\_1'}, calculate\_charge = True, calculate\_formula = True.

### 1.1.2 Addition of Beck BOF to iML1515

Beck et al. 2018 introduce a BOF for *E. Coli*, defined by six reactions (Table 1). These reactions were added to iML1515. Macromolecule fraction stoichiometric coefficients are expressed in  $\text{mol}_{\text{macromolecule}}/100\text{kg}_{\text{DW}}$ . These coefficients were converted in  $\text{mmol}_{\text{macromolecule}}\cdot\text{g}_{\text{DW}}^{-1}$  to allow comparison with BioModTool BOF. Moreover, in Beck et al. 2018 polymer lengths for the macromolecular synthesis reactions was set to 10 NTPs for RNA, 10 glucose-1-phosphates for polysaccharides and 100 amino acids for proteins. Stoichiometric coefficients were adjusted so that all polymers are one monomer long.

**Table 1. Beck BOF reactions.**

Reaction ID	Equation
Biomass reaction	5.05029378916591 DNA + 8.40374146379437 RNA + 5.01641118949985 protein + 13.9019957911911 lipid + 4.02851950095506 polysaccharide = 1 biomass
DNA synthesis reaction	0.246 dATP + 0.254 dCTP + 0.254 dGTP + 0.246 dTTP = 1 DNA + 1 diphosphate
RNA synthesis reaction	2.0262150129207 ATP + 3.1523397366045 CTP + 2.25131542077898 GTP + 2.57012982969582 UTP = 1 RNA + 10 diphosphate
Protein synthesis reaction	10.1 Alanine + 1.2 Cysteine + 5.38333333333333 Aspartate + 6.88039215686274 Glutamate + 3.9 Phenylalanine + 10.9 Glycine + 1.6 Histidine + 4.8 Isoleucine + 5.8 Lysine + 7.2 Leucine + 2.8 Methionine + 4.11666666666667 Asparagine + 4 Proline + 5.21960784313726 Glutamine + 5 Arginine + 4.9 Serine + 5.3 Threonine + 6.8 Valine + 1.5 Tryptophan + 2.8 Tyrosine + 100 ATP + 198 GTP + 197.8 H <sub>2</sub> O = 1 protein + 100 AMP + 100 diphosphate + 198 GDP + 198 phosphate + 298 H <sup>+</sup>
Lipid synthesis reaction	0.76530612244898 Phosphatidylethanolamine + 0.183673469387755 Phosphatidylglycerol + 0.0510204081632653 Cardiolipin = 1 lipid
Polysaccharide synthesis reaction	10 Glucose-1-phosphate + 10 ATP = 1 polysaccharide + 10 ADP + 10 diphosphate

### 1.1.3 Predict maximum growth rate using pFBA

Flux distributions were performed using open-source COBRApy package (Ebrahim et al. 2013); version 0.26.3), with Python version 3.11.4. Maximum growth rate was predicted by Parsimonious Flux Balance Analysis (pFBA) with respectively original Beck BOF and BioModTool\_Ecoli BOF as objective functions. Default iML1515 constrains were kept unchanged for these simulations.

### 1.2 Results

BioModTool was used to add the generate a new BOF for *E. coli* from (Beck, Hunt et Carlson 2018) biomass composition data. Final coefficients, pseudo-metabolites formula and charge calculated by BioModTool were compared to (Beck, Hunt et Carlson 2018) calculations (Table 1). Maximum growth rate was predicted using pFBA with respectively Beck BOF and BioModTool\_Ecoli BOF as objective functions. A maximum growth rate of 1.1301 h<sup>-1</sup> was predicted for both BOFs. All reactions constituting BioModTool\_Ecoli BOF are mass and charge balanced.

**Table 2. Comparison of BioModTool calculations to Beck et al., 2018 for *E. coli*.**

	Beck et al., 2018 calculations				BioModTool calculations		
	Coefficient [mol polymer/100 kg cdw]	Coefficient [mmol.gDW <sup>-1</sup> ]	Formula	Charge	Coefficient [mmol.gDW <sup>-1</sup> ]	Formula	Charge **
DNA	5.05	5.05E-02	C <sub>9.75</sub> H <sub>11.25</sub> N <sub>3.75</sub> O <sub>6</sub> P <sub>1</sub>	-1	5.05E-02	C <sub>9.746</sub> H <sub>11.246</sub> O <sub>6.0</sub> N <sub>3.754</sub> P <sub>1.0</sub>	-1
RNA *	8.40	8.40E-02	C <sub>94.28</sub> H <sub>107.43</sub> N <sub>35.98</sub> O <sub>70.54</sub> P <sub>10</sub>	-10	8.40E-01	C <sub>9.428</sub> H <sub>10.743</sub> O <sub>7.0543</sub> N <sub>3.598</sub> P <sub>0.1</sub>	-1
protein *	5.02	5.02E-02	C <sub>481.30</sub> H <sub>753.37</sub> N <sub>135.04</sub> O <sub>147.06</sub> S <sub>4</sub>	-1.46	5.03	C <sub>4.803</sub> H <sub>7.514</sub> O <sub>1.466</sub> N <sub>1.348</sub> S <sub>0.04</sub>	0
lipid	1.39E+01	1.39E-01	C <sub>40.99</sub> H <sub>78.73</sub> N <sub>0.18</sub> O <sub>10.01</sub> P <sub>1.05</sub>	-2.86E-01	1.40E-01	C <sub>40.061</sub> H <sub>78.074</sub> O <sub>8.826</sub> N <sub>0.765</sub> P <sub>1.051</sub>	0
polysaccharide *	4.03	4.03E-02	C <sub>60</sub> H <sub>100</sub> O <sub>50</sub>	0	4.03E-01	C <sub>6.0</sub> H <sub>10.0</sub> O <sub>5.0</sub>	0
<b>Biomass formula</b>	C <sub>4059</sub> H <sub>6236</sub> N <sub>1009</sub> O <sub>13685</sub> P <sub>104</sub> S <sub>20</sub>				C <sub>40.593</sub> H <sub>62.339</sub> O <sub>16.851</sub> N <sub>10.097</sub> P <sub>1.038</sub> S <sub>0.201</sub>		
<b>Biomass formula normalized to C</b>	CH <sub>1.54</sub> N <sub>0.25</sub> O <sub>0.42</sub> P <sub>0.03</sub> S <sub>0.005</sub>				C <sub>1</sub> H <sub>1.54</sub> O <sub>0.42</sub> N <sub>0.25</sub> P <sub>0.03</sub> S <sub>0.005</sub>		

\* In (Beck et al. 2018) polymer lengths for the macromolecular synthesis reactions was set to 10 NTPs for RNA, 10 glucose-1-phosphates for polysaccharides and 100 amino acids for proteins. On the other hand, in BioModTool all polymers are considered to be one monomer long. This difference explains the observed factor of 10 and 100 between the results of Beck and the results of BioModTool for DNA and polysaccharide, and proteins respectively. \*\* In GEM models, charge must be an integer. In BioModTool charges are therefore rounded to the nearest integer value.

## 2. Addition of 2 levels BOF to CNA\_Alicyclo *A. acidocaldarius* GEM

### 2.1 Methods

#### 2.1.1 BOF addition using BioModTool

BioModTool was applied to CNA\_Alicyclo model (Beck, Hunt et Carlson 2018). Biomass composition data were obtained from (Beck, Hunt et Carlson 2018) (File S4\_Biomass\_Composition). Data available allow definition of a two levels BOF. Level 1 pseudo-reaction is defined as:



Level 2 is composed of five pseudo-reactions, one for each aforementioned macromolecules.

Raw data were used to fill BioModTool Excel (Supplemental File S5). Since no formula are available in CNA\_Alicyclo, no data conversion, requiring molecular weight which is calculated from metabolite formula, can be performed by BioModTool. Biomass macromolecular composition was therefore given in final unit: mmol.gDW<sup>-1</sup>. BioModTool was used to generate and add a new BOF to CNA\_Alicyclo (Beck, Hunt et Carlson 2018) with following parameters: dict\_structure = {'POLYSACCHARIDES': 'level\_2', 'DNA': 'level\_2', 'RNA': 'level\_2', 'PROTEINS': 'level\_2', 'LIPIDS': 'level\_2', 'BIOMASS': 'level\_1'}, calculate\_charge = False, calculate\_formula = False.

### 2.1.2 Predict maximum growth rate using pFBA

Flux distributions were performed using open-source COBRApy package (Ebrahim et al. 2013) version 0.26.3), with Python version 3.11.4. Maximum growth rate was predicted by Parsimonious Flux Balance Analysis (pFBA) with respectively original B0 BOF and BioModTool\_Aacidocaldarius BOF as objective functions. Default model constrains were kept unchanged for these simulations.

## 2.2 Results

BioModTool was used to add the generate a new BOF for *A. acidocaldarius* from (Beck, Hunt et Carlson 2018) biomass composition data. BioModTool\_Aacidocaldarius BOF coefficients were compared to (Beck, Hunt et Carlson 2018) calculations (Table 3). Maximum growth rate was predicted using pFBA with respectively original B0 BOF and BioModTool\_Aacidocaldarius BOF as objective functions. A maximum growth rate of 35.842 h<sup>-1</sup> was predicted for both BOFs.

**Table 3. Comparison of BioModTool coefficients to Beck et al., 2018 for *A. acidocaldarius*.**

	Beck et al., 2018		BioModTool
	Coefficient [mol polymer/100 kg cdw]	Coefficient [mmol.gDW-1]	Coefficient [mmol.gDW-1]
DNA	3.49E+00	3.49E-02	3.49E-02
RNA *	7.91E+00	7.91E-02	7.91E-01
Protein *	5.46E+00	5.46E-02	5.46
lipid	6.67E+00	6.67E-02	6.67E-02
Polysaccharides *	5.86E+00	5.86E-02	5.86E-01

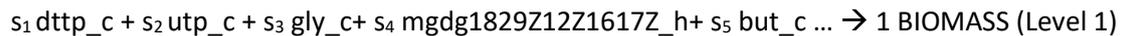
\* In (Beck et al. 2018) polymer lengths for the macromolecular synthesis reactions was set to 10 NTPs for RNA, 10 glucose-1-phosphates for polysaccharides and 100 amino acids for proteins. On the other hand, in BioModTool all polymers are considered to be one monomer long. This difference explains the observed factor of 10 and 100 between the results of Beck and the results of BioModTool for DNA and polysaccharide, and proteins respectively.

### 3. Addition of single level BOF to iRC1080 *C. reinhardtii* GEM

#### 3.1 Methods

##### 3.1.1 BOF addition using BioModTool

BioModTool was applied to iRC1080 model of *Chlamydomonas reinhardtii* (Chang et al. 2011). Biomass composition data were obtained from (Chang et al. 2011) (msb201152-s13.xlsx file, Autotrophic data). Data available allow definition of a BOF with only one level. BOF reaction is defined as:



Since formula is missing for several metabolites consumed in BOF reaction, no data conversion (requiring molecular weight which is calculated from metabolite formula) can be performed by BioModTool. All coefficients were therefore given in mmol.gDW<sup>-1</sup> in BioModTool Excel (Supplemental File S6). BioModTool was used to generate and add the new BOF to iRC1080 with following parameters: dict\_structure = {"BIOMASS": "level\_1"}, calculate\_charge = False, calculate\_formula = False.

##### 3.1.2 Predict maximum growth rate using pFBA

Flux distributions were performed using open-source COBRAPy package (Ebrahim et al. 2013); version 0.26.3), with Python version 3.11.4. Maximum growth rate was predicted by Parsimonious Flux Balance Analysis (pFBA) with respectively original Biomass\_Chlamy\_auto BOF and BioModTool\_Creinhardtii BOF as objective functions. Default model constrains were kept unchanged for these simulations.

##### 3.1.3 Results

Maximum growth rate was predicted using pFBA with respectively original Biomass\_Chlamy\_auto BOF and BioModTool BOF as objective functions. A maximum growth rate of 6.157 h<sup>-1</sup> was predicted for both BOFs.

### 4. Conclusion

Taken together, these three application examples validate the efficiency of BioModTool in generating functional BOF from user data. These examples also underscore the flexibility of BioModTool in terms of (1) updatable GEM, (2) generated BOFs structure (with one, two, or three levels) and (3) BOF complexity including basic, intermediate and advanced levels metabolites.

## References

Beck, Ashley; Hunt, Kristopher; Carlson, Ross (2018): Measuring Cellular Biomass Composition for Computational Biology Applications. In *Processes* 6 (5), p. 38. DOI: 10.3390/pr6050038.

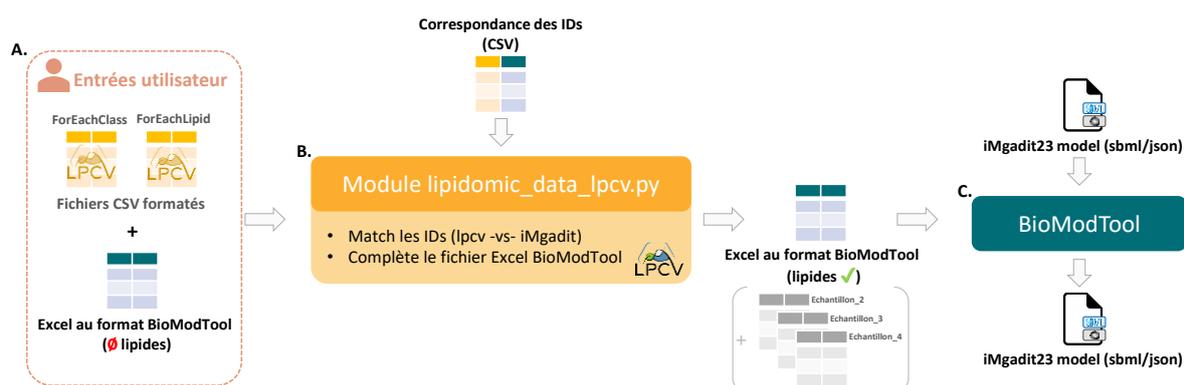
Chang, Roger L.; Ghamsari, Lila; Manichaikul, Ani; Hom, Erik F. Y.; Balaji, Santhanam; Fu, Weiqi et al. (2011): Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. In *Molecular systems biology* 7, p. 518. DOI: 10.1038/msb.2011.52.

Monk, Jonathan M.; Lloyd, Colton J.; Brunk, Elizabeth; Mih, Nathan; Sastry, Anand; King, Zachary et al. (2017): iML1515, a knowledgebase that computes *Escherichia coli* traits. In *Nature biotechnology* 35 (10), pp. 904–908. DOI: 10.1038/nbt.3956.

## 3.5 Module spécifique aux données lipidomiques du LPCV

### 3.5.1 Présentation du module lipidomic\_data\_lpcv.py

L'objectif du module lipidomic\_data\_lpcv.py est d'automatiser l'intégration des données lipidomiques au format du laboratoire, facilitant ainsi l'ajout automatique de nouvelles fonctions de biomasse à iMgadi23 (Figure 3.1).



**Figure 3.1. Représentation schématique de la fonctionnalité du module lipidomic\_data\_lpcv.py.** (A) L'utilisateur fournit des données lipidomiques au format LPCV ainsi que le fichier Excel de BioModTool à compléter. (B) Le module charge un fichier CSV permettant d'établir des correspondances entre les identifiants du LPCV et les identifiants de iMgadi23, puis complète automatiquement les données lipidiques du fichier BioModTool. Le module offre la possibilité d'intégrer les lipidomes de plusieurs échantillons simultanément. Un fichier Excel est alors généré pour chaque échantillon. (C) Le fichier Excel ainsi complété peut être utilisé par BioModTool pour ajouter une fonction de biomasse à iMgadi23.

Le module nécessite plusieurs entrées de l'utilisateur, notamment :

- Un fichier de données de composition de biomasse au format BioModTool, structuré pour créer une fonction de biomasse adaptée à iMgadi23. Dans ce modèle, la fonction de biomasse est organisée en trois niveaux comprenant six macromolécules (ADN, ARN, GLUCIDES, PROTÉINES, PIGMENTS et LIPIDES) ainsi que dix classes lipidiques (PC, PE, PG, PI, DAG, TAG, DGTS, MGDG, DGDG et SQDG). Les sections non liées aux lipides peuvent être complétées avant ou après l'utilisation du module lipidomic\_data\_lpcv.py
- Un glycérolipidome de *M. gaditana* au format du LPCV, comprenant le profil des différentes classes de glycérolipides (fichier CSV ForEachClass, Table 3.3) et le profil détaillé de toutes les entités lipidiques de *M. gaditana* (fichier CSV ForEachLipid, Table 3.4). Les fichiers peuvent contenir les profils lipidiques de plusieurs échantillons.

**Table 3.3. Format d'un fichier ForEachClass du LCPV pour M. gaditana.**

Les données peuvent être exprimées en nanogrammes, en picomoles ou en pourcentage molaire de lipides totaux.

	A	B	C	D	E	F	G	H	I	J	K
1	Sample Name	SQDG	MGDG	DGDG	PG	PI	PE	PC	DGTS	DAG	TAG
2	Sample_1	1.08E-01	1.07E-01	5.88E-02	5.57E-02	8.20E-02	5.02E-02	9.37E-02	5.69E-02	6.08E-03	4.71E-02
3	Sample_2	1.06E-01	1.02E-01	6.58E-02	6.18E-02	7.90E-02	5.26E-02	1.02E-01	5.87E-02	7.67E-03	5.35E-02
4	Sample_3	1.06E-01	1.04E-01	7.13E-02	4.96E-02	6.37E-02	5.78E-02	9.40E-02	7.36E-02	9.05E-03	1.86E-01
5	Sample_4	5.53E-02	7.47E-02	6.98E-02	2.26E-02	3.29E-02	3.88E-02	5.58E-02	7.37E-02	1.16E-02	3.86E-01

**Table 3.4. Format d'un fichier ForEachLipid du LCPV pour M. gaditana.**

Les données peuvent être exprimées en nanogrammes, en picomoles ou en pourcentage molaire de classe lipidique.

	A	B	C	D	E	F	G
1	Compound Group	Compound Name	Transition	Sample_1	Sample_2	Sample_3	Sample_4
2	SQDG Nanno	SQDG-30-0	765.0 -> 225.0	1.21E+01	1.17E+01	1.16E+01	1.38E+01
3	SQDG Nanno	SQDG-30-1	763.0 -> 225.0	1.97E+00	1.86E+00	1.51E+00	9.63E-01
4	SQDG Nanno	SQDG-32-1	791.0 -> 225.0	8.35E+01	8.41E+01	8.42E+01	8.03E+01
5	SQDG Nanno	SQDG-34-1	819.0 -> 225.0	7.09E-01	6.41E-01	7.50E-01	9.55E-01
6	SQDG Nanno	SQDG-34-2	817.0 -> 225.0	5.18E-01	4.26E-01	5.38E-01	5.29E-01
7	SQDG Nanno	SQDG-34-3	815.0 -> 225.0	3.05E-01	2.48E-01	2.28E-01	2.57E-01
8	SQDG Nanno	SQDG-36-5	839.0 -> 225.0	4.11E-01	4.65E-01	6.16E-01	1.23E+00
9	SQDG Nanno	SQDG-36-6	837.0 -> 225.0	4.94E-01	5.35E-01	6.10E-01	1.56E+00
10							
11							
12							
13				Sample_1	Sample_2	Sample_3	Sample_4
14	MGDG Nanno	MGDG-32-1	746.0 -> 567.0	6.67E+00	7.27E+00	3.87E+00	2.07E+00
15	MGDG Nanno	MGDG-32-2	744.0 -> 565.0	5.39E+00	5.10E+00	1.80E+00	0.00E+00
16	MGDG Nanno	MGDG-34-4	768.0 -> 589.0	5.29E+00	5.17E+00	4.92E+00	5.14E+00
17	MGDG Nanno	MGDG-34-5	766.0 -> 587.0	4.37E+01	4.36E+01	4.39E+01	4.22E+01
18	MGDG Nanno	MGDG-36-4	796.0 -> 617.0	1.13E+00	1.24E+00	1.42E+00	9.98E-01
19	MGDG Nanno	MGDG-36-5	794.0 -> 615.0	8.91E+00	9.46E+00	9.95E+00	1.52E+01
20	MGDG Nanno	MGDG-36-6	792.0 -> 613.0	1.15E+01	1.13E+01	1.04E+01	1.16E+01
21	MGDG Nanno	MGDG-40-10	840.0 -> 661.0	1.49E+01	1.46E+01	1.96E+01	1.76E+01
22	MGDG Nanno	MGDG-40-9	842.0 -> 663.0	2.79E+00	2.65E+00	3.87E+00	4.16E+00
23	...						

Le module lipidomic\_data\_lpcv.py complète automatiquement toutes les données lipidiques du fichier BioModTool, à l'exception de la teneur en lipides totaux, qui doit être saisie manuellement dans le fichier de composition de biomasse. Cette saisie manuelle est nécessaire pour assurer que l'unité de mesure est cohérente avec celle des autres macromolécules de la biomasse. Ce module permet de remplir un même fichier Excel pour plusieurs échantillons, générant ainsi un fichier Excel BioModTool distinct pour chaque échantillon.

## 3.5.2 Utilisation du module lipidomic\_data\_lpcv.py

Le module est utilisable en lignes de code et via une interface Tkinter disponible depuis un exécutable.

### 3.5.2.1 Utilisation en lignes de code

#### Imports

In [1]:

```
import BioModTool.main_add_biomass_objective_function
```

#### 1 - User choice

1.a) Path to data

In [2]:

```
path_to_lpcv_data_class = "path_to_data_repository\\LipidData_ForEachClass.csv"  
path_to_lpcv_data_lipid = "path_to_data_repository\\ForEachLipid.csv"
```

1.b) Path to BioModTool Excel to update

In [3]:

```
path_to_empty_excel =  
"path_to_data_repository\\BiomassComposition_iMgadit_emptyLipids.xlsx"
```

1.c) Chose unit

In [4]:

```
dict_unit = {"g": "g per ...", "mol" : "mol per ..."}
```

In [5]:

```
lpcv_unit_ForEachClass = dict_unit["mol"]  
lpcv_unit_ForEachLipid = dict_unit["mol"]
```

#### 2 - Update BioModTool Excel

In [6]:

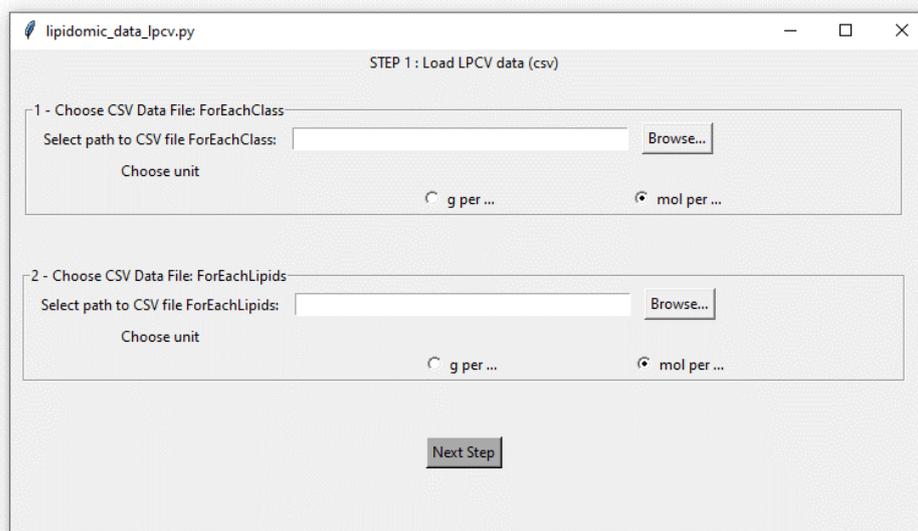
```
for sample_id in ["sample1", "sample2"]:  
df_lpcv_class = load_LPCV_data_ForEachClass(path_to_lpcv_data_class, sample_id)  
df_lpcv_lip = load_LPCV_data_ForEachLipid(path_to_lpcv_data_lipid, sample_id)
```

```
update_BioModTool_excel(df_lpcv_class,  
lpcv_unit_ForEachClass,  
df_lpcv_lip,  
lpcv_unit_ForEachLipid,  
path_to_empty_excel,  
sample_id,  
update_ForEachClass= True,  
update_ForEachLipid= True)
```

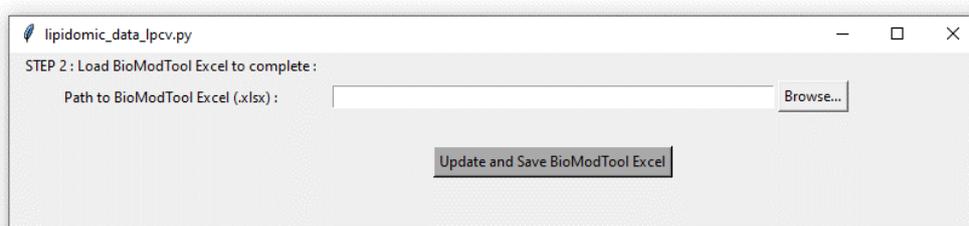
### 3.5.2.2 Utilisation via l'interface graphique

L'interface graphique d'utilisation du module lipidomic\_data\_lpcv.py est en cours de développement.

**Etape 1 :** Charger les données de lipidomique au format CSV et renseigner l'unité des données.



**Etape 2 :** Charger le fichier Excel BioModTool à compléter.



Une fois les deux étapes réalisées, BioModTool peut être utilisé de façon classique comme présenté dans la section 3.4.

### 3.5.3 Validation du module lipidomic\_data\_lpcv.py

Le module a été testé et validé sur plusieurs jeux de données du LPCV. La correspondance entre les données d'entrée au format du LPCV et les données de sortie du module lipidomic\_data\_lpcv.py au format BioModTool a été manuellement vérifiée. Les fichiers de sorties du module sont utilisables par BioModTool et permettent de générer des réactions de biomasse fonctionnelles. Le module combiné à BioModTool a notamment été utilisé pour ajouter les réactions de biomasse BIOMASS\_biomass\_WT526\_c et BIOMASS\_biomass\_MgACSBG31\_c à iMgadi23.

## **3.6 Avantages et limites des solutions techniques implémentées dans BioModTool**

Divers besoins associés à l'ajout de fonctions de biomasse aux GEM ont été identifiés (section 3.2), des solutions techniques ont été proposées (section 3.3) et implémentées dans BioModTool. Les avantages et les limites des solutions techniques retenues pour l'implémentation de BioModTool ainsi que des pistes d'amélioration sont présentés dans la Table 3.5.

**Table 3.5. Avantages, limites et propositions d'améliorations des solutions techniques implémentées dans BioModTool.**

Spécifications	Solutions techniques proposées	Avantages	Limites	Pistes d'amélioration (rapides à mettre en œuvre)
Automatisation de l'ajout d'une fonction de biomasse à un GEM.	Développement d'un script python, utilisable par lignes de commandes permettant d'automatiser les étapes de calculs et de création d'une fonction de biomasse.	Python est un langage communément utilisé en bioinformatique. La librairie COBRApy est open source et très bien documentée.	Dépendance de BioModTool à la librairie COBRApy.	
	Possibilité d'ajouter plusieurs fonctions de biomasse à un même modèle. Chaque fonction de biomasse est identifiée par un suffixe.			
Exploitation automatique des données lipidomiques au format du LPCV.	Développement d'un module spécifique convertissant les données du format LPCV au format BioModTool. Le module est accompagné d'un fichier CSV établissant la correspondance entre la nomenclature du LPCV et celle de iMgadi23.	L'outil est adapté aux besoins du laboratoire. L'utilisateur n'a pas besoin de renseigner les identifiants des lipides de iMgadi23. Même si pour le moment le module n'est applicable qu'à <i>M. gaditana</i> , il est facilement adaptable à d'autres organismes.	L'outil n'est applicable qu'aux données lipidomiques au format du LPCV.	
Origine et format des données flexibles.	Les données sont renseignées manuellement.	Aucune contrainte quant à l'origine des données.	Nécessite de regrouper manuellement l'ensemble des données de composition de biomasse, ce qui prend du temps et peut être source d'erreur. La recherche de données est également manuelle et peut se révéler fastidieuse.	
Flexibilité des unités des données de composition de la biomasse.	Choix parmi deux unités : - g per ... - mol per ...	Permet d'utiliser toutes les unités de concentrations massiques et molaires. Cette flexibilité du choix de l'unité évite à l'utilisateur de devoir convertir ses données (souvent source d'erreur).	Potentielles pertes d'informations (si l'utilisateur ne spécifie pas dans la colonne "Commentaires" l'unité précise de ses données.	
	Ces unités doivent être identiques au sein d'une pseudo-réaction mais peuvent différer entre les pseudo-réactions.		L'utilisateur doit potentiellement faire quelques conversions afin d'unifier les unités des métabolites consommés par une même pseudo-réaction.	

Spécifications	Solutions techniques proposées	Avantages	Limites	Pistes d'amélioration (rapides à mettre en œuvre)
Flexibilité de la structure et de la composition de la fonction de biomasse.	Possibilité de générer une fonction de biomasse comportant jusqu'à trois niveaux hiérarchiques.	Permet la création de fonctions de biomasse conformes aux standards observés dans les GEM.	Complexifie l'outil, l'utilisateur doit renseigner le niveau de chaque pseudo-réaction.	
	Si la fonction de biomasse est organisée en plusieurs niveaux, les macromolécules consommées sont flexibles. Une pseudo-réaction de niveau 2 doit être ajoutée au modèle pour chaque macromolécule consommée dans la fonction de biomasse. L'ajout d'une pseudo-réaction est associée à la création d'une nouvelle feuille de calcul dans le fichier Excel.	Permet à BioModTool d'être utilisé pour différents modèles et divers organismes.	Complexifie l'utilisation de BioModTool et peut être source d'erreur.	
	Ajout d'une deuxième table dans le fichier Excel permettant d'ajouter des métabolites de niveau intermédiaire et avancé (Table 2).	Permet de générer des réactions de biomasses plus ou moins complètes et complexes en fonction des données disponibles.	Les coefficients de la Table 2 ne sont pas convertis par BioModTool. Les données doivent être renseignées dans des unités précises. L'unité n'est pas la même selon le niveau de la pseudo-réaction : mmol/g <sub>DW</sub> pour le niveau 1, mol/mol <sub>macromolécule</sub> pour les autres niveaux. Cela peut être source d'erreur de la part de l'utilisateur.	Les unités suivantes : g/g <sub>DW</sub> pour le niveau 1 et g/g <sub>pool</sub> pour les niveaux 2 et 3 pourraient être proposées à l'utilisateur. Les conversions d'unité pourraient être réalisées par BioModTool en utilisant la masse molaire des métabolites. Cette fonctionnalité ne serait disponible que pour les modèles dans lesquels la formule des métabolites est renseignée.
Outil facile d'utilisation.	Disponible sur GitHub : <a href="https://github.com/Total-RD/BioModTool">https://github.com/Total-RD/BioModTool</a>	L'installation est simple ce qui permet de maximiser le nombre d'utilisateurs potentiels.		
	Développement de scripts Jupyter Notebook.	Permet une certaine automatisation (ajout de plusieurs réactions en simultané). Plus reproductible qu'un ajout de biomasse réalisé via l'interface.		
	Développement d'une interface graphique utilisable directement à partir d'un exécutable.	BioModTool peut être utilisé sans aucune ligne de code. Cela permet de minimiser les erreurs de la part des utilisateurs et par conséquent de maximiser le nombre d'utilisateurs potentiels.	Complexifie l'outil.	

Spécifications	Solutions techniques proposées	Avantages	Limites	Pistes d'amélioration (rapides à mettre en œuvre)
Outil facile d'utilisation. <i>(suite)</i>	Développement d'un fichier Excel structuré et modifiable que l'utilisateur doit compléter avec ses données.	Simple d'utilisation. Présente un compromis entre les besoins de flexibilité de l'outil et les besoins d'automatisation de la création de réactions de biomasse.	Le remplissage du fichier Excel se fait manuellement et peut être source d'erreur. La modification de l'organisation d'une feuille de calcul (déplacements des Tables 1 et 2, renommage des colonnes, modification des dépendances entre les cellules etc.) peut empêcher le bon fonctionnement de BioModTool. Cela limite l'automatisation de l'utilisation de l'outil (un fichier Excel doit être créé pour chaque fonction de biomasse à ajouter au modèle).	
Limiter les erreurs des utilisateurs.	Pour essayer de limiter les erreurs possibles, le fichier Excel est structuré et son remplissage est guidé. Le choix des unités se fait par liste déroulante. Les cellules à modifier sont identifiables par un code couleur. Les cellules destinées à recevoir les coefficients sont soumises à des critères de validation. Les cellules qui n'ont pas à être modifiées par l'utilisateur sont verrouillées.			
	Création de plusieurs fonctions de test des entrées de l'utilisateur.	Identifie les erreurs classiques et en informe l'utilisateur.		

Spécifications	Solutions techniques proposées	Avantages	Limites	Pistes d'amélioration (rapides à mettre en œuvre)
Applicables à divers GEM.	Les identifiants des métabolites doivent être renseignés par l'utilisateur en fonction du modèle choisi.	Permet l'application de l'outil à de nombreux GEM.	Nécessite une interaction utilisateur qui peut être source d'erreur. L'utilisateur doit connaître les identifiants du modèle correspondant aux métabolites. Cette étape peut s'avérer fastidieuse.	
	Possibilité d'ajouter une fonction de biomasse aux modèles comportant ou non des formules et des charges associées aux métabolites.	Permet d'élargir le nombre de modèles avec lesquels BioModTool peut être utilisé.	L'absence de formule empêche le calcul de la masse molaire réduisant les possibilités de normalisation et de conversion des données par BioModTool. Pour les modèles dans lesquels la formule des métabolites n'est pas renseignée l'utilisateur doit lui-même effectuer la conversion de certaines données dans les bonnes unités et effectuer la normalisation des coefficients de la réaction de niveau 1. Ce qui complexifie l'utilisation de BioModTool et peut être source d'erreur.	Une piste d'amélioration serait de permettre à l'utilisateur de renseigner des valeurs de masse molaires des métabolites consommés dans la biomasse afin d'avoir l'ensemble des données nécessaires aux différents calculs de normalisation et de conversion d'unité permettant ainsi de rendre BioModTool totalement fonctionnel avec les GEM sans formules.

## 3.7 Discussion et perspectives

La définition précise de la fonction de biomasse est une étape cruciale dans le développement des GEM (Bernstein et al. 2021). Ces dernières années, plusieurs outils et logiciels dédiés à la reconstruction automatique des GEM ont été développés, accélérant et facilitant le processus de reconstruction de ces modèles (Mendoza et al. 2019). Cependant, les GEM reconstruits automatiquement contiennent généralement une fonction de biomasse qui n'est pas spécifique à l'organisme d'intérêt. Au lieu de cela, la fonction de biomasse d'un organisme modèle, qui peut être plus ou moins proche phylogénétiquement, est souvent utilisée comme référence. Par exemple, les fonctions de biomasse d'*Escherichia coli* et de *Bacillus subtilis* sont respectivement fréquemment utilisées dans les GEM de bactéries à Gram négatif et à Gram positif (Bernstein et al. 2021).

De nombreuses études ont mis en évidence l'impact significatif de la composition de la biomasse sur les prédictions des GEM (Pramanik et Keasling 1998; Nookaew et al. 2008; Lavoie et al. 2020; Schulz et al. 2021). De plus, les GEM supposent généralement une composition de biomasse constante, tandis qu'*in vivo*, la composition de la biomasse varie en réponse aux conditions environnementales changeantes, telles que la disponibilité des nutriments, la température, le stress, etc. (Schaechter, Maaløe et Kjeldgaard 1958; Cotner, Makino et Biddanda 2006; Ferrer-Ledo et al. 2023; Simionato et al. 2013; Marudhupandi, Sathishkumar et Kumar 2016). Récemment des protocoles expérimentaux ont été proposés pour répondre aux besoins de définition des fonctions de biomasse (Beck, Hunt et Carlson 2018; Szélieová et al. 2020; Simensen et al. 2022). Cependant, une fois les données générées, leur conversion dans des unités appropriées, l'instanciation d'une fonction de biomasse et son ajout au modèle d'intérêt requièrent non seulement une certaine expertise dans l'utilisation des outils de manipulation des GEM, mais également un investissement de temps.

BioModTool est un outil informatique polyvalent permettant de répondre à ces problématiques.

### 3.7.1 BioModTool permet la valorisation de données variées

La polyvalence de BioModTool se manifeste à travers sa compatibilité avec une grande variété de données, qu'elles proviennent de différentes sources ou qu'elles soient exprimées dans diverses unités.

Les données peuvent être obtenues expérimentalement par l'utilisateur ou être issues de la littérature. Avec l'essor des technologies à haut débit, une quantité croissante de données omiques devient accessible. Au cours des dernières années, les données omiques ont été intégrées dans les GEM dans le but d'améliorer leur qualité (Haas et al. 2017; Lachance et al. 2019; Cunha et al. 2023).

Ces données peuvent notamment être utilisées pour déterminer la composition de la fonction de biomasse. BioModTool ne propose pas de fonctionnalité permettant d'exploiter directement des données omiques. Cette fonctionnalité est partiellement proposée par le package BOFdat (Lachance et al. 2019). En effet, BOFdat permet de créer des fonctions de biomasse (de niveaux basiques à avancés) à partir de différents jeux de données omiques. Le calcul des coefficients stœchiométriques par BOFdat est organisé en trois étapes. La première étape correspond à la définition de la composition basique de la biomasse. Elle requiert la connaissance de la composition de la biomasse en termes de fractions massiques de l'ADN, de l'ARN, des protéines et des lipides. Ces fractions massiques doivent être fournies par l'utilisateur. La composition en monomères de chaque macromolécule est ensuite calculée à partir : du génome pour l'ADN ; de l'annotation du génome (au format GenBank) et respectivement de données transcriptomiques et protéomiques pour l'ARN et les protéines ; et des données lipidomiques pour les lipides. Des données de croissance et de consommation de substrat sont utilisées pour déterminer l'énergie de maintenance. La deuxième étape permet d'identifier les coenzymes, les ions inorganiques à inclure dans la biomasse. Enfin, lors de la troisième étape les métabolites spécifiques à l'organisme sont déterminés, nécessitant notamment des données d'essentialité de gènes. L'ensemble des données requit par BOFdat n'est pas toujours disponible pour l'organisme d'intérêt, ni dans les conditions (conditions de culture, ou souche mutante) souhaitées. Cette limitation restreint ainsi les domaines d'application de BOFdat. Cependant, les trois étapes de BOFdat sont indépendantes. De plus, lors de l'étape 1, les calculs des compositions en monomères de chaque macromolécule sont effectués au moyen d'une fonction distincte et autonome. Par exemple la composition en acides aminés peut être déterminée à partir de l'annotation GenBank, d'un jeu de données protéomique et du modèle métabolique de l'organisme d'intérêt via la fonction *generate\_protein\_coefficients*. En conséquence, un utilisateur cherchant à élaborer une fonction de biomasse peut utiliser les fonctions pertinentes de BOFdat pour obtenir certains coefficients stœchiométriques, puis compléter les données manquantes à l'aide de données expérimentales complémentaires dont il dispose. Il peut ensuite utiliser BioModTool pour normaliser ses données et créer une fonction de biomasse.

BioModTool offre ainsi la possibilité de valoriser directement ou indirectement de nombreux jeux de données. Une perspective d'amélioration de BioModTool relativement simple à mettre en œuvre serait de permettre l'utilisation de certaines fonctions de BOFdat. Ces fonctions pourraient notamment être utilisées pour déterminer automatiquement la composition nucléotidique des acides nucléiques et la composition en acides aminés des protéines à partir de données omiques. Cette intégration pourrait être accomplie en important ou en adaptant des fonctions de BOFdat dans BioModTool.

### 3.7.2 BioModTool est applicable à divers GEM d'organismes variés

Une très grande majorité des GEM appartiennent au règne bactérien (Gu et al. 2019; Ye et al. 2022). De nombreux outils sont donc adaptés voire spécifiques aux GEM de bactéries. C'est le cas de BOFdat par exemple. En effet, la détermination des ions inorganique est réalisée à partir de données issues d'une étude comparative des fonctions de biomasses de 72 GEM procaryotes (Xavier, Patil et Rocha 2017). De plus, la composition basique de la fonction de biomasse générée par BOFdat se limite à l'ADN, l'ARN, les protéines et les lipides. BOFdat n'est donc pas adapté à la génération de fonctions de biomasses pour les GEM d'organismes photosynthétiques (qu'ils soient procaryotes ou eucaryotes), nécessitant l'ajout de métabolites spécifiques notamment des carbohydrates et des pigments.

De par sa flexibilité BioModTool ne présente pas de restriction quant à la phylogénie des organismes auxquels il peut être appliqué.

BioModTool est donc applicable à une grande diversité de GEM. Toutefois, les fonctionnalités de BioModTool sont restreintes lorsque le modèle métabolique à modifier ne contient pas les formules chimiques de tous les métabolites nécessaires à la fonction de biomasse. En effet, l'utilisation de BioModTool implique de nombreux calculs de normalisation et de conversion d'unités. Ces opérations nécessitent la connaissance de la masse molaire des métabolites. BioModTool calcule la masse molaire à partir de la formule chimique des métabolites en utilisant la fonction *formula\_weight* de COBRapy. Ainsi, en l'absence d'une formule chimique pour un métabolite, il devient impossible de calculer sa masse molaire, et par conséquent, de déterminer son coefficient stœchiométrique. Une piste d'amélioration potentielle pour BioModTool consisterait à permettre à l'utilisateur de renseigner manuellement les valeurs de masse molaire des métabolites consommés dans la biomasse. Cette approche fournirait toutes les données requises pour effectuer les divers calculs de normalisation et de conversion, permettant ainsi à BioModTool de fonctionner de manière optimale même en l'absence de formules chimiques dans le modèle métabolique.

### 3.7.3 Élargissement des applications des GEM disponibles pour les non-modeleurs.

De très nombreux logiciels de reconstruction et d'utilisation des GEM sont disponibles. Certains de ces logiciels ont été conçus de manière à ne pas exiger de compétences avancées en bioinformatique de la part de leurs utilisateurs. C'est par exemple le cas d'Escher qui permet de visualiser et d'explorer le contenu des modèles ainsi que de réaliser des simulations de flux par FBA (Escher-FBA) (King et al. 2015; Rowe, Palsson et King 2018). Même s'il existe plusieurs logiciels permettant de facilement

modifier des GEM tel que MetExplore (Cottret et al. 2010; Cottret et al. 2018), le calcul des coefficients de biomasse requiert une solide compréhension des concepts et des normes associés aux GEM. BioModTool permet de pallier ces limitations. En effet, bien qu'il repose sur la librairie COBRAPy, l'utilisation de BioModTool ne requiert pas de compétences en bioinformatique. L'utilisateur remplit un fichier Excel, un logiciel tableur très largement utilisé par les biologistes, avec ses données dans des unités avec lequel il est familier. Ensuite, il peut utiliser l'interface graphique de BioModTool pour ajouter une fonction de biomasse détaillée à un modèle métabolique sans écrire de lignes de code. La mise à disposition d'une interface graphique élargit l'utilisation de BioModTool à un public plus large. Par conséquent, lorsqu'il est associé à des outils de simulation interactifs déjà existants, BioModTool a le potentiel d'étendre davantage les applications des modèles métaboliques.

### 3.7.4 Détermination des identifiants des métabolites

L'utilisation de BioModTool nécessite une étape préliminaire de collecte de données. Lorsqu'il définit la composition de la biomasse, l'utilisateur doit vérifier que l'ensemble des métabolites requis est présent dans le GEM d'intérêt et renseigner leur identifiant dans le fichier Excel. La complexité de cette étape varie significativement en fonction de la qualité du modèle métabolique, en particulier de son niveau d'annotation et de la nomenclature employée pour les identifiants. Des outils automatisant et facilitant cette recherche sont disponibles mais ils présentent des limitations quant aux modèles auxquels ils peuvent être appliqués. Par exemple, Poupin et al ont récemment proposé un outil permettant d'établir des correspondances entre des identifiants de données et les identifiants des modèles métaboliques, notamment pour les lipides (Poupin et al. 2020). Cependant ce package nécessite la présence des identifiants ChEBI dans les annotations du modèle.

### 3.7.5 Perspectives d'application à d'autres organismes du laboratoire

Le module `lipidomic_data_lpcv.py` permet de compléter automatiquement le fichier Excel de BioModTool à partir de fichiers standardisés de données lipidomiques. Actuellement, ce module n'est compatible qu'avec iMgadit23, modèle métabolique de *M. gaditana*. Grâce à ce module, les profils lipidiques de diverses souches de *M. gaditana* cultivées dans différentes conditions ont pu être intégrés à iMgadit23.

D'autres organismes font l'objet d'analyses régulières au sein du LPCV tels que *Phaeodactylum*, *Chlamydomonas* ou *Arabidopsis*. Des GEM sont disponibles pour ces organismes (Cheung et al. 2013; Levering et al. 2016; Mora Salguero et al. 2018). Le module `lipidomic_data_lpcv.py` pourrait facilement

être adapté pour permettre l'intégration automatique des données lipidomiques du LPCV dans les fonctions de biomasse des GEM de ces organismes.

### 3.7.6 Perspectives d'automatisation de BioModTool

Les données relatives à la composition de la biomasse doivent être introduites manuellement dans le fichier Excel de BioModTool. Cependant, il est possible de développer des fonctions et des modules supplémentaires pour automatiser cette tâche en complétant automatiquement le fichier Excel à partir de données standardisées. Le module LPCV constitue un bon exemple de l'automatisation potentielle de BioModTool.

## 3.8 Conclusion

En permettant la génération de fonctions objectives de biomasse, BioModTool constitue un outil précieux pour la reconstruction et l'amélioration des GEM. Les conditions environnementales et génétiques influent significativement sur la composition de la biomasse des organismes. Des données de composition de biomasse sont souvent disponibles pour un même organisme dans différentes conditions. BioModTool permet de valoriser ces données, indépendamment de leur format, et de représenter au sein d'un même modèle des compositions de biomasse différentes.

BioModTool offre une flexibilité non seulement en termes de sources utilisables pour obtenir des données de composition de la biomasse (littérature, mesures expérimentales etc.), mais aussi en termes de structure et de composition de la fonction de biomasse. Cette flexibilité rend BioModTool utilisable dans de nombreuses situations. De plus, la disponibilité d'une interface graphique étend l'utilisation de BioModTool aux non-modélisateurs.

## 4 Chapitre 4

---

### Conclusion générale et Perspectives

Les microalgues représentent des ressources au potentiel intéressant pour diverses applications industrielles. Parmi ces organismes, des espèces oléagineuses telles que *Microchloropsis gaditana*, suscitent un intérêt particulier dans le domaine de la production de biocarburants en raison de leur capacité à accumuler des lipides, notamment des TAG, en réponse à une carence en azote ou en phosphate. Cependant, notre connaissance et notre compréhension du métabolisme de cette souche présentent de nombreuses lacunes à combler pour répondre aux exigences de l'industrie (Lupette et Maréchal 2018).

Dans ce contexte, l'objectif de cette thèse était de développer et d'appliquer des approches bioinformatiques pour approfondir notre compréhension du métabolisme de *M. gaditana*, en mettant particulièrement l'accent sur son métabolisme lipidique. La thèse a abouti à un modèle métabolique de *M. gaditana* : iMgadi23, qui décrit précisément les connaissances les plus avancées sur le métabolisme lipidique de cette souche. De plus, BioModTool, un package Python, a été développé pour faciliter l'intégration de fonctions de biomasse dans des modèles métaboliques. Une version spécialement adaptée au format des données lipidomiques du LPCV, nommée BioModTool-LPCV, a également été conçue. BioModTool-LPCV a été utilisé pour incorporer le profil lipidique détaillé d'un mutant généré en laboratoire, le mutant *acsbg*, à iMgadi23. Ce mutant présente un profil lipidique intéressant, notamment avec une accumulation de triacylglycérols (TAG) huit fois plus élevée que celle de la souche sauvage (Billey et al. 2021). Des simulations de distributions de flux ont été réalisées pour ce mutant et comparées aux flux prédits pour la souche sauvage. Les résultats de cette analyse *in silico* sont cohérents avec les conclusions de l'analyse expérimentale effectuée sur le mutant *acsbg* par Billey et al..

Dans ce chapitre, nous commencerons par présenter comment les résultats obtenus contribuent à répondre aux différents objectifs de la thèse. Les avantages et les limites de chaque outil pris individuellement ayant déjà été discutés précédemment (dans le [Chapitre 2](#) pour iMgadi23 et dans le [Chapitre 3](#) pour BioModTool), nous ne les reprendrons pas en détail ici. Ensuite, nous explorerons diverses perspectives d'applications notamment pour le modèle métabolique.

## 4.1 Conclusions relatives aux objectifs du projet de thèse

Cette section présente comment les résultats obtenus au cours de la thèse contribuent à répondre aux objectifs du projet, à savoir :

- Intégrer les connaissances les plus avancées du métabolisme lipidique de *M. gaditana* provenant des experts du LPCV et de la littérature, dans le modèle métabolique reconstruit par SilicoLife.
- Permettre aux biologistes du LPCV et de TotalEnergies d'exploiter le modèle métabolique de *M. gaditana*. Les biologistes doivent pouvoir facilement visualiser le contenu du modèle, modifier les compositions de biomasse selon les profils lipidiques mesurés, et réaliser des simulations de flux dans différentes conditions (différentes conditions de culture, pour différentes souches).
- Exploiter le modèle métabolique pour interpréter des profils lipidiques détaillés de *M. gaditana* et de ses mutants, et relier ces profils à des hypothèses sur leur fonctionnement métabolique.

Des perspectives d'applications à court terme, reposant majoritairement sur les approches développées au cours de la thèse sont également présentées.

### 4.1.1 Correction du métabolisme lipidique de iMgadit et intégration de données lipidomiques

Avant le début de cette thèse, un modèle métabolique de *M. gaditana* avait été reconstruit par SilicoLife. Cependant, en raison du manque de caractérisation du métabolisme lipidique de cette souche et du processus de reconstruction majoritairement automatique, la première version du modèle (iMgadit\_SilicoLife) présentait de nombreuses incohérences. Le premier objectif de cette thèse était donc d'intégrer dans le modèle métabolique de *M. gaditana*, les connaissances les plus avancées de son métabolisme. Les corrections apportées se sont concentrées sur le métabolisme des acides gras et des glycérolipides, car ces lipides sont les plus abondants et donc les mieux caractérisés chez *M. gaditana*. La première étape de cette correction a consisté à établir une nomenclature claire pour les identifiants des métabolites et des réactions, en se basant sur les normes couramment utilisées dans la communauté Lipide (IUPAC 1978; Liebisch et al. 2020), tout en respectant les normes et directives de BiGG (King et al. 2016). Par la suite, toutes les voies métaboliques de synthèse et de dégradation des acides gras et des glycérolipides ont été manuellement corrigées en s'appuyant sur l'expertise des biologistes du LPCV et sur la littérature scientifique. Cette correction a nécessité l'ajout

de nombreuses réactions manquantes et la rectification des localisations intracellulaires des différentes voies métaboliques. La diversité des acides gras et des glycérolipides en termes de longueur de chaînes et de degré d'insaturation, a également été minutieusement corrigée à partir de profils lipidomiques détaillées de *M. gaditana* (Jouhet et al. 2017). Dans la version finale du modèle, la modélisation du métabolisme lipidique intègre ainsi un total de 1017 réactions. De plus, la photosynthèse, le cycle du glyoxylate, la photorespiration, le métabolisme de l'azote ont également été vérifiés et corrigés. Ainsi, l'ensemble des corrections réalisées au cours de la thèse a nécessité l'ajout de 1108 réactions, 682 métabolites et 58 gènes. 668 réactions, 604 métabolites et 33 gènes ont été supprimés. Les identifiants de 302 réactions et 224 métabolites ont été modifiés, pour correspondre à la nomenclature du métabolisme lipidique proposée dans le [Chapitre 2](#). Enfin, la composition de la biomasse a été corrigée en utilisant des données provenant du laboratoire, notamment un glycérolipidome détaillé (Billey et al. 2021) et des données extraites de la littérature scientifique (Volkman et al. 1993; Radakovits et al. 2012; Simionato et al. 2013; Scholz et al. 2014; Medina et al. 2015; Tibbetts, Bjornsson et McGinn 2015; Finkel et al. 2016; Shah et al. 2017; Janssen, Kastenhofer et al. 2018; Lamminen et al. 2019). Ces diverses corrections font du métabolisme lipidique décrit dans iMgadi23 l'un des plus complets et détaillés parmi les modèles de microalgues (Correa et al. 2020).

La validation expérimentale des prédictions générées par iMgadi23 a mis en évidence sa bonne précision prédictive et son potentiel en tant qu'outil permettant de formuler des hypothèses pour orienter la conception de futures d'expériences, et aider à une meilleure compréhension du métabolisme de *M. gaditana*.

L'outil BioModTool permet par ailleurs d'intégrer des données de composition de biomasse, y compris des données lipidomiques, au sein de iMgadi sous forme de fonction de biomasse. Des fonctions de biomasse spécifiques de contextes donnés peuvent ainsi être générées et utilisées comme objectif de simulation, élargissant ainsi les perspectives d'utilisation du modèle métabolique.

De plus, afin d'étendre son potentiel applicatif à un plus large panel d'utilisateurs, une carte métabolique détaillée du contenu de iMgadi23 a été générée.

#### 4.1.2 Potentiel applicatif des méthodes développées pour les biologistes

Malgré leur popularité, les modèles métaboliques à base de contrainte ne sont pas faciles à comprendre et à utiliser pour les non-spécialistes (Zhang et Hua 2015). Les codes et scripts complexes couramment utilisés pour les applications et les analyses informatiques basées sur ces modèles ne sont ainsi pas aisément accessibles à la communauté des biologistes.

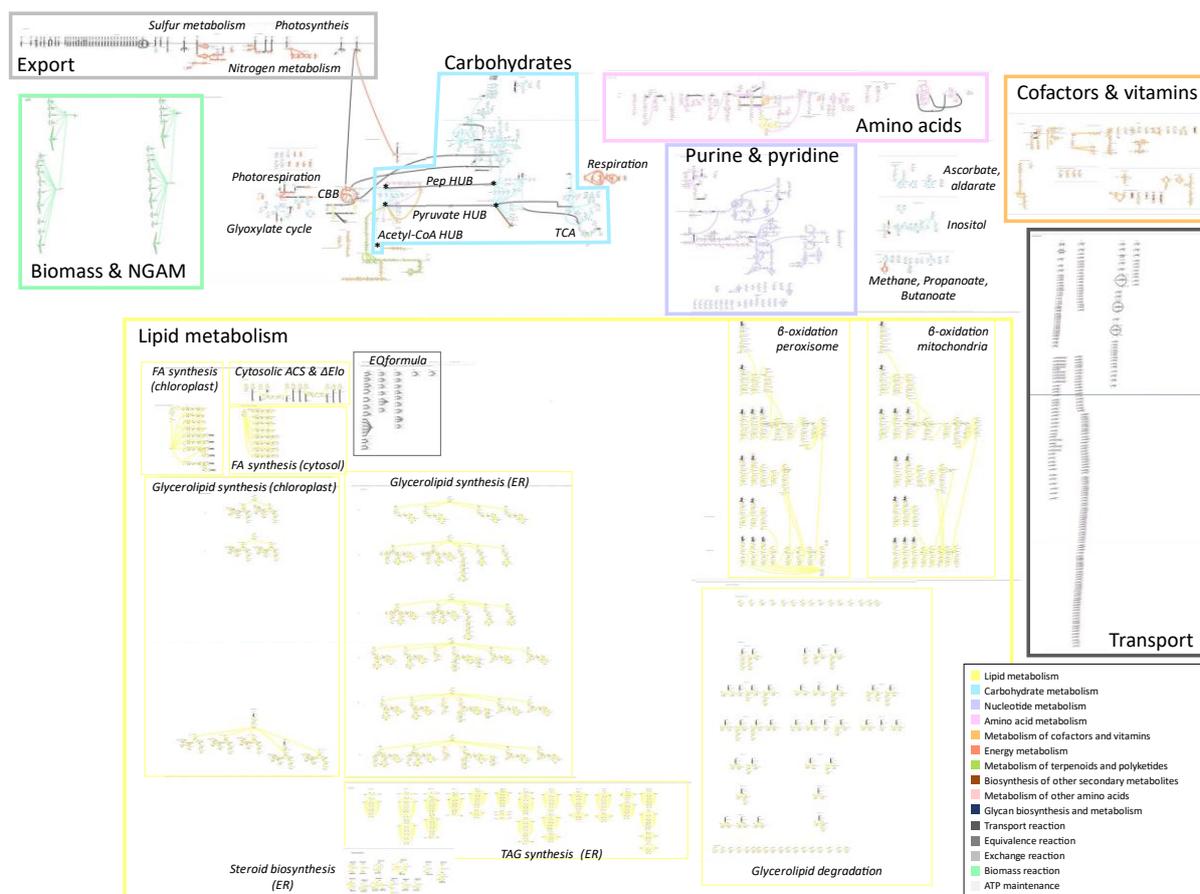
Ainsi, un autre objectif de la thèse consistait à rendre iMgadit ainsi que BioModTool accessibles aux biologistes, notamment de TotalEnergies et du LPCV. La prise en main du modèle par des experts du métabolisme, devrait favoriser une meilleure appréhension des limites de ce dernier. Cette prise en main du modèle englobe l'exploration de son contenu ainsi que la possibilité de réaliser et d'interpréter des distributions de flux. La création d'une carte métabolique exhaustive d'iMgadit23 et le développement de l'interface graphique de BioModTool ont été réalisés pour répondre à cet objectif.

#### 4.1.2.1 La carte métabolique permet d'explorer le contenu de iMgadit

Les GEM comportent des milliers de composants. Par exemple, iMgadit23 compte 1977 métabolites impliqués dans 2330 réactions interconnectées et associées à 889 gènes. La représentation visuelle de réseaux métaboliques est donc essentielle pour faciliter leur analyse et leur exploitation. Les cartes métaboliques bidimensionnelles (2D) sont couramment employées comme outils de visualisation du métabolisme. Plusieurs de ces cartes métaboliques sont devenues des ressources incontournables au sein de la communauté scientifique, telles que celles proposées par BioCyc (Caspi et al. 2018), ChloroKB (Gloaguen et al. 2017; Gloaguen et al. 2021), KEGG pathway maps (Kanehisa et Goto 2000b), Roche Biochemical Pathways (Gerhard 2018) et WikiPathways (Kelder et al. 2012; Kutmon et al. 2016). Ces cartes offrent une représentation fonctionnelle des réseaux métaboliques, décrivant chaque réaction et les métabolites impliqués, pouvant également inclure des informations sur les enzymes catalysant ces réactions, leur(s) inhibiteur(s) potentiels et la localisation chromosomique des gènes codant pour ces enzymes.

En ce qui concerne les GEM, l'application Web Escher est largement utilisée pour la visualisation de leur contenu (King et al. 2015). Escher permet de créer des cartes métaboliques à partir des informations contenues dans un modèle. Le format d'entrée d'Escher est le format JSON (JavaScript Object Notation), qui est l'un des formats standards de sauvegarde des GEM, proposé notamment par les bibliothèques COBRA (Ebrahim et al. 2013; Heirendt et al. 2019; Norsigian et al. 2020)). Escher étant une application web, son déploiement est rapide, éliminant ainsi le besoin de télécharger une application ou un plug-in de navigateur. De plus, lorsqu'une carte métabolique est disponible, Escher offre une gamme de possibilités pour visualiser divers types de données sur cette carte ainsi que la possibilité d'effectuer des prédictions de distributions de flux de manière "dynamique" (on-the-fly) (Rowe, Palsson et King 2018). Ces fonctionnalités sont détaillées dans les sections suivantes. L'utilisation d'Escher est intuitive et très bien documentée : <https://escher.readthedocs.io/en/latest/>. Par conséquent, Escher a été sélectionnée pour la création et l'exploitation d'une carte métabolique exhaustive du contenu de iMgadit23, organisée en voies métaboliques (Figure 4.1). Cette carte offre

une visualisation complète de l'ensemble des réactions biochimiques constituant le modèle, ainsi que des différents gènes associés à ces réactions. Elle offre ainsi aux experts en métabolisme du LPCV et de TotalEnergies la possibilité d'explorer le contenu du modèle pour identifier des axes de correction et d'amélioration. Elle présente également un potentiel pédagogique, permettant aux non-spécialistes de se familiariser avec le métabolisme de *M. gaditana*, en particulier son métabolisme lipidique qui n'est pas simple à appréhender. Des cartes plus restreintes, obtenues par subdivision de la carte principale, sont également disponibles.



**Figure 4.1. Vue d'ensemble de la carte métabolique de iMgadit23.**

Les réactions sont colorées par voies métaboliques (sous-système). Pour le métabolisme lipidique, la localisation des voies métaboliques est précisée.

La carte métabolique de iMgadit, constitue ainsi un très bon outil pour les biologistes sans compétence bioinformatique avancée qui souhaiteraient étudier le contenu du réseau métabolique développé dans le cadre de cette thèse. L'exploration du contenu du modèle par des experts du métabolisme devrait contribuer à l'identification de potentielles corrections et pistes d'amélioration à apporter à ce dernier. De plus, avec les progrès/découvertes dans la caractérisation du métabolisme de *M. gaditana*, des modifications devront être apportées au modèle dans le futur. Ces modifications

peuvent être réalisables *via* la librairie COBRA, tel que cela a été réalisé au cours de la thèse. Toutefois, des outils plus conviviaux sont également disponibles et accessibles à la communauté scientifique. C'est le cas par exemple de MetExplore qui permet de corriger et d'annoter des modèles métaboliques dans un environnement collaboratif simple d'utilisation, permettant à un large éventail d'utilisateurs d'apporter des modifications aux modèles (Cottret et al. 2010; Cottret et al. 2018). Cet outil pourrait être utilisé par les collaborateurs de TotalEnergies et du LPCV pour l'amélioration du contenu de iMgadit.

En plus de permettre la visualisation et l'exploration de iMgadit, la carte métabolique permet de visualiser différents jeux de données omiques.

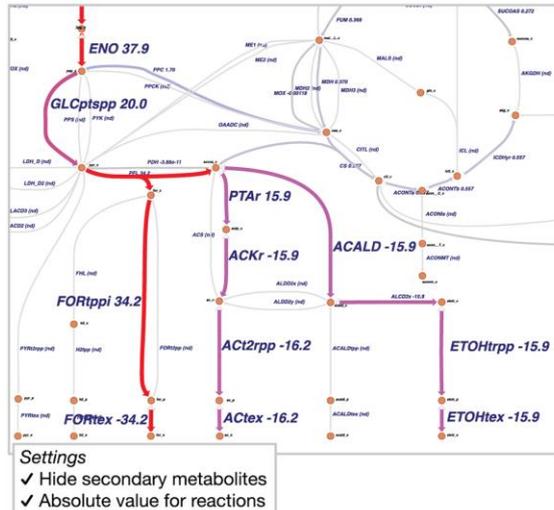
#### 4.1.2.2 La carte métabolique : support pour la visualisation, et l'interprétation de différents jeux de données

L'avènement des technologies omiques a conduit à une abondance croissante de données biologiques à grande échelle. En conséquence, il existe une demande importante d'outils et de méthodes de visualisation de ces données visant à faciliter leur interprétation (VIZBI).

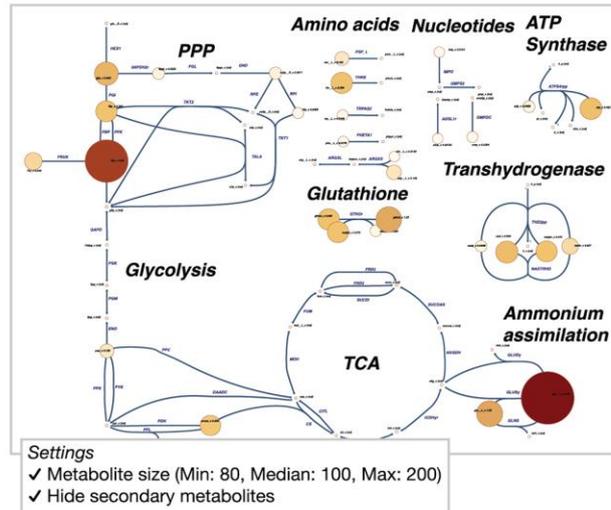
Dans le domaine des réseaux métaboliques, Escher permet la visualisation de diverses données omiques sur une carte métabolique. En effet, Escher peut-être utilisée pour visualiser des données sur les réactions (flux métaboliques obtenus expérimentalement ou prédits *in silico*) et sur les métabolites (données métabolomiques) (Figure 4.2 A et B). Escher reliant les gènes et les enzymes aux réactions qu'ils catalysent, des données génomiques (données transcriptomiques) peuvent également être visualisées (Figure 4.2C) (King et al. 2015). Toutefois, pour visualiser ces données, une étape de formatage est nécessaire, notamment pour établir un lien entre les identifiants des gènes, des réactions ou des métabolites des données et ceux du modèle.

Ainsi, la carte métabolique de iMgadit23 créée au cours de la thèse peut être exploitée pour toutes ces applications. Elle offre notamment des perspectives de visualisation et d'interprétation de données transcriptomiques de *M. gaditana* actuellement disponibles (Gong et al. 2020), ainsi que des données lipidomiques du LPCV. La carte permet également l'interprétation de données de distributions de flux préalablement prédites par FBA, ou d'autres algorithmes de simulation de flux, avec iMgadit.

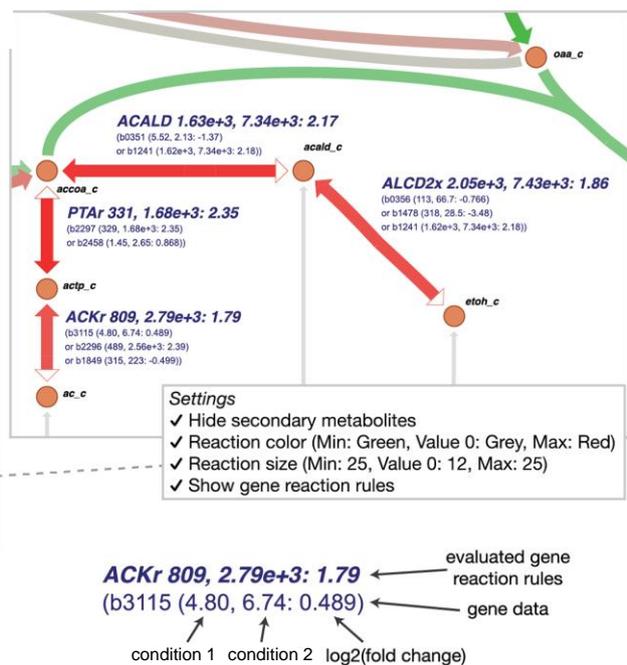
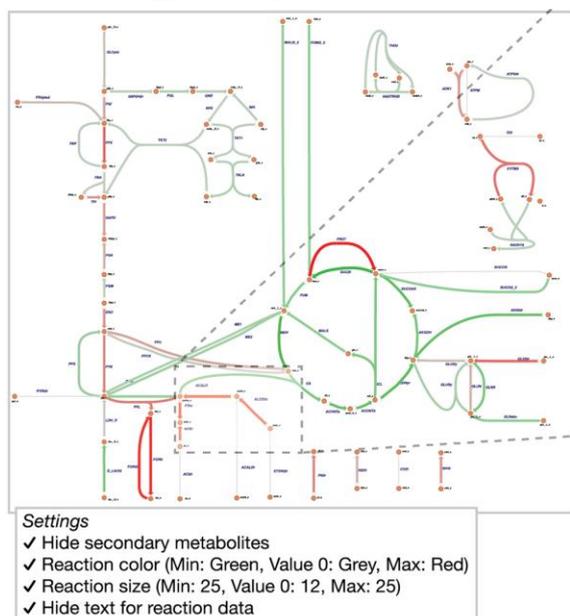
### A Visualizing reaction flux



### B Visualizing metabolite concentration



### C Visualizing transcriptomic



**Figure 4.2. Visualisation de données omiques via Escher (source: King et al. 2015).**

(A) Résultats d'une simulation de flux *in silico* visualisés sur les réactions. L'intensité des flux est codée par une échelle de couleur. (B) Données métabolomiques pour la croissance aérobie d'*E. coli* visualisées sur les métabolites. (C) Données transcriptomiques (RNA-Seq) comparant des conditions aérobie et anaérobie chez *E. coli* visualisées sur les réactions. En fonction de la valeur calculée du log<sub>2</sub>(Fold Change), les réactions régulées à la baisse sont représentées en vert, celles régulées à la hausse sont en rouge. L'ensemble des paramètres de visualisations (indices visuels tels que la taille et la couleur) sont modifiables par l'utilisateur.

4.1.2.3 La carte métabolique : support à la simulation de distributions de flux par FBA  
 Des distributions de flux peuvent être générées directement en exploitant iMgadi et sa carte métabolique, sans aucune ligne de code. En effet, Escher-FBA permet la réalisation "dynamique" de simulations de flux sans écriture de code ni téléchargement de logiciel (Rowe, Palsson et King 2018).

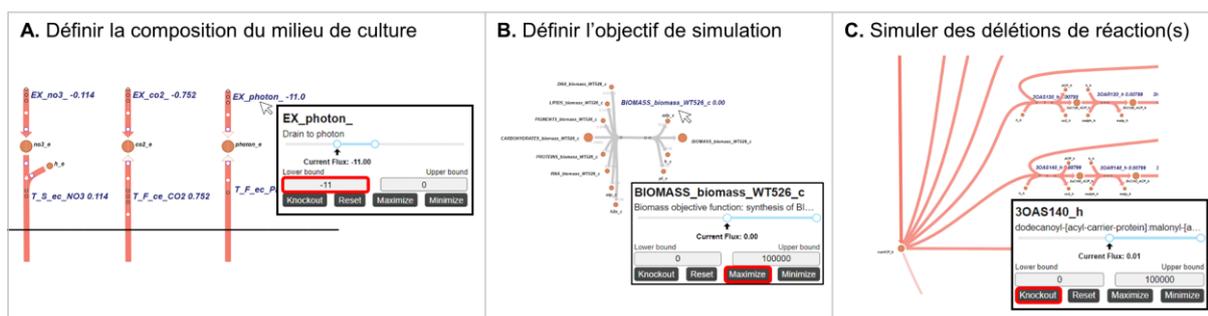
Escher-FBA peut donc être utilisée par les biologistes du laboratoire pour réaliser de telles simulations. Après avoir chargé iMgadi23 et la carte métabolique dans Escher-FBA, l'utilisateur peut réaliser des simulations FBA en jouant sur les paramètres de simulation de manière intuitive et en visualisant immédiatement l'impact sur les résultats de la simulation. Différentes options sont possibles permettant la réalisation de différentes analyses :

(1) L'utilisateur a la possibilité d'appliquer des contraintes de simulation en définissant les limites de flux (lower et upper bounds) des réactions du modèle. Ainsi, il peut simuler la composition du milieu de culture en définissant la disponibilité et les taux de consommation de différents substrats (Figure 4.3A). Par exemple, il peut ajuster le taux de consommation de photons ou de CO<sub>2</sub> disponibles, choisir une source d'azote parmi celles proposées dans le modèle : NO<sub>2</sub>, NO<sub>3</sub>, NH<sub>4</sub>, et urée, etc., définir le type de culture (autotrophie, mixotrophie ou hétérotrophie) en définissant la disponibilité de diverses sources de carbone (glucose, fructose, galactose etc.). Les contraintes appliquées aux autres réactions du modèle sont également modifiables.

(2) L'utilisateur a la possibilité de déterminer l'objectif de simulation. Il peut, par exemple, choisir de maximiser l'une des réactions de biomasse disponibles dans le modèle, de minimiser la consommation d'un substrat spécifique, de minimiser l'ATP maintenance (NGAM, Non Growth Associated ATP Maintenance) etc. (Figure 4.3B).

(3) L'utilisateur peut simuler la délétion d'une ou de plusieurs réactions (en contraignant leur flux à une valeur nulle) (Figure 4.3C).

Après chaque modification d'un paramètre par l'utilisateur, Escher-FBA recalcule automatiquement la distribution de flux. Les résultats (valeur de la fonction objective et valeur du flux pour chaque réaction du modèle) sont immédiatement visibles sur la carte. L'utilisateur peut ensuite analyser l'impact des différents paramètres de simulation sur le phénotype prédit par le modèle.



**Figure 4.3. Simulations par FBA via Escher-FBA.**

Escher-FBA permet de définir les paramètres de simulation et de directement visualiser leurs impacts sur la distribution de flux prédite par FBA sur l'ensemble des réactions présentes sur la carte. L'utilisateur peut ainsi (A) modifier les limites inférieure et supérieure (lower/upper bound) des réactions pour définir le milieu de culture, (B) définir l'objectif de simulation : la réaction du modèle à maximiser ou minimiser, (C) simuler des délétions de réaction(s) (Knockout).

D'autres outils en lignes permettent, à partir d'un modèle métabolique et d'une carte au format JSON, de réaliser des analyses similaires en utilisant d'autres algorithmes de simulation de flux. Par exemple, Caffeine permet de prédire des distributions de flux par pFBA (DD-DeCaF), et IMFLER permet l'analyse de la variabilité des flux par FVA, en plus de proposer la simulation par FBA (Petrovs, Stalidzans et Pentjuss 2021). Toutes ces possibilités de simulations sont réalisables avec les deux fonctions de biomasse de iMgadi23. De plus, grâce à BioModTool, disposant d'une interface graphique, les utilisateurs peuvent facilement ajouter une nouvelle fonction de biomasse à iMgadi23. Ils peuvent ensuite intégrer cette réaction à la carte métabolique et l'utiliser comme nouvel objectif pour la simulation des flux.

Pour conclure, l'exploitation du modèle iMgadi23 ainsi que de sa carte métabolique à travers les différentes applications web présentées dans cette section, notamment Escher et Escher-FBA, offre la possibilité à des utilisateurs non-spécialistes d'explorer le contenu du modèle, d'effectuer et visualiser des simulations de flux, et de visualiser diverses données omiques. Ces applications ouvrent des perspectives d'analyse par un large panel d'utilisateurs, ce qui devrait contribuer *in fine* à améliorer notre connaissance du métabolisme de *M. gaditana*. En effet, les GEM peuvent révéler des prédictions incorrectes, ce qui permet d'identifier les lacunes ou les inexactitudes dans les connaissances métaboliques (Bordbar et al. 2014). Par conséquent, les divergences entre les prédictions de iMgadi23, les données expérimentales et les hypothèses des experts pourront être utilisées pour formuler de nouvelles hypothèses et concevoir des analyses (expérimentales et *in silico*) ciblées visant à corriger ces imprécisions dans les connaissances métaboliques.

Par ailleurs, les outils développés au cours de la thèse peuvent également être exploités pour des analyses plus complexes.

#### 4.1.3 Exploitation du modèle métabolique et de BioModTool pour l'interprétation de profils lipidiques détaillés

La composition de la biomasse de *M. gaditana*, notamment son profil lipidique, est fortement variable en fonction des conditions de culture (Simionato et al. 2013; Kim et al. 2014; Jouhet et al. 2017; Janssen, Kastenhofer et al. 2018; Janssen, Driessen et al. 2018a). Comprendre les raisons et les mécanismes de ces variations est essentiel, notamment pour optimiser les conditions de culture en vue d'améliorer la productivité et les rendements des procédés de culture. Par ailleurs, l'ingénierie métabolique de différentes souches de microalgues, notamment des genres *Nannochloropsis* et *Microchloropsis*, dans le but d'augmenter la synthèse et l'accumulation de lipides est actuellement un

domaine de recherche en plein essor (Al-Hoqani, Young et Purton 2017; Poliner, Farré et Benning 2018; Ghosh et al. 2016; Park, Nguyen et Jin 2019; Ranjbar et Malcata 2022). Ainsi, des souches mutantes de *M. gaditana* ont été générées ces dernières années dans le but d'affiner notre compréhension du métabolisme lipidique de cette souche, et d'améliorer ses rendements de production de lipides (Billey et al. 2021; Dolch et al. 2017; Bo et al. 2021; Cecchin et al. 2020). La plateforme lipidomique du LPCV permet d'effectuer des analyses détaillées du profil lipidique de divers organismes, y compris *M. gaditana* pour répondre à ces différentes problématiques. De nombreux glycérolipidomes de souches de *M. gaditana* (sauvage et mutantes) cultivées dans différentes conditions sont ainsi disponibles et de nouvelles données sont régulièrement générées. Cependant, en raison de la complexité et de la variabilité du métabolisme lipidique, l'analyse et l'interprétation de ces données s'avèrent être des tâches complexes qui demandent une expertise dans le domaine, et requièrent la réalisation de diverses expériences coûteuses en termes de temps et en ressources financières.

Ainsi, le troisième objectif de la thèse consistait à exploiter iMgadiT pour interpréter ces profils lipidiques détaillés de *M. gaditana*. Nous avons choisi d'ajouter ces données au modèle en les intégrant à des fonctions de biomasse. BioModTool, ainsi que le module `lipidomic_data_lpcv.py`, présentés dans le [Chapitre 3](#), ont été développés en réponse à cet objectif. BioModTool offre la possibilité de générer rapidement de nouvelles fonctions de biomasse à partir de données de composition de biomasse que l'utilisateur renseigne dans un fichier Excel. Le module `lipidomic_data_lpcv.py` permet, quant à lui, de compléter automatiquement ce fichier Excel d'entrée de BioModTool en utilisant les données lipidomiques au format standardisé du LPCV.

L'ajout d'une fonction de biomasse représentant la composition de la biomasse d'un mutant du laboratoire : le mutant *acsbg*, ainsi que l'analyse phénotypique *in silico* de ce mutant a permis de démontrer l'utilité de ces outils.

#### 4.1.3.1 Analyse phénotypique du mutant *acsbg*

BioModTool et le module `lipidomic_data` ont été utilisés pour ajouter deux fonctions de biomasse dans iMgadiT23 : l'une pour la souche sauvage et l'autre pour le mutant *acsbg*. La composition de ces fonctions de biomasse repose notamment sur des glycérolipidomes générés par le LPCV et précédemment publiés (Billey et al. 2021). Ces deux fonctions de biomasse ont permis de réaliser une analyse différentielle de flux *in silico* entre les deux souches. Les résultats de cette analyse sont cohérents avec les conclusions de l'analyse expérimentale du mutant *acsbg* effectuée par Billey et al., suggérant une spécificité de l'enzyme mutée MgACSBG pour les acides gras à longue chaîne carbonée.

Les prédictions du modèle confortent notamment le rôle proposé pour MgACSBG dans la production de 16:1-CoA précédant l'incorporation de ces acyl-CoA à la position *sn*-2 des glycérolipides via la voie Kennedy. Ainsi, cette analyse a renforcé la validation expérimentale du modèle et a démontré le potentiel d'application de iMgadit23 et de BioModTool dans la compréhension de phénotypes lipidiques complexes, en particulier pour la caractérisation de souches mutantes.

Cette approche d'analyse différentielle offre la possibilité d'explorer diverses questions biologiques. Par exemple, elle peut être utilisée pour (1) caractériser d'autres souches de *M. gaditana* génétiquement modifiées ("mutant") en comparant le phénotype du mutant à celui d'une souche de référence, ainsi que pour (2) analyser l'impact des conditions de culture en comparant les phénotypes de souches identiques cultivées dans des conditions différentes. De plus, les molécules lipidiques de *M. gaditana* étant similaires à celles des microalgues appartenant aux genres *Microchloropsis* et *Nannochloropsis*, le modèle ainsi que ses prédictions peuvent être extrapolés à ces microalgues (Mühlroth et al. 2017).

#### 4.1.3.2 Perspectives d'application de la méthodologie à d'autres questions biologiques

Étant donné le potentiel d'application de cette méthodologie et la disponibilité des données au LPCV, les scripts développés pour l'analyse du mutant *acsbg* ont été adaptés pour simplifier la génération de résultats dans d'autres contextes d'application. Des scripts "génériques" sous forme de Jupyter Notebook ont ainsi été élaborés et mis à la disposition du laboratoire.

L'exploitation de ces scripts, de iMgadit, de BioModTool, et de données appropriées permet ainsi d'obtenir de façon semi-automatique des résultats de comparaison de distribution de flux entre deux conditions. Ainsi, à partir de données de composition de biomasse (notamment un glycérolipidome) dans deux conditions : une condition test et une condition contrôle (ctr), BioModTool peut être utilisé pour générer deux fonctions de biomasse et les ajouter à iMgadit. Les scripts génériques peuvent ensuite être utilisés afin de (1) prédire différentes distributions de flux : par pFBA pour le contrôle (pFBA\_ctr) et par pFBA et/ou MOMA pour la condition test (pFBA\_test et MOMA\_test), (2) comparer les flux prédits pour chaque réaction entre le contrôle et le test via le calcul du  $\log_2(\text{Fold Change})$  :

$$\log_2(FC) = \log_2\left(\frac{\text{flux condition test}}{\text{flux condition contrôle}}\right) \quad (\text{Eq. 4.1})$$

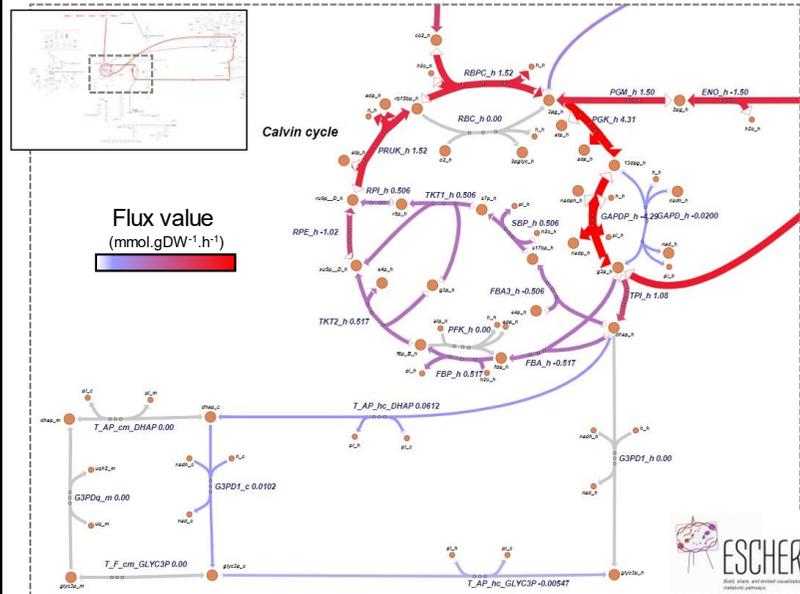
(3) Générer différentes visualisations des résultats (valeurs de flux, et  $\log_2(FC)$  pour chaque réaction) (Figure 4.4). L'ensemble des visualisations des résultats sont générées automatiquement par les scripts, afin de permettre l'analyse et l'interprétation des résultats. Ainsi, chaque distribution de flux (pFBA\_ctr, pFBA\_test et MOMA\_test) est sauvegardée au format JSON afin d'être visualisable *via*

Escher (Figure 4.4A). De même, les valeurs de  $\log_2(\text{FC})$  [ $\log_2(\text{pFBA\_test}/\text{pFBA\_ctr})$  et  $\log_2(\text{MOMA\_test}/\text{pFBA\_ctr})$ ] sont sauvegardées au format JSON afin d'être visualisables sur les réactions de la carte métabolique (Figure 4.4B). Enfin des "Heat Maps" récapitulant les résultats de l'ensemble des réactions du métabolisme lipidique en fonction des couples de chaînes *sn-1* et *sn-2* sont générées automatiquement (C). Pour chaque réaction, ces Heat-Maps sont disponibles pour différents résultats :  $\log_2(\text{pFBA condition test} / \text{pFBA contrôle})$ ,  $\log_2(\text{MOMA condition test} / \text{pFBA contrôle})$ , ainsi que la comparaison entre les deux méthodes. L'ensemble des résultats est également sauvegardé sous forme de tableur CSV. Ces données peuvent être chargées dans TIBCO Spotfire®, une plateforme d'analyse et de visualisation des données (Figure 4.4D). Un fichier d'analyse TIBCO Spotfire® a été créé pour permettre la visualisation et l'exploration dynamique de ces résultats. Une fois les données chargées dans TIBCO Spotfire®, les visualisations sont immédiatement disponibles, explorables dynamiquement et enregistrables sous différents formats.

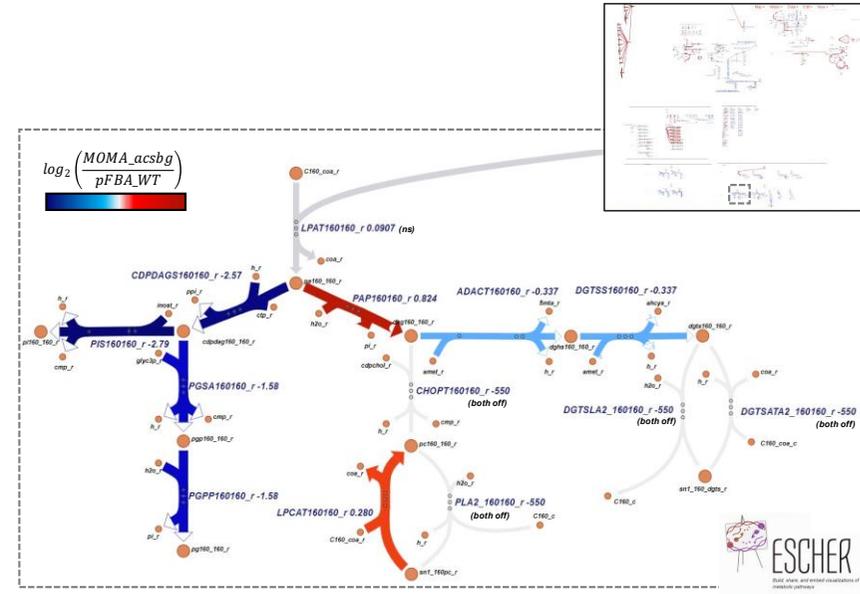
Les scripts ont été développés et organisés de manière à réduire les interactions avec l'utilisateur, simplifiant ainsi leur utilisation pour des personnes non spécialisées en bio-informatique. Les scripts sont disponibles sous la forme de Jupyter Notebooks détaillés et commentés afin de guider l'utilisateur qui n'a que quelques paramètres à renseigner. Ces scripts peuvent reproduire la même analyse que celle effectuée pour le mutant *acsbg*, tout en permettant à l'utilisateur de jouer sur la composition du milieu (qui peut être spécifiée sous forme de tableau CSV) ainsi que sur les deux fonctions de biomasse à contraindre, à savoir celle représentant la condition contrôle et celle de la condition test. De plus, les scripts ont été annotés afin d'explicitier les différentes étapes de génération des résultats et ainsi permettre à des utilisateurs plus expérimentés de faire davantage varier les paramètres de simulation (contraintes, choix des fonctions objectives, des algorithmes de simulation etc.), étendant le potentiel applicatif de ces scripts. Des scripts permettant de réaliser des calculs de la variabilité des flux par FVA sont également disponibles.

La méthodologie développée offre la perspective d'analyse de profils lipidomiques complexes, mais plus largement de données de composition de biomasse. Elle présente ainsi un potentiel intéressant pour aider à mieux comprendre l'impact des variations génétiques et environnementales sur le phénotype de *M. gaditana*. Cette méthodologie pourrait par exemple être appliquée pour analyser l'impact d'une carence nutritionnelle comme la carence en azote.

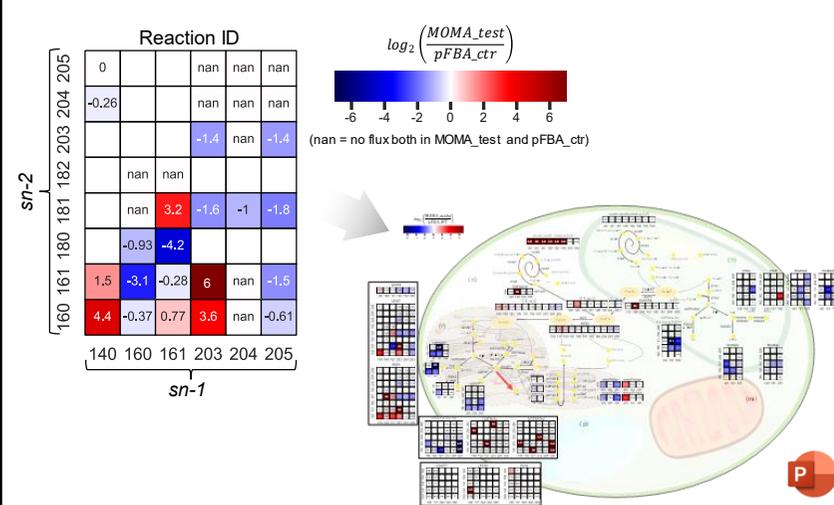
### A. Visualisation des distributions de flux (MOMA\_test)



### B. Visualisation des résultats de log<sub>2</sub>(Fold Change).



### C. Visualisation des résultats du métabolisme lipidique



### D. Exploration des résultats dans TIBCO Spotfire®

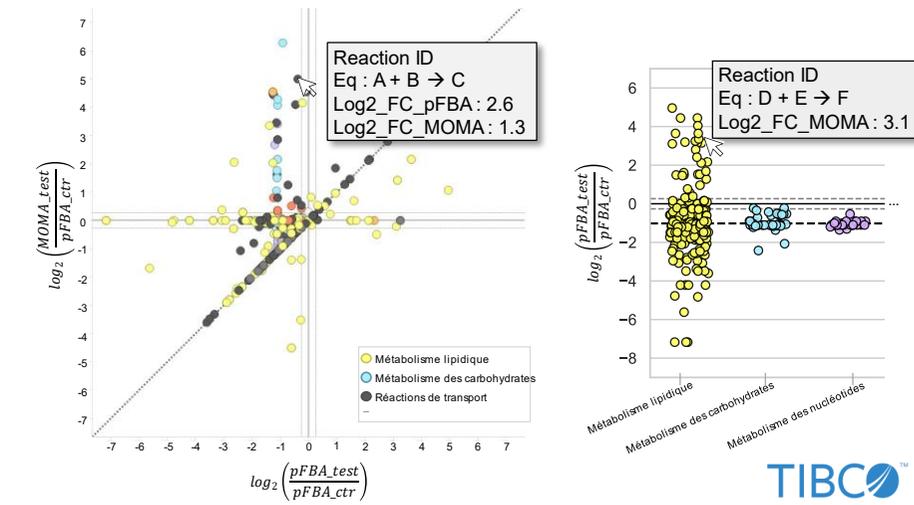


Figure 4.4. Récapitulatif des visualisations des résultats d'analyse différentielle. (Texte de la légende : page suivante)

Les résultats de l'analyse différentielle de flux sont automatiquement mis en forme via les scripts Jupyter Notebooks et sont sauvegardés dans des formats compatibles à leur visualisation via différentes applications. (A) Chaque distribution de flux est visualisable via Escher. (B) Les calculs de  $\log_2(\text{pFBA\_test}/\text{pFBA\_ctr})$  et  $\log_2(\text{MOMA\_test}/\text{pFBA\_ctr})$  pour chaque réaction peuvent être visualisés sur Escher. (C) Tous les résultats numériques sont visualisables dans TIBCO Spotfire®. Cette plateforme propose un système de survol de la souris (mouse-over tool), qui facilite l'exploration dynamique des données en affichant différentes informations contextuelles lorsque l'utilisateur passe la souris sur un point du graphique (identifiant et équation de la réaction,  $\log_2(\text{pFBA\_test}/\text{pFBA\_ctr})$  [Log2\_FC\_pFBA] et  $\log_2(\text{MOMA\_test}/\text{pFBA\_ctr})$  [Log2\_FC\_MOMA]). Les résultats sont colorés par voie métabolique. (D) Des "Heat Maps" présentant les  $\log_2(\text{MOMA\_test}/\text{pFBA\_ctr})$  pour chaque couple sn-1 (abscisse) et sn-2 (ordonnée) de chaque réaction du métabolisme lipidique sont générées. Les différentes Heat Map peuvent être rassemblées sur le schéma bilan du métabolisme lipidique disponible sur PowerPoint®. Des Heat Maps sont également disponibles pour les  $\log_2(\text{pFBA\_test}/\text{pFBA\_ctr})$  et pour la comparaison de  $\log_2(\text{pFBA\_test}/\text{pFBA\_ctr})$  et des  $\log_2(\text{MOMA\_test}/\text{pFBA\_ctr})$  (non représentées).

#### 4.1.3.3 Perspectives d'application : impact d'une carence en azote

Les carences nutritionnelles, notamment en azote et en phosphate, induisent des remodelages lipidiques chez les microalgues. En raison de son impact rapide (environ deux jours) sur la production de lipides, la carence en azote est la carence nutritionnelle la plus largement étudiée. En situation de carence en azote, les microalgues continuent d'assimiler le carbone, mais leur incapacité à synthétiser des métabolites contenant de l'azote, tels que les acides aminés et les acides nucléiques, entraîne l'arrêt de la division cellulaire sans bloquer la photosynthèse (Breuer et al. 2012; Simionato et al. 2013; Zhu et al. 2014; Lupette et Maréchal 2018; Tibocha-Bonilla et al. 2021). Elles convertissent alors le carbone fixé en métabolites de stockage, tels que les lipides pour des souches oléagineuses (*Nannochloropsis*, *Microchloropsis*, *Phaeodactylum*) et/ou l'amidon pour des algues vertes telles que *Chlorella* ou *Chlamydomonas* qui accumulent également des TAG mais dans des quantités moindres (Ho et al. 2013; Condor et al. 2022). Ce phénomène constitue un obstacle majeur à l'application industrielle des microalgues, car il limite leur production continue. Ainsi, l'un des défis majeurs actuels dans l'exploitation des microalgues oléagineuses à des fins de production de biocarburants réside dans l'identification des mécanismes qui permettront de réguler la production de lipides, principalement sous forme de TAG, sans compromettre la division cellulaire. L'exploration des processus métaboliques impliqués dans cette accumulation depuis la photosynthèse, jusqu'au partitionnement des voies métaboliques conduisant aux TAG représente donc un domaine de recherche de grande importance.

Différents modèles de microalgues telles que *Microchloropsis gaditana*, *Microchloropsis salina*, *Thalassiosira pseudonana* (Shah et al. 2017; Loira et al. 2017; van Tol et Armbrust 2021) et de levures oléagineuses *Yarrowia lipolytica*, *Rhodospiridium toruloides*, *Cutaneotrichosporon oleaginosus*,

*Papiliotrema laurentii* (Castañeda et al. 2018; Pham et al. 2021; Ventorim et al. 2022; Kavšček et al. 2015; Kim et al. 2019) ont été utilisés pour analyser l'impact de la carence en azote sur l'accumulation de lipides. Ces analyses reposent sur diverses stratégies de simulation de flux : le FBA et des stratégies dérivées du FBA telles que la dynamic FBA (Mahadevan, Edwards et Doyle 2002), l'environnemental MOMA (ou eMOMA) (Kim et al. 2019). Elles reposent sur l'optimisation d'une fonction objective. Comme présenté dans le chapitre introductif de la thèse, la fonction objective la plus couramment utilisée en FBA (et ses méthodes dérivées) est la maximisation de la production de biomasse, qui implique que l'organisme a évolué de manière à maximiser sa croissance. Bien que cette hypothèse soit probablement correcte dans certaines conditions, elle ne s'applique pas universellement (Schuetz, Kuepfer et Sauer 2007; Schuster, Pfeiffer et Fell 2008; Bernstein et al. 2021). Par exemple, dans des conditions de carence en nutriments, l'objectif d'une cellule n'est probablement pas la production de biomasse, mais plutôt l'optimisation du taux de production de métabolite de stockage pour une utilisation ultérieure. La simulation d'une carence en azote repose ainsi sur l'optimisation de pseudo-réactions consommant des lipides (réactions "lipides" ou "tag").

Des analyses similaires peuvent ainsi être réalisées avec iMgadiT. Des données de compositions de biomasse, ainsi que des profils lipidiques détaillés de *M. gaditana* soumise à une carence en azote sont disponibles au laboratoire. BioModTool-LPCV pourrait ainsi être utilisé pour ajouter des fonctions de biomasse à iMgadiT à partir de ces données. La structure de la fonction de biomasse générée par défaut par BioModTool-LPCV est structurée en trois niveaux. Des pseudo-réactions "LIPIDS" et "TAG" sont donc disponibles pour ces fonctions de biomasse et peuvent être utilisées comme objectif de simulation. Par ailleurs, la flexibilité de BioModTool permet de facilement modifier la structure de la fonction de biomasse générée. Les lipides de la fonction de biomasse pourraient ainsi être séparés en "lipides de stockages" et en "lipides structuraux". La pseudo-réaction correspondant aux lipides de stockages pourrait ainsi être utilisée comme objectif de simulation dans le cas d'une carence en azote. Différentes stratégies de simulation de la carence en azote peuvent ainsi être envisagées avec iMgadiT. Les visualisations présentées dans la section précédente pourront être générées et exploitées pour l'interprétation des résultats obtenus *via* ces simulations. L'ensemble de ces approches offre un potentiel intéressant pour accroître notre compréhension de l'influence de la carence sur la redistribution des flux de la photosynthèse jusqu'au métabolisme lipidique chez *M. gaditana*.

#### 4.1.3.4 Ajouter une biomasse "idéale" au modèle pour identifier des stratégies d'optimisation de son métabolisme

Dans les exemples d'application précédemment présentés, des données expérimentales sont utilisées pour contraindre le modèle. Les prédictions du modèle sont ensuite utilisées pour analyser ces

données et essayer de comprendre le métabolisme de *M. gaditana*. L'approche inverse peut également être imaginée : une composition "idéale" de la biomasse peut être définie : par exemple, une biomasse comportant un grand pourcentage de TAG avec des profils d'acides gras intéressants pour les applications industrielles. BioModTool peut être utilisé pour ajouter la fonction de biomasse basée sur cette composition "idéale" à iMgadit. Cette fonction de biomasse peut être définie comme objectif de simulation afin de réaliser des simulations de flux. Une analyse différentielle de flux peut par exemple être réalisée afin de comparer les flux prédits aux flux d'une condition contrôle. L'analyse des résultats obtenus permettra potentiellement d'aider à cibler des stratégies d'optimisation de souche, ou des conditions de culture à explorer pour obtenir *in vivo* cette composition de biomasse souhaitée.

## 4.2 Autres perspectives d'applications pour iMgadit

Diverses approches sont envisagées pour manipuler le métabolisme des microalgues afin de promouvoir l'accumulation de lipides sans ralentir la croissance : notamment l'optimisation des souches et de leurs conditions de culture (Lupette et Maréchal 2018; Park, Nguyen et Jin 2019; Sundaram et al. 2023). Le modèle métabolique de *M. gaditana*, en plus d'être un outil utile pour aider à l'interprétation des phénotypes, présente des potentiels applicatifs importants dans ces deux contextes.

### 4.2.1 Perspectives pour l'ingénierie métabolique

L'ingénierie métabolique est une stratégie d'optimisation de souches, consistant à manipuler les voies métaboliques d'un organisme en vue d'améliorer ses performances. Elle a été employée pour la modification de plusieurs souches de microalgues dans le but d'augmenter la synthèse et l'accumulation de lipides (Al-Hoqani, Young et Purton 2017; Poliner, Farré et Benning 2018; Ghosh et al. 2016; Park, Nguyen et Jin 2019; Ranjbar et Malcata 2022). Plusieurs systèmes d'édition génomique, permettant d'effectuer des modifications génétiques ciblées, sont disponibles pour *M. gaditana* telles que les systèmes TALEN ou CRISPR-Cas9 (Poliner, Farré et Benning 2018; Kurita et al. 2023) rendant ainsi possible l'ingénierie métabolique de cette souche. Néanmoins, malgré la disponibilité de ces techniques, la génération de souches mutantes demeure une entreprise complexe qui ne garantit pas systématiquement l'obtention des souches envisagées.

Le modèle métabolique développé au cours de la thèse offre des perspectives d'applications intéressantes pour identifier et cibler des manipulations pertinentes afin d'aider à répondre à ces problématiques.

#### 4.2.1.1 Simuler *in silico* l'impact de stratégies d'ingénierie métabolique

Dans un premier temps, iMgadiT peut être utilisé pour modéliser les stratégies d'ingénierie de souche envisagées par les biologistes et prédire leur impact sur le phénotype de *M. gaditana*.

Les modèles métaboliques permettent notamment de prédire les conséquences de trois types de modifications : (1) la suppression de gènes ou de réactions. La suppression d'un gène entraîne l'inactivation des réactions associées (les contraintes supérieures et inférieures des flux de ces réactions sont fixées à zéro). Lorsqu'une réaction est délétée, son flux est fixé à zéro (Heirendt et al. 2019). (2) La surexpression ou la sous-expression de gènes en ajustant les limites de flux des réactions du modèle, et (3) l'incorporation de nouvelles réactions ou de nouvelles voies métaboliques dans le modèle (qui *in vivo* peut être réalisée par l'expression de gène(s) exogène(s) dans la souche d'intérêt). La modélisation de ces stratégies d'ingénierie métabolique permet d'évaluer l'effet de ces manipulations sur le phénotype d'un organisme, notamment de prédire l'essentialité d'un gène ou d'une réaction, d'évaluer l'impact d'une modification génétique sur la production de métabolites d'intérêt, d'identifier les goulots d'étranglement de voies métaboliques etc. L'exploitation de iMgadiT offre ainsi la possibilité d'évaluer la pertinence d'une cible d'ingénierie métabolique avant de tenter de la réaliser expérimentalement.

Cependant, il est essentiel de garder à l'esprit que, bien que le génome de *M. gaditana* ait été largement annoté, il subsiste encore des zones d'ombre, ce qui complique l'ingénierie de souche mais également les analyses *in silico*. Les imprécisions dans les annotations du génome se traduisent par des lacunes dans les associations gène-réaction (GPR) du modèle. Lorsque les gènes codant pour une enzyme n'ont pas encore été identifiés chez *M. gaditana*, aucune correspondance gène-réaction n'est établie pour les réactions catalysées par cette enzyme (les GPR demeurent vides). C'est par exemple le cas des réactions catalysées par une phosphatidic acid phosphatase (PAP). De plus, en raison (1) d'un manque d'informations sur la spécificité des enzymes, en particulier dans le métabolisme lipidique de *M. gaditana*, et (2) de l'absence d'identification de complexes protéiques lors de la ré-annotation fonctionnelle du génome par SilicoLife, une même réaction est souvent associée à plusieurs gènes (séparé par l'opérateur logique "OR"). L'association de plusieurs gènes à une même réaction peut ainsi aussi bien refléter ce manque d'informations, que la plasticité biologique de *M. gaditana*. Le raffinement et la correction des GPR constitue une perspective d'amélioration pour le

modèle. Ces lacunes et approximations ont un impact significatif sur les prédictions d'essentialité génique du modèle. Le modèle peut avoir tendance à sous-estimer l'impact de la délétion d'un gène. Par conséquent, il peut prédire que certains gènes ne sont pas essentiels, alors qu'ils le sont en réalité *in vivo*. Ainsi, la stratégie à privilégier pour modéliser des stratégies d'ingénierie métabolique dans iMgadiT, en particulier pour les voies métaboliques moins bien caractérisées sur le plan génétique, consiste à évaluer l'impact des modifications apportées aux réactions plutôt qu'aux gènes. Si la suppression ou l'augmentation/diminution du flux d'une réaction entraîne la prédiction d'un phénotype intéressant dans le modèle, une étape de caractérisation génomique peut être entreprise. Des alignements de séquences par BLAST (Basic Local Alignment Search Tool) peuvent être réalisés contre le génome de *M. gaditana* pour identifier les gènes codant pour ces enzymes, notamment à partir de séquences protéiques ou génomiques d'enzymes homologues identifiées et caractérisées chez d'autres espèces (si possible phylogénétiquement proches telles que *Microchloropsis salina* ou du genre *Nannochloropsis*, sinon de microalgues plus éloignées mais mieux caractérisées, telles que *Chlorella vulgaris* ou *Chlamydomonas reinhardtii* ou encore de plantes supérieures telle que *Arabidopsis thaliana*). Si le(s) gène(s) codant le(s) enzymes d'intérêt sont identifiés, ils pourront alors être mutés expérimentalement, et être ajouté(s) au modèle. La mise à jour des GPR du modèle pourrait être effectuée progressivement à mesure de l'avancement de la caractérisation génomique de *M. gaditana*, ce qui permettrait une amélioration continue du modèle.

Par ailleurs, iMgadiT offre également le potentiel d'identifier des nouvelles cibles d'ingénierie métabolique, allant au-delà des propositions des biologistes.

#### 4.2.1.2 Identification de nouvelles cibles d'ingénierie métabolique

De nombreuses méthodes de conception de souches *in silico* (computational strain optimization methods ou CSOM) existent. Ces méthodes ont fait l'objet de diverses revues (Lewis, Nagarajan et Palsson 2012; Maia, Rocha et Rocha 2016; Machado et Herrgård 2015; Vieira et al. 2019). Les CSOM peuvent être utilisés pour identifier des stratégies d'édition génomique telles que (1) la délétion de gènes, (2) l'ajout de gènes, (3) up-/down-régulation de l'expression génique et (4) l'addition de voies métaboliques exogènes (Ye et al. 2022). L'application de ces différents algorithmes à iMgadiT ouvre la porte à l'identification de nouvelles cibles d'ingénierie métabolique, potentiellement non intuitives et en dehors du champ de vision des biologistes. Néanmoins, la plupart des algorithmes de conception de souches *in silico* sont conçus pour optimiser des métabolites sécrétés par la cellule ou l'organisme d'intérêt, tels que le succinate ou le lactate, et non des constituants de la biomasse tels que les lipides. Par conséquent, la détermination de la ou des meilleures applications à mettre en œuvre exige une

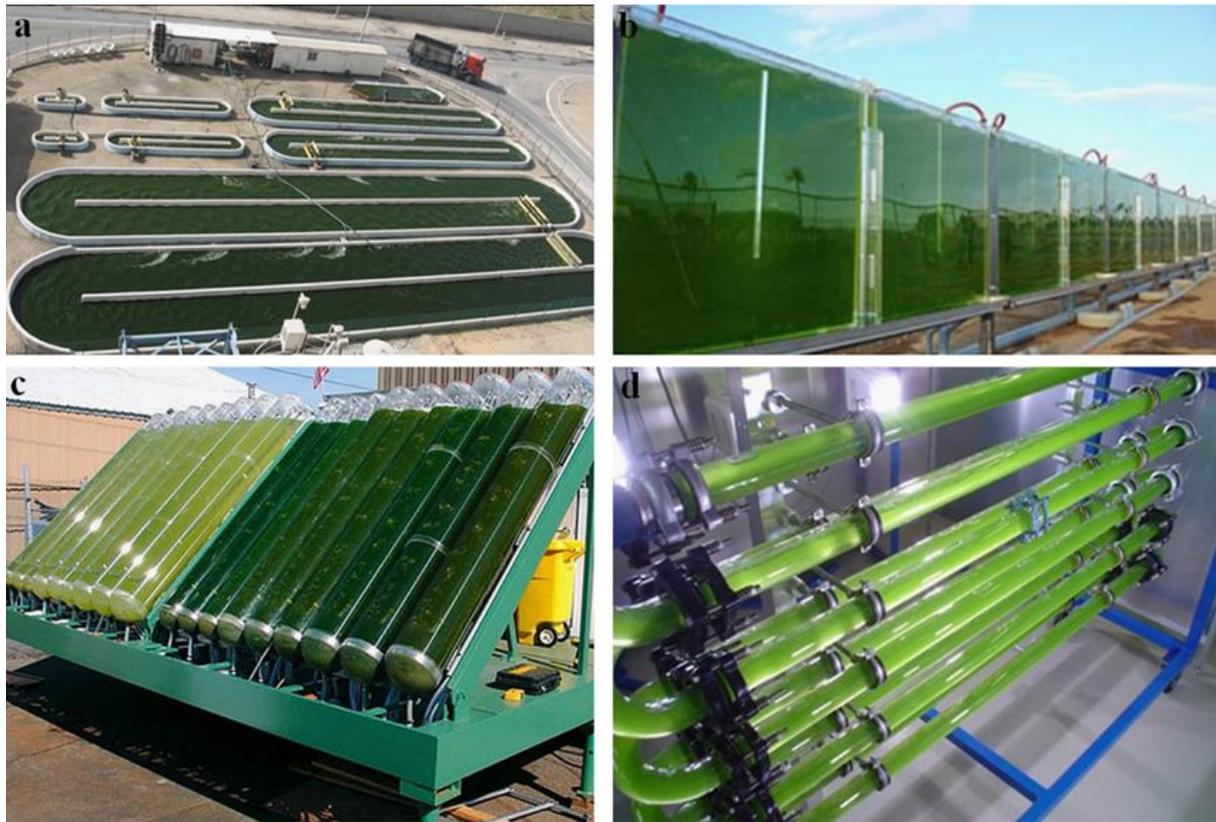
revue bibliographique approfondie des diverses méthodes disponibles, ainsi que la définition de la réaction du modèle à optimiser (telle qu'une pseudo-réaction "TAG") et de sa composition.

Par ailleurs, lorsqu'une souche intéressante a été générée génétiquement et validée expérimentalement au laboratoire, son potentiel applicatif à l'échelle industrielle reste à confirmer par des étapes successives de montées en échelle.

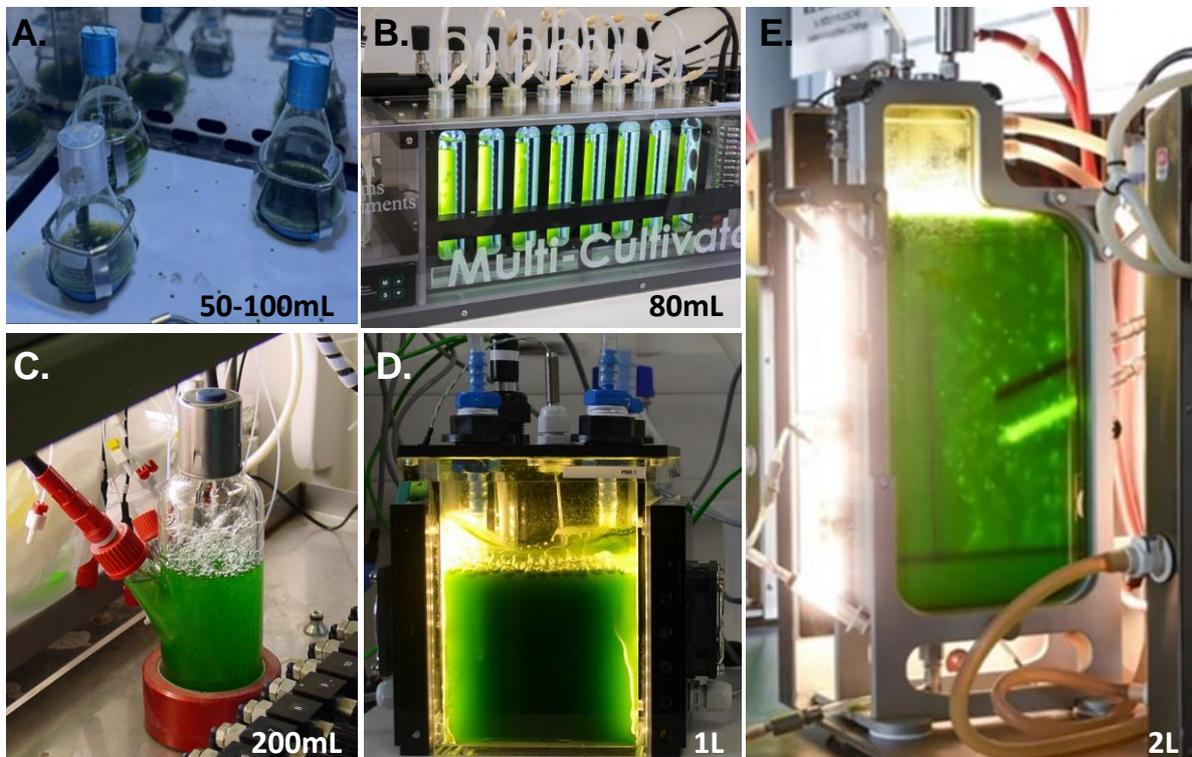
#### 4.2.2 Modélisation multi-échelle : passage à l'échelle du réacteur

La montée en échelle (ou scale-up) des procédés de culture des microalgues représente l'un des défis majeurs pour l'exploitation industrielle de ces organismes. En milieu industriel, les microalgues sont cultivées dans des bassins ou des photobioréacteurs (PBR) de différentes géométries et dimensions (Figure 4.5). Ainsi, elles sont exposées à des variations significatives de lumière, température, pH, et autres paramètres de culture (Bitog et al. 2011; Legrand, Artu et Pruvost 2021). Ces conditions de culture diffèrent considérablement des environnements hautement contrôlés typiques des laboratoires, qui sont généralement réalisés dans des volumes relativement restreints (Figure 4.6).

Par conséquent, la reproduction de résultats obtenus en laboratoire à une échelle industrielle peut s'avérer extrêmement complexe. Pour réussir cette montée en échelle, il est généralement recommandé de prendre en compte les contraintes et objectifs finaux dès les premières étapes de recherche et développement (Straathof et al. 2019; Noorman et Heijnen 2017). Des démarches expérimentales peuvent être réalisées pour optimiser les procédés de culture des microalgues. Cependant une telle optimisation nécessiterait de nombreuses années et serait extrêmement coûteuse. Ainsi l'utilisation de modèles bioinformatiques pour une évaluation rapide et économique des conditions de croissance optimales revêt un intérêt particulier pour l'exploitation industrielle des microalgues.



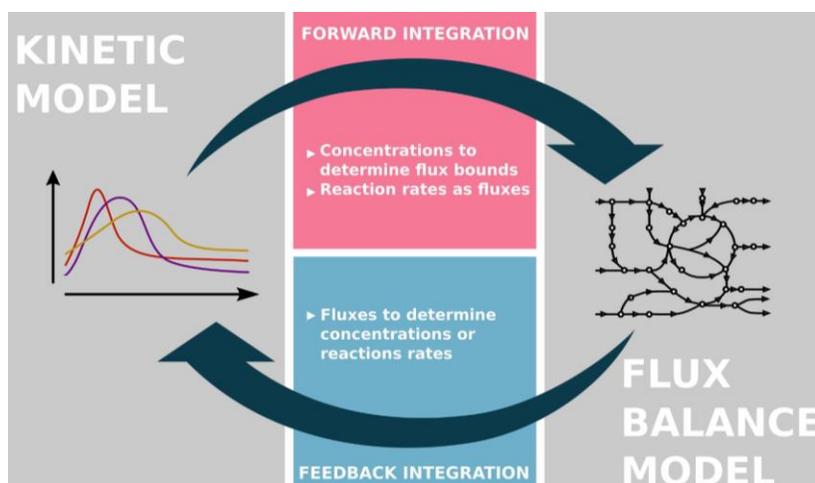
**Figure 4.5. Exemples de photobioreacteurs (PBR) utilisés pour la culture de microalgues.**  
 (a) Bassin Raceway, (b) panneau plat, (c) tubulaires inclinés ou (d) horizontaux. (source : Bitog et al. 2011).



**Figure 4.6. Exemples de PBR de laboratoire.**  
 (A) Fioule Erlenmeyer (Benner et al. 2022), (B) Multicultivateur (Photon Systems Instruments®),  
 (C) colonne à bulles (Benner et al. 2022), (D) MicroPharos (Xanthella®), (E) Labfors (Infors®).

Il existe de nombreux modèles de croissance de microalgues, y compris des genres *Nannochloropsis* et *Microchloropsis* (Lima et al. 2022), permettant la prédiction dynamique des performances d'un PBR. Des revues telles que (Darvehei, Bahri et Moheimani 2018; Lee, Jalalizadeh et Zhang 2015) répertorient différents modèles disponibles dans la littérature. Ces modèles, plus ou moins complexes, intègrent divers paramètres : des conditions de culture telles que la lumière, la température, le pH, la concentration en nutriments, en oxygène, en dioxyde de carbone, etc., aux propriétés physiologiques de la souche modélisée : notamment son absorption de l'énergie lumineuse, différents rendements de production (sucre/photon, biomasse/photon, etc.). Tous ces paramètres sont calibrés à partir de données expérimentales. Ainsi, ces modèles, souvent qualifiés de "boîtes noires", ne prennent pas en compte la variation du métabolisme du microorganisme modélisé. Cependant, le comportement global des cultures de microorganismes est fortement influencé par le fonctionnement métabolique à l'échelle cellulaire. L'établissement d'une connexion directe entre ces deux échelles par la modélisation représente l'un des principaux défis de la biologie des systèmes, un domaine encore largement ouvert et sujet à de nombreuses recherches en cours (Noorman et Heijnen 2017; Straathof et al. 2019).

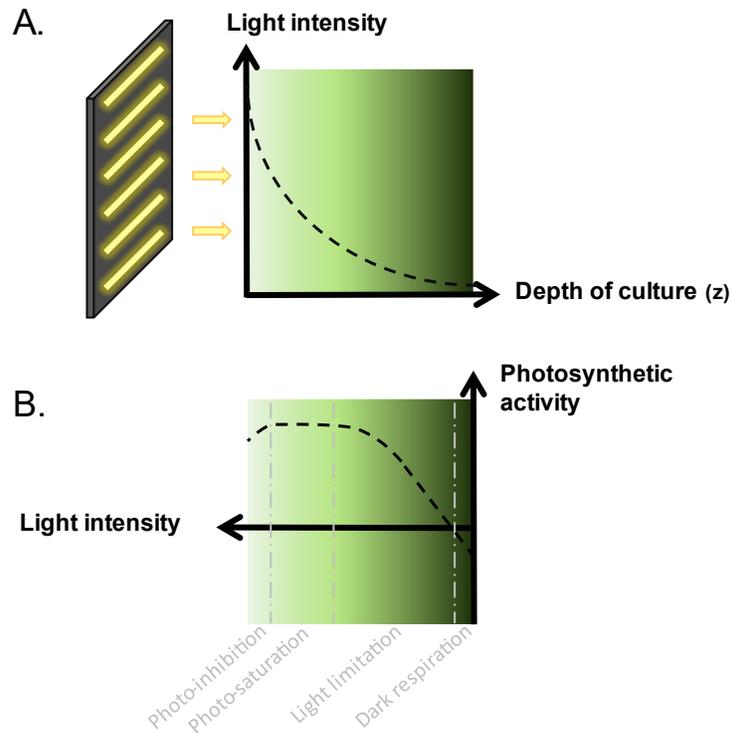
C'est pourquoi l'une des perspectives d'application intéressante des GEM réside dans leur utilisation pour la prédiction de bioprocédés à l'échelle des bioréacteurs. Cependant, afin de modéliser la culture de microalgues, la variation dynamique des conditions de culture doit pouvoir être prise en compte. En effet, les microalgues sont soumises à des conditions de culture fluctuantes qui influencent leur métabolisme. Par exemple, lorsqu'elles sont cultivées en extérieur elles sont soumises au cycle jour/nuit. Elles accumulent ainsi de l'énergie et du carbone durant le jour pour pouvoir continuer leur croissance et satisfaire leur maintenance cellulaire durant la nuit (Bernard 2011). Les approches de modélisation utilisées dans ce contexte doivent donc s'affranchir de l'hypothèse d'état stationnaire du FBA. Plusieurs approches dynamiques ou pseudo-dynamiques de la FBA ont été développées. Du point de vue computationnel, elles sont utilisables même pour de larges réseaux métaboliques (Tibocha-Bonilla et al. 2021; Sahu et al. 2021). Ces approches résultent du couplage entre les formalismes FBA et des modèles cinétiques (systèmes d'équations différentielles ordinaires (EDO)) (Figure 4.7). Dans un premier temps, les variables calculées pour un temps "t" par les EDO sont utilisées pour contraindre le modèle métabolique. Une simulation de distribution de flux est alors réalisée. Les résultats prédits par le GEM sont utilisés pour calculer les valeurs des paramètres des EDO au temps "t+1".



**Figure 4.7. Principe de la modélisation multi-échelle : couplage FBA-EDO (Sahu et al. 2021).**

Par ailleurs, la modélisation à l'échelle d'un PBR nécessite non seulement la prise en compte de la dynamique des conditions de culture, mais aussi la position de la microalgue dans ce dernier. En effet, l'absorption de la lumière est un facteur important à prendre en compte lors de la culture de microalgues. Un phénomène courant observé dans les PBR est l'atténuation de la lumière : les microalgues en périphérie du bioréacteur sont exposées à une forte luminosité quand celles situées au centre ne reçoivent que le flux restant (Figure 4.8A). Par conséquent, l'intensité de la lumière diminue au fur et à mesure qu'elle pénètre dans le système de culture influençant ainsi l'activité photosynthétique des microalgues (Figure 4.8B) (Pruvost et Cornet 2012; Cornet, Dussap et Gros 1995).

L'agitation ou le flux dans un PBR déplace en permanence les microalgues entre la "zone lumineuse" (où la lumière est suffisante) et la "zone sombre" (où la lumière est considérablement atténuée), ce qui peut entraîner une transition du métabolisme autotrophe vers le métabolisme hétérotrophe. Par conséquent, la modélisation d'une culture en PBR dépend à la fois de l'espace et du temps, ce qui rend le développement de tels modèles encore plus complexes que pour des organismes hétérotrophes. Quelques exemples de modélisation multi-échelles de microalgues ont ainsi été développés ces dernières années. Un modèle métabolique de la cyanobactérie *Synechocystis* 6803 couplé à un modèle d'atténuation lumineuse, à un modèle de mouvement cellulaire et à un modèle cinétique de croissance a été publié (He et al. 2015). Un modèle hybride de *Chlorella vulgaris* couplant un GEM (iCZ843) à un modèle d'atténuation lumineuse, à des modèles cinétiques de consommation de photon, de carbone et d'azote, à un modèle de photo-inhibition et à un modèle d'allocation du carbone (accumulation/consommation des carbohydrates et des lipides) a également été proposé (Tibocha-Bonilla et al. 2021).



**Figure 4.8. Atténuation lumineuse et activité photosynthétique dans un photobioréacteur. (Adapté de Pruvost et Cornet 2012).**

(A) L'intensité lumineuse diminue au fur et à mesure de sa pénétration dans le système de culture (z). (B) L'activité photosynthétique est directement liée à l'intensité lumineuse disponible dans le milieu de culture. Lorsque la lumière reçue par les microalgues est inférieure au "point de compensation de la respiration" (dark respiration), l'activité photosynthétique (ou évolution de l'oxygène) est négative. Lorsque l'intensité lumineuse est supérieure à ce point de compensation, l'activité photosynthétique devient linéairement dépendante de la lumière reçue, jusqu'à une intensité de saturation. Pour des intensités lumineuses plus élevées, des processus de photoinhibition peuvent se produire.

Des modèles similaires pourraient être couplés à iMgadiT. Dans cette optique, des modèles d'atténuation lumineuse de *M. gaditana* ont été développés au cours de la thèse. Cependant, faute de temps, ces modèles n'ont pas pu être validés sur plusieurs jeux de données, ni être utilisés pour contraindre iMgadiT lors de simulations.

À l'avenir, ce type de modèle multi-échelles pourrait être utilisé pour l'optimisation de procédés de culture et pour la conception de PBR afin d'augmenter les rendements de production.

### 4.3 Conclusion générale

Les méthodes développées dans cette thèse, ainsi que les résultats obtenus ont un potentiel important pour accroître notre compréhension du métabolisme lipidique de *M. gaditana*. Il s'agit d'un des premiers modèles incorporant l'intégralité de nos connaissances du métabolisme des lipides à ce jour. Il pourra donc être utilisé pour toutes les applications traditionnellement effectuées avec les modèles métaboliques à l'échelle du génome.

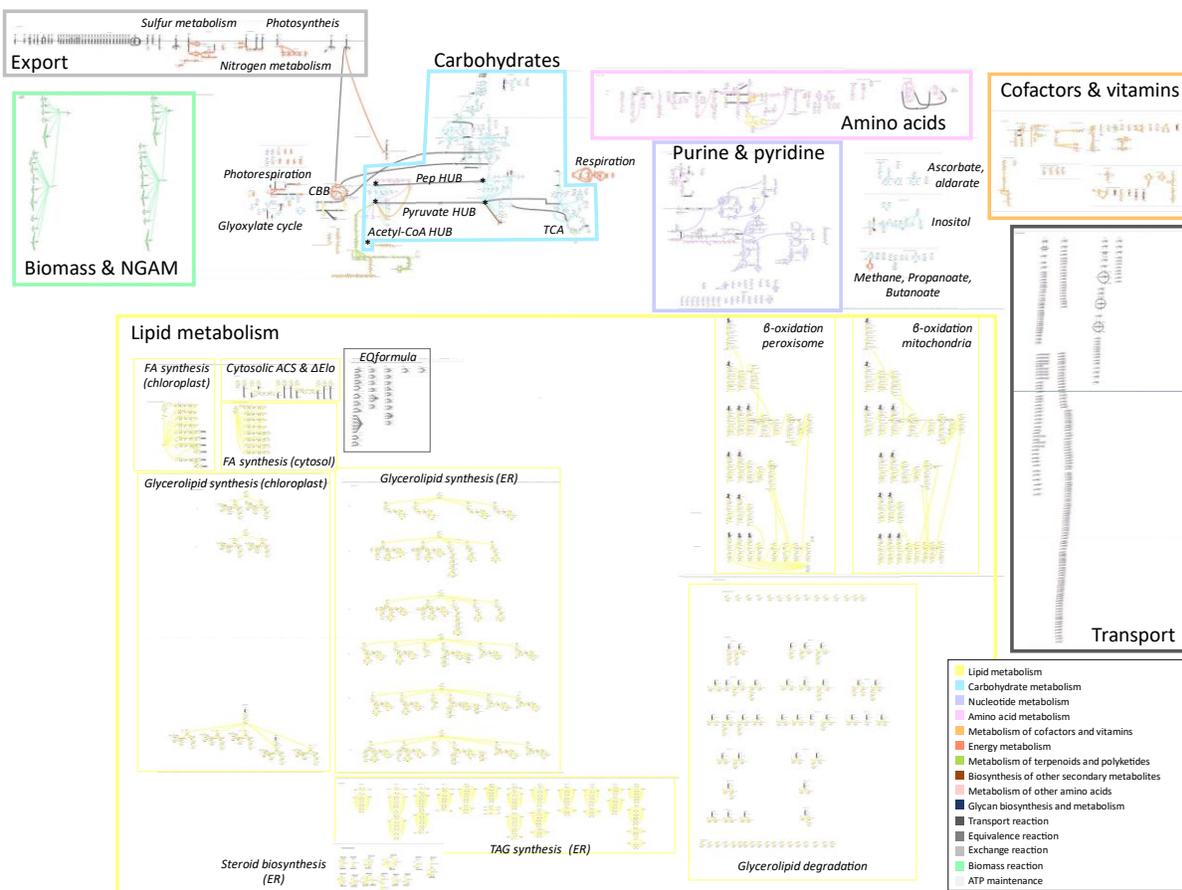
Une meilleure connaissance du métabolisme des lipides et des approches *in silico* sont nécessaires à l'exploitation des microalgues pour des applications industrielles telles que la production de biocarburants. Le modèle métabolique développé au cours de cette thèse va permettre d'identifier les potentiels verrous qui corrélient de façon négative la production de biomasse et de TAG et ainsi de concevoir des souches plus performantes. Il servira aussi de socle pour un couplage à d'autres modèles (cinétique, atténuation lumineuse, allocation du carbone etc.) pour l'optimisation des procédés de culture, qui sont des axes de recherches et de développement importants pour TotalEnergies et le CEA.

# Annexes

## A. Article 1: Supplemental files

### Supplemental Text and Figures

#### 1. iMgadi23 content



**Figure S1. Overview of the 2D-pathway map of iMgadi23.**

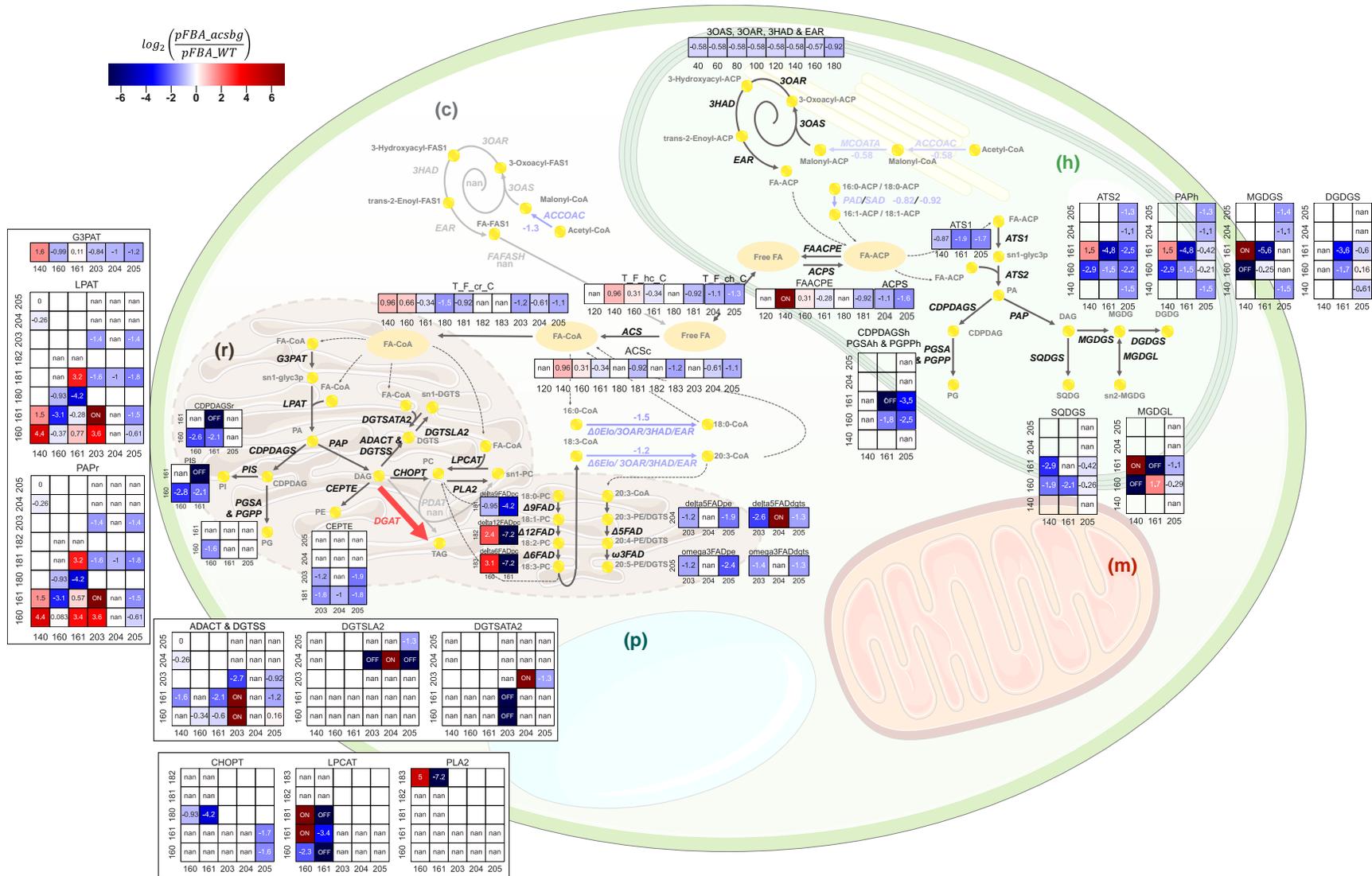
Reactions are colored by subsystem. For lipid metabolism, detailed compartmentalized pathways are indicated.

## 2. Differential flux distribution analysis between WT and *acsbg* strains

### 2.1 pFBA

**Table S1.  $\log_2(\text{pFBA}_{\text{acsbg}}/\text{pFBA}_{\text{WT}})$  results.**

$\log_2(\text{pFBA}_{\text{acsbg}}/\text{pFBA}_{\text{WT}})$	Number of reactions	% of reactions with flux (770)
Both off	1526	
$\log_2(\text{FC}) < -0.2630$	606	78.70
OFF	36	4.68
$\log_2(\text{FC})$ in $[-0.2630;0.2630]$	38	4.94
$\log_2(\text{FC}) > 0.2630$	50	6.49
ON	40	5.19
Change sign: positive to negative	0	0
Change sign: negative to positive	0	0



**Figure S2. Results of pFBA differential flux analysis for lipid metabolism.**

For each reaction, results are shown for all *sn*-1 (*x*-axis) and *sn*-2 (*y*-axis) couples. For each specific reaction,  $\log_2(pFBA\_acsbg/pFBA\_WT)$  value is indicated. Red colors indicate reactions upregulated in *acsbg*. Blue colors indicate reactions downregulated in *acsbg*. "nan" indicates that reaction was predicted to have no flux both in *pFBA\_acsbg* and *pFBA\_WT*. Finally, empty value indicates that reaction does not exist in *iMgadit23*.

## 2.2 MOMA

**Table S2.  $\log_2(\text{MOMA\_acsb}/\text{pFBA\_WT})$  results.**

$\log_2(\text{MOMA\_acsb}/\text{pFBA\_WT})$	Number of reactions	% of reactions with flux (1009)
Both off	1287	
$\log_2(\text{FC}) < -0.2630$	373	36.97
OFF	16	1.59
$\log_2(\text{FC})$ in $[-0.2630;0.2630]$	253	25.07
$\log_2(\text{FC}) > 0.2630$	80	7.93
ON	279	27.65
Change sign: positive to negative	5	0.50
Change sign: negative to positive	3	0.30



# Supplemental Material and Methods

## 1. Addition of lipid degradation and acyl editing in iMgadi23

Acyl editing and glycerolipid degradation pathways addition in iMgadi23 was based on the following arguments.

### 1.1 TAG/DAG/MAG degradation

Since several genes of *M. gaditana* B-31 genome are annotated as TAG or MAG lipases, TAG, DAG and MAG degradation was added to iMgadi23.

#### **MAG lipases:**

Putative monoacylglycerol (MAG) lipases (Naga\_100049g29 and Naga\_100130g4) were identified in *M. gaditana* genome.

#### **TAG lipases:**

Several putative TAG lipases (Naga\_100677g1, Naga\_100055g28, Naga\_102685g1, Naga\_100052g28, Naga\_100718g2, Naga\_100045g8, Naga\_100016g73, Naga\_100039g12, Naga\_100171g1, Naga\_100104g14, Naga\_100529g6, Naga\_100043g24, Naga\_100101g8, Naga\_100133g7, Naga\_100426g5, Naga\_100091g1, Naga\_100271g1, Naga\_100008g63, Naga\_100101g12, Naga\_100011g88, Naga\_100013g62, Naga\_100530g1, Naga\_100144g12, Naga\_100241g4, Naga\_100247g4, Naga\_100046g19, Naga\_100012g35) were identified in *M. gaditana* genome.

**Table S3. TAGL coding genes in NagaB31\_1.0.**

Gene	Identified as homologous to a TAGL	Associated to TAG degradation in Shah GSM model	Annotated as "Lipase, class 3" in NagaB31_1.0	Annotated as "triacylglycerol lipase" in NagaB31_1.0	Annotated as lipase in NagaB31_1.0
Naga_100677g1	homologous to Arabidopsis SDP1 and Phaeodactylum Tgl1	✓			
Naga_100055g28	homologous to Arabidopsis SDP1 and Phaeodactylum Tgl1	✓			
Naga_102685g1	TAGL-Like-A				
Naga_100052g28	Homologous to Human PNPLA3				
Naga_100247g4		✓			PNPLA domain
Naga_100171g1		✓	✓		
Naga_100104g14		✓	✓		
Naga_100529g6		✓	✓		
Naga_100043g24		✓	✓		
Naga_100101g8		✓	✓		
Naga_100133g7		✓	✓		
Naga_100426g5		✓	✓		
Naga_100091g1		✓	✓		
Naga_100271g1		✓	✓		
Naga_100008g63		✓	✓		
Naga_100101g12		✓	✓		
Naga_100144g12		✓	✓		
Naga_100045g8		✓		✓	
Naga_100013g62		✓		✓	
Naga_100011g88		✓			✓
Naga_100530g1		✓			✓
Naga_100241g4		✓			✓
Naga_100046g19		✓			✓
Naga_100012g35		✓			✓
Naga_100718g2		✓			✓
Naga_100016g73		✓			✓
Naga_100039g12		✓			✓

### 1.2 Plastid Galactoglycerolipid Degradation / Acyl editing on MGDG

Complete acyl editing cycle of MGDG was reported for *Chlamydomonas reinhardtii*. Plastid Galactoglycerolipid Degradation 1 (PGD1) catalyzes the hydrolysis of acyl chains at the *sn*-1 position of the glycerol backbone of the plastidial MGDG forming a lyso-MGDG (Li, Benning et Kuo 2012). Recently, CrLAT1 enzyme catalyzing the reacylation of lyso-MGDG was characterized in *C. reinhardtii* (Hoffmann et Shachar-Hill 2023). Lyso-MGDG were also identified in *Emiliana huxleyi* (Malitsky et al. 2016).

### 1.3 Phospholipases

One gene of in *M. gaditana* genome: Naga\_100454g3 is annotated as phospholipase D and contains a PLD domain. This gene was associated with hydrolysis of PC, producing PA and a free choline group.

### 1.4 Acyl editing on PC and DGTS

Evidence of acyl editing on PC or DGTS was reported for microalgae in general (Hoffmann et Shachar-Hill 2023).

## 2. Biomass Objective Functions

### 2.1 Macromolecules stoichiometric coefficients

#### **Cultivation of *M. gaditana* cells**

For DNA and protein analyses, *M. gaditana* CCMP52 was cultivated in artificial seawater (ESAW) using 10 times enriched nitrogen and phosphate sources ( $5.49 \cdot 10^{-3}$  M  $\text{NaNO}_3$  and  $2.24 \cdot 10^{-4}$   $\text{NaH}_2\text{PO}_4$ , (Dolch et al. 2017)). A volume of 80 mL liquid medium was inoculated at  $2.5 \cdot 10^6$  cells.mL<sup>-1</sup> and cultivation was achieved in small photobioreactors (Multi-Cultivator MC 1000, Photon Systems Instruments, Czech Republic) at 24°C, in continuous light (illumination at  $60 \mu\text{mol}_{\text{photon}}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ). Culture mixing was achieved by gas bubbling in which CO<sub>2</sub> levels were maintained constant at 0.5% as in air-lift photobioreactors. Precise and constant level of CO<sub>2</sub> was supplied by a Gas Mixing System GMS 150 (Photon Systems Instruments, Czech Republic) following manufacturer's instructions.

For pigment composition, *M. gaditana* CCMP526 was cultivated in GB medium containing anhydrous salts (solution I : 10 mM  $\text{NaHCO}_3$ , 445 mM  $\text{NaCl}$ , 3.5 mM  $\text{Na}_2\text{SO}_4$ ), hydrated salts (solution II : 3 mM  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ , 2.5 mM  $\text{CaCl}_2\cdot 6\text{H}_2\text{O}$ ), nitrate and phosphate sources (respectively 33.6 mM  $\text{KNO}_3$  and 2.5 mM  $\text{K}_2\text{HPO}_4$ ), iron (28  $\mu\text{M}$   $\text{Fe-EDTA}$ ), trace metals (80  $\mu\text{M}$   $\text{Na}_2\text{EDTA}\cdot 2\text{H}_2\text{O}$ , 4  $\mu\text{M}$   $\text{ZnSO}_4\cdot 7\text{H}_2\text{O}$ , 1.2  $\mu\text{M}$   $\text{CoCl}_2\cdot 6\text{H}_2\text{O}$ , 19  $\mu\text{M}$   $\text{MnCl}_2\cdot 4\text{H}_2\text{O}$ , 0.1  $\mu\text{M}$   $\text{Na}_2\text{MoO}_4\cdot 2\text{H}_2\text{O}$ , 1.3  $\mu\text{M}$   $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ ) and vitamins (0.1  $\mu\text{M}$  vitamin H, 0.1  $\mu\text{M}$  vitamin B12, 3.3  $\mu\text{M}$  vitamin B1), pH was set by addition of 100 mM HEPES buffer pH 7.5. A volume of 1.8 L liquid medium was inoculated at  $2.6 \cdot 10^7$  cells.mL<sup>-1</sup> and cultivation was achieved in flat panel photobioreactors (Labfors 5 Lux, Infors HT, Switzerland) at 26°C under a day:night cycle of 16:8 hours. The incident light intensity increased daily, by keeping the outgoing light intensity  $30\text{-}40 \mu\text{mol}_{\text{photon}}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ , up to  $636 \mu\text{mol}_{\text{photon}}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ . Culture was mixed by filter sterilized air with 2% CO<sub>2</sub> with a flow rate of 1 L.min<sup>-1</sup>.

#### **Pigments extraction and dosage**

Concentration of photosynthetic pigments in *M. gaditana* grown under the aforementioned conditions were estimated spectrometrically in terms of mass of pigments per unit volume of suspension ( $\mu\text{g}/\text{mL}$ ). A volume of 1 mL of culture was first centrifuged at 13,400 rpm for 10 min. The medium was discarded and the cells were resuspended by vortexing in 1.5 mL pure methanol.

Pigments were extracted for 45min at 44°C in the dark and the extract was centrifuged at 13,400 rpm for 10 min. The optical density  $OD_{\lambda}$  of the supernatant, was measured at wavelengths 750, 665, 652, and 480 nm using a spectrophotometer (BioSpectrometer, Eppendorf). Chlorophyll a and PhotoProtective Carotenoids (PPC) were estimated according to the correlations:

$$C_{chl a} = 13.9 (DO_{665} - DO_{750}) \cdot V_2 \cdot \ell^{-1} \cdot V_1^{-1} \text{ (Eq.1)}$$

$$C_{PPC} = (4 (DO_{480} - DO_{750})) \cdot V_2 \cdot \ell^{-1} \cdot V_1^{-1} \text{ (Eq.2)}$$

where  $V_1$  and  $V_2$  are respectively volumes of culture and solvent (methanol) ,  $\ell$  is the thickness of the cuvette used to measure the optical density.

### **Proteins extraction and dosage**

Proteins extraction was performed by adding 100  $\mu$ l of Lysis buffer composed of Tris HCl 50 mM pH 6,8, SDS 4%, protease inhibitor (complete, Roche Diagnostics) to  $2 \cdot 10^8$  *M. gaditana* cells harvested by centrifugation at 3,500 g for 10 minutes. The mix was nitrogen frozen and thawed 3 times before incubation for 30 minutes at room temperature. For protein precipitation 5 volumes (v/v) of 4°C 80% acetone was added to the mix. After 2 hours at -20°C, the proteins were collected by centrifugation at 13,000 g for 15 minutes at 4°C and the pellet rinsed once with cold acetone 80% before to be dried for 30 minutes at RT. Finally, proteins were resuspended in 100  $\mu$ l of Lysis buffer. For protein dosage a BCA protein assay kit (BioVision) was used following supplier instructions, bovine serum albumin was used for establishing the standard curve.

### **DNA extraction and dosage**

Cell lysis was run by vortexing a  $8 \cdot 10^8$  cells pellet collected by centrifugation (10 minutes at 3,500 g) into 500  $\mu$ l of lysis buffer (250 mM Tris pH8.2, 100 mM EDTA, 2% SDS, 100mM NaCl). The solution was then incubated at 60°C for 15 minutes before being vortexed again. DNA extraction was then performed by a classical phenol:chloroform:isoamyl alcohol protocol (25:24:1) used at 1:1 (v/v) and washed twice with chloroform. DNA was precipitated with 30  $\mu$ l Sodium-Acetate 3M pH5 and 1.5 volume of absolute Ethanol (v/v). After centrifugation and drying, the pellet was resuspended in 30  $\mu$ l of MilliQ water for photospectrometric dosage (Biospectrometer, Eppendorf).

### **Lipids, carbohydrates and RNA coefficients**

Total lipid, carbohydrate and RNA coefficients were obtained from literature. Lipid and carbohydrates coefficients were calculated based on *M. gaditana* data respectively reported by (Billey et al. 2021) and (Janssen et al. 2018; Billey et al. 2021). RNA coefficient was calculated from *Ochrophyta* data reviewed by (Finkel et al. 2016).

### **Specific metabolites coefficients**

DNA composition was calculated from %GC of *M. gaditana* genome. RNA composition was calculated from rRNA sequences of *M. gaditana* genome. Detailed carbohydrate composition was based on (Radakovits et al. 2012; Scholz et al. 2014; Shah et al. 2017). Detailed pigment composition was calculated from (Simionato et al. 2013). Amino acid composition of proteins was calculated based on data from (Lamminen et al. 2019; Medina et al. 2015; Volkman et al. 1993; Tibbetts, Bjornsson et McGinn 2015). Finally, lipidomic profiles in terms of glycerolipid classes and acyl-chains was detailed based on experimental characterization “untransformed WT” and “MgACSBG#31” strains from (Billey et al. 2021).

### **ATP maintenance**

ATP requirements for DNA, RNA, proteins and pigments formation were respectively set to 1.372, 0.4, 4.306 and 2.0 mol ATP per mol of macromolecule (Neidhart F. C. 1996; Oliveira, Nielsen et Förster 2005; Kliphuis et al. 2012; Fachet et al. 2020). The Growth Associated Maintenance (GAM) requirement was based on *Chlamydomonas reinhardtii* GEM (29.89 mmol<sub>ATP.gDW<sup>-1</sup>.h<sup>-1</sup></sub>) (Boyle et Morgan 2009). Energetic requirement for non-growth associated maintenance (NGAM), was added to iMgadi23 as an independent reaction (NGAM\_c) with a lower\_bound constrained of 2.2 mmol<sub>ATP.gDW<sup>-1</sup>.h<sup>-1</sup></sub> based on *Nannochloropsis* sp. data (Zhang et al. 2014; Shah et al. 2017).

Detailed biomass composition, original data and sources and final stoichiometric coefficients are detailed in Supplemental Tables S1 and S2.

## **3. Flux simulation**

### *3.1 Default medium composition*

Minimum defined medium, representing photoautotrophic growth with NO<sub>3</sub> as nitrogen source was used. Unlimited uptake rates of CO<sub>2</sub>, O<sub>2</sub>, H<sub>2</sub>O, protons, phosphate, sulfates, magnesium, mannose, and 6-deoxy-galactose were allowed. This entailed setting the lower bounds of the corresponding exchange reactions to - 100000, as we used the usual convention of writing the exchange reactions in such way that production corresponds to positive fluxes and consumption to negative ones. Complete default constraints are listed in Table S4. In default medium conditions photon uptake was limited to 23 mmol<sub>photon.gDW<sup>-1</sup>.h<sup>-1</sup></sub> resulting in a maximum growth predicted with pFBA 0.027 h<sup>-1</sup> to be consistent with average growth rate reported in literature for *M. gaditana* (Gentile et Blanch 2001; Ren et Ogden 2014; Kim et al. 2014; Kareya et al. 2020).

**Table S4. Default constraints within iMgadit23.**

Reaction ID	Lower bound	Upper bound
EX_5mta_	0	100000
EX_ac_	0	100000
EX_ade_	0	100000
EX_arg__L_	0	100000
EX_ca2_	0	100000
EX_chol_	0	100000
EX_co2_	-100000	100000
EX_cobalt2_	0	100000
EX_fe2_	0	100000
EX_for_	0	100000
EX_fru_	0	100000
EX_fuc__L_	-100000	0
EX_fum_	0	100000
EX_gal_	0	100000
EX_glc__D_	0	100000
EX_gln__L_	0	100000
EX_glu__L_	0	100000
EX_glyc_	0	100000
EX_glyc3p_	0	100000
EX_gua_	0	100000
EX_h_	-100000	100000
EX_h2o_	-100000	100000
EX_hco3_	0	100000
EX_his__L_	0	100000
EX_inost_	0	100000
EX_leu__L_	0	100000
EX_lys__L_	0	100000
EX_man_	-100000	0
EX_mg2_	-100000	0
EX_na1_	0	100000
EX_nh4_	0	0
EX_no2_	0	100000
EX_no3_	-100000	100000
EX_o2_	-100000	100000
EX_phe__L_	0	100000
EX_photon_abs_	-23	0
EX_pi_	-100000	100000
EX_ppi_	0	100000
EX_pro__L_	0	100000
EX_sbt__D_	0	100000
EX_so3_	-100000	100000
EX_so4_	-100000	100000
EX_trp__L_	0	100000
EX_tyr__L_	0	100000
EX_urea_	0	100000
EX_val__L_	0	100000
EX_xylt_	0	100000

### 3.2 Obtention of (Rafay et al. 2020) data for quantitative validation

WebPlotDigitizer (<https://apps.automeris.io/wpd/>) was used to extract volatile suspended solids (VSS;  $\text{gVSS.L}^{-1}$ ) and fixed suspended solids (FSS;  $\text{gFSS.L}^{-1}$ ) from (Rafay et al. 2020); Figure 3). Dry weight (DW;  $\text{gDW.L}^{-1}$ ) was calculated as:

$$DW = VSS + FSS$$

Linear correlation was predicted between DW and VSS ( $r^2 = 0.99$ ), with:

$$DW = 0.89 VSS$$

This relation was used to convert maximum specific nitrate uptake rate ( $\text{mg}_{\text{NO}_3}\cdot\text{gVSS}^{-1}\cdot\text{d}^{-1}$ ) into unit required to constrain iMgadit23 uptake rate:  $\text{mmol}_{\text{NO}_3}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$  (Table S5).

**Table S5. Nitrate uptake data used to constrain iMgadit23 for quantitative validation.**

Initial nitrate concentration ( $\text{mg.L}^{-1}$ ) (Rafay et al., 2020)	Nitrate uptake rate ( $\text{mg}_{\text{NO}_3}\cdot\text{gVSS}^{-1}\cdot\text{d}^{-1}$ ) (Rafay et al., 2020)	Nitrate uptake rate ( $\text{mmol}_{\text{NO}_3}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$ )
100	223	1.33E-01
40	199	1.19E-01
56	208	1.24E-01
<b>avg</b>	<b>210</b>	<b>1.25E-01</b>
std	12.12	7.24E-03

## 4. References

- Billey, Elodie; Magneschi, Leonardo; Leterme, Sébastien; Bedhomme, Mariette; Andres-Robin, Amélie; Poulet, Laurent et al. (2021) Characterization of the Bubblegum acyl-CoA synthetase of *Microchloropsis gaditana*. In : *Plant physiology*, vol. 185, n° 3, p. 815–835. DOI: 10.1093/plphys/kiaa110.
- Boyle, Nanette R.; Morgan, John A. (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. In : *BMC systems biology*, vol. 3, p. 4. DOI: 10.1186/1752-0509-3-4.
- Dolch, Lina-Juana; Lupette, Josselin; Tourcier, Guillaume; Bedhomme, Mariette; Collin, Séverine; Magneschi, Leonardo et al. (2017) Nitric Oxide Mediates Nitrite-Sensing and Acclimation and Triggers a Remodeling of Lipids. In : *Plant physiology*, vol. 175, n° 3, p. 1407–1423. DOI: 10.1104/pp.17.01042.
- Fachet, Melanie; Witte, Carina; Flassig, Robert J.; Rihko-Struckmann, Liisa K.; McKie-Krisberg, Zaid; Polle, Jürgen E. W.; Sundmacher, Kai (2020) Reconstruction and analysis of a carbon-core metabolic network for *Dunaliella salina*. In : *BMC bioinformatics*, vol. 21, n° 1, p. 313. DOI: 10.1186/s12859-019-3325-0.
- Finkel, Zoe V.; Follows, Mick J.; Liefer, Justin D.; Brown, Chris M.; Benner, Ina; Irwin, Andrew J. (2016) Phylogenetic Diversity in the Macromolecular Composition of Microalgae. In : *PloS one*, vol. 11, n° 5, e0155977. DOI: 10.1371/journal.pone.0155977.

- Gentile, M. P.; Blanch, H. W. (2001) Physiology and xanthophyll cycle activity of *Nannochloropsis gaditana*. In : *Biotechnology and bioengineering*, vol. 75, n° 1, p. 1–12. DOI: 10.1002/bit.1158.
- Hoffmann, Danielle Yvonne; Shachar-Hill, Yair (2023) Do betaine lipids replace phosphatidylcholine as fatty acid editing hubs in microalgae? In : *Frontiers in plant science*, vol. 14, p. 1077347. DOI: 10.3389/fpls.2023.1077347.
- Janssen, Jorijn H.; Kastenhofer, Jens; Hoop, Jacob A. de; Lamers, Packo P.; Wijffels, René H.; Barbosa, Maria J. (2018) Effect of nitrogen addition on lipid productivity of nitrogen starved *Nannochloropsis gaditana*. In : *Algal Research*, vol. 33, p. 125–132. DOI: 10.1016/j.algal.2018.05.009.
- Kareya, Mukul Suresh; Mariam, Iqra; Shaikh, Kashif Mohd; Nesamma, Asha Arumugam; Jutur, Pannaga Pavan (2020) Photosynthetic Carbon Partitioning and Metabolic Regulation in Response to Very-Low and High CO<sub>2</sub> in *Microchloropsis gaditana* NIES 2587. In : *Frontiers in plant science*, vol. 11, p. 981. DOI: 10.3389/fpls.2020.00981.
- Kim, Chul Woong; Sung, Min-Gyu; Nam, Kibok; Moon, Myoungsoon; Kwon, Jong-Hee; Yang, Ji-Won (2014) Effect of monochromatic illumination on lipid accumulation of *Nannochloropsis gaditana* under continuous cultivation. In : *Bioresource technology*, vol. 159, p. 30–35. DOI: 10.1016/j.biortech.2014.02.024.
- Kliphuis, Anna M. J.; Klok, Anne J.; Martens, Dirk E.; Lamers, Packo P.; Janssen, Marcel; Wijffels, René H. (2012) Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. In : *Journal of applied phycology*, vol. 24, n° 2, p. 253–266. DOI: 10.1007/s10811-011-9674-3.
- Lamminen, M.; Halmemies-Beauchet-Filleau, A.; Kokkonen, T.; Jaakkola, S.; Vanhatalo, A. (2019) Different microalgae species as a substitutive protein feed for soya bean meal in grass silage based dairy cow diets. In : *Animal Feed Science and Technology*, vol. 247, p. 112–126. DOI: 10.1016/j.anifeedsci.2018.11.005.
- Li, Xiaobo; Benning, Christoph; Kuo, Min-Hao (2012) Rapid triacylglycerol turnover in *Chlamydomonas reinhardtii* requires a lipase with broad substrate specificity. In : *Eukaryotic cell*, vol. 11, n° 12, p. 1451–1462. DOI: 10.1128/EC.00268-12.
- Malitsky, Sergey; Ziv, Carmit; Rosenwasser, Shilo; Zheng, Shuning; Schatz, Daniella; Porat, Ziv et al. (2016) Viral infection of the marine alga *Emiliania huxleyi* triggers lipidome remodeling and induces the production of highly saturated triacylglycerol. In : *The New phytologist*, vol. 210, n° 1, p. 88–96. DOI: 10.1111/nph.13852.
- Medina, Camila; Rubilar, Mónica; Shene, Carolina; Torres, Simonet; Verdugo, Marcela (2015) Protein Fractions with Techno-Functional and Antioxidant Properties from *Nannochloropsis gaditana* Microalgal Biomass. In : *Journal of Biobased Materials and Bioenergy*, vol. 9, n° 4, p. 417–425. DOI: 10.1166/jbmb.2015.1534.
- Neidhart F. C. (1996) *Escherichia coli* and *Salmonella typhimurium*. In : *Cellular and Molecular Biology*, vol. 1, p. 1225. En ligne : <https://ci.nii.ac.jp/naid/10005158520/>.
- Oliveira, Ana Paula; Nielsen, Jens; Förster, Jochen (2005) Modeling *Lactococcus lactis* using a genome-scale flux model. In : *BMC microbiology*, vol. 5, p. 39. DOI: 10.1186/1471-2180-5-39.
- Radakovits, Randor; Jinkerson, Robert E.; Fuerstenberg, Susan I.; Tae, Hongseok; Settlage, Robert E.; Boore, Jeffrey L.; Posewitz, Matthew C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. In : *Nature communications*, vol. 3, p. 686. DOI: 10.1038/ncomms1688.
- Rafay, Ramis; Uratani, Joao M.; Hernandez, Hector H.; Rodríguez, Jorge (2020) Growth and Nitrate Uptake in *Nannochloropsis gaditana* and *Tetraselmis chuii* Cultures Grown in Sequential Batch Reactors. In : *Frontiers in Marine Science*, vol. 7. DOI: 10.3389/fmars.2020.00077.

Ren, Ming; Ogden, Kimberly (2014) Cultivation of *Nannochloropsis gaditana* on mixtures of nitrogen sources. In : *Environmental Progress & Sustainable Energy*, vol. 33, n° 2, p. 551–555. DOI: 10.1002/ep.11818.

Scholz, Matthew J.; Weiss, Taylor L.; Jinkerson, Robert E.; Jing, Jia; Roth, Robyn; Goodenough, Ursula et al. (2014) Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall. In : *Eukaryotic cell*, vol. 13, n° 11, p. 1450–1464. DOI: 10.1128/EC.00183-14.

Shah, Ab Rauf; Ahmad, Ahmad; Srivastava, Shireesh; Jaffar Ali, B. M. (2017) Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. In : *Algal Research*, vol. 26, p. 354–364. DOI: 10.1016/j.algal.2017.08.014.

Simionato, Diana; Block, Maryse A.; La Rocca, Nicoletta; Jouhet, Juliette; Maréchal, Eric; Finazzi, Giovanni; Morosinotto, Tomas (2013) The response of *Nannochloropsis gaditana* to nitrogen starvation includes de novo biosynthesis of triacylglycerols, a decrease of chloroplast galactolipids, and reorganization of the photosynthetic apparatus. In : *Eukaryotic cell*, vol. 12, n° 5, p. 665–676. DOI: 10.1128/EC.00363-12.

Tibbetts, Sean M.; Bjornsson, William J.; McGinn, Patrick J. (2015) Biochemical composition and amino acid profiles of *Nannochloropsis granulata* algal biomass before and after supercritical fluid CO<sub>2</sub> extraction at two processing temperatures. In : *Animal Feed Science and Technology*, vol. 204, p. 62–71. DOI: 10.1016/j.anifeedsci.2015.04.006.

Volkman, John; Brown, Malcolm; Dunstan, Greame; Jeffrey S. (1993) The biochemical composition of marine microalgae from the class eustigmatophyceae. In : *Phycologia*, vol. 29, p. 69–78.

Zhang, Dongmei; Yan, Fei; Sun, Zhongliang; Zhang, Qinghua; Xue, Shengzhang; Cong, Wei (2014) On-line modeling intracellular carbon and energy metabolism of *Nannochloropsis* sp. in nitrogen-repletion and nitrogen-limitation cultures. In : *Bioresource technology*, vol. 164, p. 86–92. DOI: 10.1016/j.biortech.2014.04.083.

## B. Liste des contributions et communications

### Journaux internationaux à comité de lecture

- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Séverine Collin, Eric Maréchal, Elodie Billey, Gilles Curien, Maxime Durot, Juliette Jouhet. "Deciphering lipid mutant phenotypes with a new genome-scale metabolic model of *Microchloropsis gaditana* encompassing detailed lipid metabolism." In: *Submitted*
- **Clémence Dupont-Thibert**, Gilles Curien, Maxime Durot. "BioModTool: from biomass composition data to structured biomass objective functions for genome scale metabolic models." In: *Under preparation*
- Dimitri Tolleter, Edward N. Smith, **Clémence Dupont-Thibert**, Clarisse Uwizeye, Denis Vile, Pauline Gloaguen, Denis Falconet, Giovanni Finazzi, Yves Vandenbrouck, Gilles Curien. "An evolutive quantitative cellular/subcellular atlas of Arabidopsis reference leaf 6." In: *Submitted*

### Conférences nationales et internationales

- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Eric Maréchal, Elodie Billey, Séverine Collin, Juliette Jouhet, Gilles Curien, Maxime Durot. "Reconstruction of a genome-scale metabolic model to improve lipid production in *Microchloropsis gaditana*." (2021) Poster, Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), colloque virtuel, France.
- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Eric Maréchal, Elodie Billey, Séverine Collin, Juliette Jouhet, Gilles Curien, Maxime Durot. "Reconstruction of a genome-scale metabolic model of *Microchloropsis gaditana* with detailed lipid metabolism." (2022), Poster, Young Algaeneers Symposium (YAS), Symposium virtuel, France.
- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Eric Maréchal, Elodie Billey, Séverine Collin, Juliette Jouhet, Gilles Curien, Maxime Durot. "Reconstruction of a genome-scale metabolic model of *Microchloropsis gaditana* with detailed lipid metabolism." (2022) Communication orale, 25<sup>th</sup> International Symposia on Plant Lipids (ISPL), Grenoble, France.
- **Clémence Dupont-Thibert**, Gilles Curien, Maxime Durot. "BioModTool: a python package to add biomass objective functions to genome-scale metabolic models from user data.", (2023), Communication orale, 15<sup>èmes</sup> journées scientifiques de Réseau Francophone de Métabolomique et Fluxomique (RFMF), Perpignan, France.
- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Eric Maréchal, Elodie Billey, Séverine Collin, Juliette Jouhet, Gilles Curien, Maxime Durot. "Reconstruction of a genome-scale metabolic model to improve lipid production in *Microchloropsis gaditana*.", (2023) Poster, 31<sup>st</sup> Intelligent Systems For Molecular Biology and the 22<sup>nd</sup> European Conference on Computational Biology (ISMB/ECCB), Lyon, France.

## Communications scientifiques

- **Clémence Dupont-Thibert**, Florian Delrue, Gilles Curien, Maxime Durot. "Development of multi-scale models to improve lipid production in the microalgae *Microchloropsis gaditana*." (2021) Poster, Applied Mathematics, Scientific Computing, Data Science and Artificial Intelligence (Mathias Days - TotalEnergies), Paris, France.
- **Clémence Dupont-Thibert**, Sónia Carneiro, Bruno Pereira, Rafael Carreira, Paulo Vilaça, Giovanni Finazzi, Eric Maréchal, Elodie Billey, Séverine Collin, Juliette Jouhet, Gilles Curien, Maxime Durot. "Reconstruction of a genome-scale metabolic model of *Microchloropsis gaditana* with detailed lipid metabolism." (2022), Communication Orale, Journée des Doctorants EDCSV (JDCSV), Grenoble, France.

## 5 Bibliographie

---

Adadi, Roi; Volkmer, Benjamin; Milo, Ron; Heinemann, Matthias; Shlomi, Tomer (2012) Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. In : PLoS computational biology, vol. 8, n° 7, e1002575. DOI: 10.1371/journal.pcbi.1002575.

Aimo, Lucila; Liechti, Robin; Hyka-Nouspikel, Nevena; Niknejad, Anne; Gleizes, Anne; Götz, Lou et al. (2015) The SwissLipids knowledgebase for lipid biology. In : Bioinformatics (Oxford, England), vol. 31, n° 17, p. 2860–2866. DOI: 10.1093/bioinformatics/btv285.

Aite, Méziane; Chevallier, Marie; Frioux, Clémence; Trottier, Camille; Got, Jeanne; Cortés, María Paz et al. (2018) Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. In : PLoS computational biology, vol. 14, n° 5, e1006146. DOI: 10.1371/journal.pcbi.1006146.

Alboresi, Alessandro; Perin, Giorgio; Vitulo, Nicola; Diretto, Gianfranco; Block, Maryse; Jouhet, Juliette et al. (2016) Light Remodels Lipid Biosynthesis in *Nannochloropsis gaditana* by Modulating Carbon Partitioning between Organelles. In : Plant physiology, vol. 171, n° 4, p. 2468–2482. DOI: 10.1104/pp.16.00599.

Al-Hoqani, Umaima; Young, Rosanna; Purton, Saul (2017) The biotechnological potential of *Nannochloropsis*. In : Perspectives in Phycology, vol. 4, n° 1, p. 1–15. DOI: 10.1127/pip/2016/0065.

Alves, Marina Amaral; Lamichhane, Santosh; Dickens, Alex; McGlinchey, Aidan; Ribeiro, Henrique Caracho; Sen, Partho et al. (2021) Systems biology approaches to study lipidomes in health and disease. In : Biochimica et biophysica acta. Molecular and cell biology of lipids, vol. 1866, n° 2, p. 158857. DOI: 10.1016/j.bbali.2020.158857.

Anaconda (2016) : Anaconda Software Distribution. En ligne : <https://anaconda.com>, consulté le 2021.

Anton, Mihail; Almaas, Eivind; Benfeitas, Rui; Benito-Vaquerizo, Sara; Blank, Lars M.; Dräger, Andreas et al. (2023) standard-GEM: standardization of open-source genome-scale metabolic models. In : Preprint bioRxiv. DOI: 10.1101/2023.03.21.512712.

Antoniewicz, Maciek R. (2015) Methods and advances in metabolic flux analysis: a mini-review. In : Journal of industrial microbiology & biotechnology, vol. 42, n° 3, p. 317–325. DOI: 10.1007/s10295-015-1585-x.

Aransiola, E. F.; Ojumu, T. V.; Oyekola, O. O.; Madzimbamuto, T. F.; Ikhu-Omoregbe, D.I.O. (2014) A review of current technology for biodiesel production: State of the art. In : Biomass and Bioenergy, vol. 61, p. 276–297. DOI: 10.1016/j.biombioe.2013.11.014.

Archibald, John M.; Keeling, Patrick J. (2002) Recycled plastids: a 'green movement' in eukaryotic evolution. In : Trends in Genetics, vol. 18, n° 11, p. 577–584. DOI: 10.1016/S0168-9525(02)02777-4.

Bansal, Parit; Morgat, Anne; Axelsen, Kristian B.; Muthukrishnan, Venkatesh; Coudert, Elisabeth; Aimo, Lucila et al. (2022) Rhea, the reaction knowledgebase in 2022. In : Nucleic acids research, vol. 50, n° D1, D693–D700. DOI: 10.1093/nar/gkab1016.

Baroukh, C.; Steyer, J. P.; Bernard, O.; Chachuat, B. (2016) dynamic Flux Balance Analysis of the Metabolism of Microalgae under a Diurnal Light Cycle. In : IFAC-PapersOnLine, vol. 49, n° 7, p. 791–796. DOI: 10.1016/j.ifacol.2016.07.285.

Beard, Daniel A.; Liang, Shou-dan; Qian, Hong (2002) Energy Balance for Analysis of Complex Metabolic Networks. In : Biophysical journal, vol. 83, n° 1, p. 79–86. DOI: 10.1016/S0006-3495(02)75150-3.

Beck, Ashley; Hunt, Kristopher; Carlson, Ross (2018) Measuring Cellular Biomass Composition for Computational Biology Applications. In : Processes, vol. 6, n° 5, p. 38. DOI: 10.3390/pr6050038.

Beg, Q. K.; Vazquez, A.; Ernst, J.; Menezes, M. A. de; Bar-Joseph, Z.; Barabási, A.-L.; Oltvai, Z. N. (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. In : Proceedings of the National Academy of Sciences, vol. 104, n° 31, p. 12663–12668. DOI: 10.1073/pnas.0609845104.

Belcour, Arnaud; Got, Jeanne; Aite, Méziane; Delage, Ludovic; Collén, Jonas; Frioux, Clémence et al. (2023) Inferring and comparing metabolism across heterogeneous sets of annotated genomes using AuCoMe. In : Genome research. DOI: 10.1101/gr.277056.122.

Benner, Philipp; Meier, Lisa; Pfeffer, Annika; Krüger, Konstantin; Oropeza Vargas, José Enrique; Weuster-Botz, Dirk (2022) Lab-scale photobioreactor systems: principles, applications, and scalability. In : Bioprocess and biosystems engineering, vol. 45, n° 5, p. 791–813. DOI: 10.1007/s00449-022-02711-1.

Bernard, Olivier (2011) Hurdles and challenges for modelling and control of microalgae for CO<sub>2</sub> mitigation and biofuel production. In : Journal of Process Control - J PROCESS CONTROL, vol. 21, p. 1378–1389. DOI: 10.1016/j.jprocont.2011.07.012.

Bernard, Thomas; Bridge, Alan; Morgat, Anne; Moretti, Sébastien; Xenarios, Ioannis; Pagni, Marco (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. In : Briefings in bioinformatics, vol. 15, n° 1, p. 123–135. DOI: 10.1093/bib/bbs058.

Bernstein, David B.; Sulheim, Snorre; Almaas, Eivind; Segrè, Daniel (2021) Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. In : Genome biology, vol. 22, n° 1, p. 64. DOI: 10.1186/s13059-021-02289-z.

Billey, Elodie; Magneschi, Leonardo; Leterme, Sébastien; Bedhomme, Mariette; Andres-Robin, Amélie; Poulet, Laurent et al. (2021) Characterization of the Bubbegum acyl-CoA synthetase of *Microchloropsis gaditana*. In : Plant physiology, vol. 185, n° 3, p. 815–835. DOI: 10.1093/plphys/kiaa110.

Bitog, J. P.; Lee, I.-B.; Lee, C.-G.; Kim, K.-S.; Hwang, H.-S.; Hong, S.-W. et al. (2011) Application of computational fluid dynamics for modeling and designing photobioreactors for microalgae production: A review. In : Computers and Electronics in Agriculture, vol. 76, n° 2, p. 131–147. DOI: 10.1016/j.compag.2011.01.015.

Bo, Davide Dal; Magneschi, Leonardo; Bedhomme, Mariette; Billey, Elodie; Deragon, Etienne; Storti, Mattia et al. (2021) Consequences of Mixotrophy on Cell Energetic Metabolism in *Microchloropsis gaditana* Revealed by Genetic Engineering and Metabolic Approaches. In : Frontiers in plant science, vol. 12, p. 628684. DOI: 10.3389/fpls.2021.628684.

Bodeit, Oliver; Ben Samir, Inès; Karr, Jonathan R.; Goelzer, Anne; Liebermeister, Wolfram (2023) RBAtools: a programming interface for Resource Balance Analysis models. In : *Bioinformatics advances*, vol. 3, n° 1, vbad056. DOI: 10.1093/bioadv/vbad056.

Boele, Joost; Olivier, Brett G.; Teusink, Bas (2012) FAME, the Flux Analysis and Modeling Environment. In : *BMC Systems Biology*, vol. 6, p. 8. DOI: 10.1186/1752-0509-6-8.

Bogen, Christian; Al-Dilaimi, Arwa; Albersmeier, Andreas; Wichmann, Julian; Grundmann, Michael; Rupp, Oliver et al. (2013) Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its genome sequence reveals characteristics suitable for biofuel production. In : *BMC Genomics*, vol. 14, n° 1, p. 926. DOI: 10.1186/1471-2164-14-926.

Bordbar, Aarash; Monk, Jonathan M.; King, Zachary A.; Palsson, Bernhard O. (2014) Constraint-based models predict metabolic and associated cellular functions. In : *Nature reviews. Genetics*, vol. 15, n° 2, p. 107–120. DOI: 10.1038/nrg3643.

Bowler, Chris; Allen, Andrew E.; Badger, Jonathan H.; Grimwood, Jane; Jabbari, Kamel; Kuo, Alan et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. In : *Nature*, vol. 456, n° 7219, p. 239–244. DOI: 10.1038/nature07410.

Boyle, Nanette R.; Morgan, John A. (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. In : *BMC systems biology*, vol. 3, p. 4. DOI: 10.1186/1752-0509-3-4.

Breuer, Guido; Lamers, Packo P.; Martens, Dirk E.; Draaisma, René B.; Wijffels, René H. (2012) The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. In : *Bioresource technology*, vol. 124, p. 217–226. DOI: 10.1016/j.biortech.2012.08.003.

Broddrick, Jared T.; Rubin, Benjamin E.; Welkie, David G.; Du, Niu; Mih, Nathan; Diamond, Spencer et al. (2016) Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, n° 51, E8344-E8353. DOI: 10.1073/pnas.1613446113.

Burgard, A. P.; Vaidyaraman, S.; Maranas, C. D. (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. In : *Biotechnology progress*, vol. 17, n° 5, p. 791–797. DOI: 10.1021/bp0100880.

Burgard, Anthony P.; Maranas, Costas D. (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. In : *Biotechnology and bioengineering*, vol. 82, n° 6, p. 670–677. DOI: 10.1002/bit.10617.

Burgard, Anthony P.; Pharkya, Priti; Maranas, Costas D. (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. In : *Biotechnology and bioengineering*, vol. 84, n° 6, p. 647–657. DOI: 10.1002/bit.10803.

Cardoso, João G. R.; Jensen, Kristian; Lieven, Christian; Lærke Hansen, Anne Sofie; Galkina, Svetlana; Beber, Moritz et al. (2018) Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. In : *ACS synthetic biology*, vol. 7, n° 4, p. 1163–1166. DOI: 10.1021/acssynbio.7b00423.

Carey, Maureen A.; Dräger, Andreas; Beber, Moritz E.; Papin, Jason A.; Yurkovich, James T. (2020) Community standards to facilitate development and address challenges in metabolic modeling. In : *Molecular systems biology*, vol. 16, n° 8, e9235. DOI: 10.15252/msb.20199235.

Caspi, Ron; Billington, Richard; Fulcher, Carol A.; Keseler, Ingrid M.; Kothari, Anamika; Krummenacker, Markus et al. (2018) The MetaCyc database of metabolic pathways and enzymes. In : *Nucleic acids research*, vol. 46, n° D1, D633-D639. DOI: 10.1093/nar/gkx935.

Caspi, Ron; Billington, Richard; Keseler, Ingrid M.; Kothari, Anamika; Krummenacker, Markus; Midford, Peter E. et al. (2020) The MetaCyc database of metabolic pathways and enzymes - a 2019 update. In : *Nucleic acids research*, vol. 48, n° D1, D445-D453. DOI: 10.1093/nar/gkz862.

Castañeda, María Teresita; Nuñez, Sebastián; Garelli, Fabricio; Voget, Claudio; Battista, Hernán de (2018) Comprehensive analysis of a metabolic model for lipid production in *Rhodospiridium toruloides*. In : *Journal of biotechnology*, vol. 280, p. 11–18. DOI: 10.1016/j.jbiotec.2018.05.010.

Castañeda, María Teresita; Nuñez, Sebastián; Voget, Claudio; Battista, Hernán de (2019) In silico optimization of lipid production in *Rhodospiridium toruloides* by gene knockout strategies. In : *IFAC-PapersOnLine*, vol. 52, n° 1, p. 94–99. DOI: 10.1016/j.ifacol.2019.06.043.

Cecchin, Michela; Berteotti, Silvia; Paltrinieri, Stefania; Vigliante, Ivano; Iadarola, Barbara; Giovannone, Barbara et al. (2020) Improved lipid productivity in *Nannochloropsis gaditana* in nitrogen-replete conditions by selection of pale green mutants. In : *Biotechnology for biofuels*, vol. 13, p. 78. DOI: 10.1186/s13068-020-01718-8.

Chaiboonchoe, Amphun; Dohai, Bushra Saeed; Cai, Hong; Nelson, David R.; Jijakli, Kenan; Salehi-Ashtiani, Kourosh (2014) Microalgal Metabolic Network Model Refinement through High-Throughput Functional Metabolic Profiling. In : *Frontiers in bioengineering and biotechnology*, vol. 2, p. 68. DOI: 10.3389/fbioe.2014.00068.

Chang, Antje; Jeske, Lisa; Ulbrich, Sandra; Hofmann, Julia; Koblitz, Julia; Schomburg, Ida et al. (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. In : *Nucleic acids research*, vol. 49, n° D1, D498-D508. DOI: 10.1093/nar/gkaa1025.

Chang, Roger L.; Ghamsari, Lila; Manichaikul, Ani; Hom, Erik F. Y.; Balaji, Santhanam; Fu, Weiqi et al. (2011) Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. In : *Molecular systems biology*, vol. 7, p. 518. DOI: 10.1038/msb.2011.52.

Cheung, C. Y. Maurice; Williams, Thomas C. R.; Poolman, Mark G.; Fell, David A.; Ratcliffe, R. George; Sweetlove, Lee J. (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. In : *The Plant journal : for cell and molecular biology*, vol. 75, n° 6, p. 1050–1061. DOI: 10.1111/tpj.12252.

Chisti, Yusuf (2007) Biodiesel from microalgae. In : *Biotechnology advances*, vol. 25, n° 3, p. 294–306. DOI: 10.1016/j.biotechadv.2007.02.001.

Christian, Nils; May, Patrick; Kempa, Stefan; Handorf, Thomas; Ebenhöf, Oliver (2009) An integrative approach towards completing genome-scale metabolic networks. In : *Mol. BioSyst.*, vol. 5, n° 12, p. 1889–1903. DOI: 10.1039/B915913B.

Chu, Wan-Loy (2012) Biotechnological applications of microalgae. In : *leJSME*, vol. 6, n° 1, S24-S37.

Clark, Teresa J.; Guo, Longyun; Morgan, John; Schwender, Jorg (2020) Modeling Plant Metabolism: From Network Reconstruction to Mechanistic Models. In : Annual review of plant biology, vol. 71, p. 303–326. DOI: 10.1146/annurev-arplant-050718-100221.

Cogne, Guillaume; Rügen, Marco; Bockmayr, Alexander; Titica, Mariana; Dussap, Claude-Gilles; Cornet, Jean-François; Legrand, Jack (2011) A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae *Chlamydomonas reinhardtii*. In : Biotechnology progress, vol. 27, n° 3, p. 631–640. DOI: 10.1002/btpr.596.

Condor, Billriz E.; Luna, Mark Daniel G. de; Chang, Yu-Han; Chen, Jih-Heng; Leong, Yoong Kit; Chen, Po-Ting et al. (2022) Bioethanol production from microalgae biomass at high-solids loadings. In : Bioresource technology, vol. 363, p. 128002. DOI: 10.1016/j.biortech.2022.128002.

Consortium, The UniProt (2022) UniProt: the Universal Protein Knowledgebase in 2023. In : Nucleic acids research, vol. 51, n° D1, D523-D531. DOI: 10.1093/nar/gkac1052.

Córdoba, Sandra Correa; Tong, Hao; Burgos, Asdrúbal; Zhu, Feng; Alseekh, Saleh; Fernie, Alisdair R.; Nikoloski, Zoran (2023) Identification of gene function based on models capturing natural variability of *Arabidopsis thaliana* lipid metabolism. In : Nature communications, vol. 14, n° 1, p. 4897. DOI: 10.1038/s41467-023-40644-9.

Cornet, J. F.; Dussap, C. G.; Gros, J. B (1995) A simplified monodimensional approach for modeling coupling between radiant light transfer and growth kinetics in photobioreactors. In : Chemical Engineering Science.

Correa, Sandra M.; Fernie, Alisdair R.; Nikoloski, Zoran; Brotman, Yariv (2020) Towards model-driven characterization and manipulation of plant lipid metabolism. In : Progress in lipid research, vol. 80, p. 101051. DOI: 10.1016/j.plipres.2020.101051.

Corteggiani Carpinelli, Elisa; Telatin, Andrea; Vitulo, Nicola; Forcato, Claudio; D'Angelo, Michela; Schiavon, Riccardo et al. (2014) Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. In : Molecular plant, vol. 7, n° 2, p. 323–335. DOI: 10.1093/mp/sst120.

Cotner, James B.; Makino, Wataru; Biddanda, Bopaiah A. (2006) Temperature Affects Stoichiometry and Biochemical Composition of *Escherichia coli*. In : Microbial Ecology, vol. 52, n° 1, p. 26–33. DOI: 10.1007/s00248-006-9040-1.

Cottret, Ludovic; Frainay, Clément; Chazalviel, Maxime; Cabanettes, Floréal; Gloaguen, Yoann; Camenen, Etienne et al. (2018) MetExplore: collaborative edition and exploration of metabolic networks. In : Nucleic acids research, vol. 46, n° W1, W495-W502. DOI: 10.1093/nar/gky301.

Cottret, Ludovic; Wildridge, David; Vinson, Florence; Barrett, Michael P.; Charles, Hubert; Sagot, Marie-France; Jourdan, Fabien (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. In : Nucleic acids research, vol. 38, n° Web Server issue, W132-7. DOI: 10.1093/nar/gkq312.

Courtot, Mélanie; Juty, Nick; Knüpfner, Christian; Waltemath, Dagmar; Zhukova, Anna; Dräger, Andreas et al. (2011) Controlled vocabularies and semantics in systems biology. In : Molecular systems biology, vol. 7, p. 543. DOI: 10.1038/msb.2011.77.

Cunha, Emanuel; Sousa, Vítor; Geada, Pedro; Teixeira, José A.; Vicente, António A.; Dias, Oscar (2023) Systems biology's role in leveraging microalgal biomass potential: Current status and future perspectives. In : *Algal Research*, vol. 69, p. 102963. DOI: 10.1016/j.algal.2022.102963.

Curtis, Bruce A.; Tanifuji, Goro; Burki, Fabien; Gruber, Ansgar; Irimia, Manuel; Maruyama, Shinichiro et al. (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. In : *Nature*, vol. 492, n° 7427, p. 59–65. DOI: 10.1038/nature11681.

Dahlqvist, A.; Stahl, U.; Lenman, M.; Banas, A.; Lee, M.; Sandager, L. et al. (2000) Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, n° 12, p. 6487–6492. DOI: 10.1073/pnas.120067297.

Dal'Molin, Cristiana Gomes de Oliveira; Quek, Lake-Ee; Palfreyman, Robin W.; Nielsen, Lars K. (2011) AlgaGEM--a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. In : *BMC Genomics*, 12 Suppl 4, n° Suppl 4, S5. DOI: 10.1186/1471-2164-12-S4-S5.

Darvehei, Pooya; Bahri, Parisa A.; Moheimani, Navid R. (2018) Model development for the growth of microalgae: A review. In : *Renewable and Sustainable Energy Reviews*, vol. 97, p. 233–258. DOI: 10.1016/j.rser.2018.08.027.

DD-DeCaF Caffeine. En ligne : [caffeine.dd-decaf.eu](http://caffeine.dd-decaf.eu), consulté le 2023.

Degtyarenko, Kirill; Matos, Paula de; Ennis, Marcus; Hastings, Janna; Zbinden, Martin; McNaught, Alan et al. (2008) ChEBI: a database and ontology for chemical entities of biological interest. In : *Nucleic acids research*, vol. 36, n° Database issue, D344-50. DOI: 10.1093/nar/gkm791.

Devoid, Scott; Overbeek, Ross; DeJongh, Matthew; Vonstein, Veronika; Best, Aaron A.; Henry, Christopher (2013) Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. In : *Methods in molecular biology (Clifton, N.J.)*, vol. 985, p. 17–45. DOI: 10.1007/978-1-62703-299-5\_2.

Dias, Oscar; Rocha, Miguel; Ferreira, Eugénio C.; Rocha, Isabel (2015) Reconstructing genome-scale metabolic models with merlin. In : *Nucleic acids research*, vol. 43, n° 8, p. 3899–3910. DOI: 10.1093/nar/gkv294.

Dolch, Lina-Juana; Maréchal, Eric (2015) Inventory of fatty acid desaturases in the pennate diatom *Phaeodactylum tricornutum*. In : *Marine Drugs*, vol. 13, n° 3, p. 1317–1339. DOI: 10.3390/md13031317.

Dolch, Lina-Juana; Rak, Camille; Perin, Giorgio; Tourcier, Guillaume; Broughton, Richard; Leterrier, Marina et al. (2017) A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in *Nannochloropsis*. In : *Plant physiology*, vol. 173, n° 1, p. 742–759. DOI: 10.1104/pp.16.01420.

Duarte, Natalie C.; Herrgård, Markus J.; Palsson, Bernhard Ø. (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. In : *Genome research*, vol. 14, n° 7, p. 1298–1309. DOI: 10.1101/gr.2250904.

Durot, Maxime; Bourguignon, Pierre-Yves; Schachter, Vincent (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. In : FEMS microbiology reviews, vol. 33, n° 1, p. 164–190. DOI: 10.1111/j.1574-6976.2008.00146.x.

Ebrahim, Ali; Lerman, Joshua A.; Palsson, Bernhard O.; Hyduke, Daniel R. (2013) COBRApy: COConstraints-Based Reconstruction and Analysis for Python. In : BMC Systems Biology, vol. 7, n° 1, p. 74. DOI: 10.1186/1752-0509-7-74.

Edwards, J. S.; Palsson, B. O. (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. In : The Journal of biological chemistry, vol. 274, n° 25, p. 17410–17416. DOI: 10.1074/jbc.274.25.17410.

Egelhofer, Volker; Schomburg, Ida; Schomburg, Dietmar (2010) Automatic assignment of EC numbers. In : PLOS Computational Biology, vol. 6, n° 1, e1000661. DOI: 10.1371/journal.pcbi.1000661.

Elsemman, Ibrahim E.; Rodriguez Prado, Angelica; Grigaitis, Pranas; Garcia Albornoz, Manuel; Harman, Victoria; Holman, Stephen W. et al. (2022) Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. In : Nature communications, vol. 13, n° 1, p. 801. DOI: 10.1038/s41467-022-28467-6.

Fahy, Eoin; Sud, Manish; Cotter, Dawn; Subramaniam, Shankar (2007) LIPID MAPS online tools for lipid research. In : Nucleic acids research, vol. 35, n° Web Server issue, W606-12. DOI: 10.1093/nar/gkm324.

Fang, Xu; Wei, Cong; Zhao-Ling, Cai; Fan, Ouyang (2004) Effects of organic carbon sources on cell growth and eicosapentaenoic acid content of *Nannochloropsis* sp. In : Journal of Applied Phycology, vol. 16, n° 6, p. 499–503. DOI: 10.1007/s10811-004-5520-1.

Fawley, Marvin W.; Jameson, Ian; Fawley, Karen P. (2015) The phylogeny of the genus *Nannochloropsis* (Monodopsidaceae, Eustigmatophyceae), with descriptions of *N. australis* sp. nov. and *Microchloropsis* gen. nov. In : Phycologia, vol. 54, n° 5, p. 545–552. DOI: 10.2216/15-60.1.

Feist, Adam M.; Palsson, Bernhard O. (2010) The biomass objective function. In : Current opinion in microbiology, vol. 13, n° 3, p. 344–349. DOI: 10.1016/j.mib.2010.03.003.

Fell, D. A.; Small, J. R. (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. In : The Biochemical journal, vol. 238, n° 3, p. 781–786. DOI: 10.1042/bj2380781.

Ferrer-Ledo, Narcís; Stegemüller, Lars; Janssen, Marcel; Wijffels, René H.; Barbosa, Maria J. (2023) Growth and fatty acid distribution over lipid classes in *Nannochloropsis oceanica* acclimated to different temperatures. In : Frontiers in plant science, vol. 14, p. 1078998. DOI: 10.3389/fpls.2023.1078998.

Finkel, Zoe V.; Follows, Mick J.; Liefer, Justin D.; Brown, Chris M.; Benner, Ina; Irwin, Andrew J. (2016) Phylogenetic Diversity in the Macromolecular Composition of Microalgae. In : PLoS one, vol. 11, n° 5, e0155977. DOI: 10.1371/journal.pone.0155977.

Flassig, Robert J.; Facht, Melanie; Höffner, Kai; Barton, Paul I.; Sundmacher, Kai (2016) Dynamic flux balance modeling to increase the production of high-value compounds in green microalgae. In : Biotechnology for biofuels, vol. 9, p. 165. DOI: 10.1186/s13068-016-0556-4.

Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R. et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. In : *Science* (New York, N.Y.), vol. 269, n° 5223, p. 496–512. DOI: 10.1126/science.7542800.

Fritzemeier, Claus Jonathan; Hartleb, Daniel; Szappanos, Balázs; Papp, Balázs; Lercher, Martin J. (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. In : *PLoS computational biology*, vol. 13, n° 4, e1005494. DOI: 10.1371/journal.pcbi.1005494.

Fu, Weiqi; Nelson, David R.; Mystikou, Alexandra; Daakour, Sarah; Salehi-Ashtiani, Kourosh (2019) Advances in microalgal research and engineering development. In : *Current opinion in biotechnology*, vol. 59, p. 157–164. DOI: 10.1016/j.copbio.2019.05.013.

Gentile, M. P.; Blanch, H. W. (2001) Physiology and xanthophyll cycle activity of *Nannochloropsis gaditana*. In : *Biotechnology and bioengineering*, vol. 75, n° 1, p. 1–12. DOI: 10.1002/bit.1158.

Gerhard, Michal (2018) Roche Biochemical Pathways Wall Charts by Gerhard Michal. <http://biochemical-pathways.com/>.

Ghosh, Ashmita; Khanra, Saumyakanti; Mondal, Madhumanti; Halder, Gopinath; Tiwari, O. N.; Saini, Supreet et al. (2016) Progress toward isolation of strains and genetically engineered strains of microalgae for production of biofuel and other value added chemicals: A review. In : *Energy Conversion and Management*, vol. 113, p. 104–118. DOI: 10.1016/j.enconman.2016.01.050.

Gianchandani, Erwin P.; Oberhardt, Matthew A.; Burgard, Anthony P.; Maranas, Costas D.; Papin, Jason A. (2008) Predicting biological system objectives de novo from internal state measurements. In : *BMC bioinformatics*, vol. 9, p. 43. DOI: 10.1186/1471-2105-9-43.

Gim, Geun Ho; Ryu, Jaewon; Kim, Moon Jong; Kim, Pyung Il; Kim, Si Wouk (2016) Effects of carbon source and light intensity on the growth and total lipid production of three microalgae under different culture conditions. In : *Journal of industrial microbiology & biotechnology*, vol. 43, n° 5, p. 605–616. DOI: 10.1007/s10295-016-1741-y.

Gloaguen, Pauline; Bournais, Sylvain; Alban, Claude; Ravanel, Stéphane; Seigneurin-Berny, Daphné; Matringe, Michel et al. (2017) ChloroKB: A Web Application for the Integration of Knowledge Related to Chloroplast Metabolic Network. In : *Plant physiology*, vol. 174, n° 2, p. 922–934. DOI: 10.1104/pp.17.00242.

Gloaguen, Pauline; Vandenbrouck, Yves; Joyard, Jacques; Curien, Gilles (2021) ChloroKB, a cell metabolism reconstruction of the model plant *Arabidopsis thaliana*. In : *Comptes rendus biologies*, vol. 344, n° 2, p. 157–163. DOI: 10.5802/crbiol.49.

Goelzer, Anne; Fromion, Vincent; Scorletti, Gérard (2011) Cell design in bacteria as a convex optimization problem. In : *Automatica*, vol. 47, n° 6, p. 1210–1218.

Goelzer, Anne; Muntel, Jan; Chubukov, Victor; Jules, Matthieu; Prestel, Eric; Nölker, Rolf et al. (2015) Quantitative prediction of genome-wide resource allocation in bacteria. In : *Metabolic engineering*, vol. 32, p. 232–243.

Gong, Yanhai; Kang, Nam K.; Kim, Young U.; Wang, Zengbin; Wei, Li; Xin, Yi et al. (2020) The NanDeSyn database for *Nannochloropsis* systems and synthetic biology. In : *The Plant journal : for cell and molecular biology*, vol. 104, n° 6, p. 1736–1745. DOI: 10.1111/tpj.15025.

Gu, Changdai; Kim, Gi Bae; Kim, Won Jun; Kim, Hyun Uk; Lee, Sang Yup (2019) Current status and applications of genome-scale metabolic models. In : *Genome biology*, vol. 20, n° 1, p. 121. DOI: 10.1186/s13059-019-1730-3.

Gu, Deqing; Zhang, Cheng; Zhou, Shengguo; Wei, Liujing; Hua, Qiang (2016) IdealKnock: A framework for efficiently identifying knockout strategies leading to targeted overproduction. In : *Computational biology and chemistry*, vol. 61, p. 229–237. DOI: 10.1016/j.compbiolchem.2016.02.014.

Gudmundsson, Steinn; Thiele, Ines (2010) Computationally efficient flux variability analysis. In : *BMC bioinformatics*, vol. 11, p. 489. DOI: 10.1186/1471-2105-11-489.

Haas, Robert; Zelezniak, Aleksej; Iacovacci, Jacopo; Kamrad, Stephan; Townsend, StJohn; Ralsler, Markus (2017) Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. In : *Current Opinion in Systems Biology*, vol. 6, p. 37–45. DOI: 10.1016/j.coisb.2017.08.009.

Hamilton, Joshua J.; Reed, Jennifer L. (2012) Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. In : *PloS one*, vol. 7, n° 4, e34670. DOI: 10.1371/journal.pone.0034670.

Hartmann, Andrés; Vila-Santa, Ana; Kallscheuer, Nicolai; Vogt, Michael; Julien-Laferrrière, Alice; Sagot, Marie-France et al. (2017) OptPipe - a pipeline for optimizing metabolic engineering targets. In : *BMC systems biology*, vol. 11, n° 1, p. 143. DOI: 10.1186/s12918-017-0515-0.

Hastings, Janna; Owen, Gareth; Dekker, Adriano; Ennis, Marcus; Kale, Namrata; Muthukrishnan, Venkatesh et al. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. In : *Nucleic acids research*, vol. 44, n° D1, D1214-9. DOI: 10.1093/nar/gkv1031.

Hawkins, Charles; Ginzburg, Daniel; Zhao, Kangmei; Dwyer, William; Xue, Bo; Xu, Angela et al. (2021) Plant Metabolic Network 15: A resource of genome-wide metabolism databases for 126 plants and algae. In : *Journal of integrative plant biology*, vol. 63, n° 11, p. 1888–1905. DOI: 10.1111/jipb.13163.

He, Lian; Wu, Stephen G.; Wan, Ni; Reding, Adrienne C.; Tang, Yinjie J. (2015) Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function. In : *Microbial cell factories*, vol. 14, p. 206. DOI: 10.1186/s12934-015-0396-0.

Heirendt, Laurent; Arreckx, Sylvain; Pfau, Thomas; Mendoza, Sebastián N.; Richelle, Anne; Heinken, Almut et al. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. In : *Nature protocols*, vol. 14, n° 3, p. 639–702. DOI: 10.1038/s41596-018-0098-2.

Hendry, John I.; Dinh, Hoang V.; Foster, Charles; Gopalakrishnan, Saratram; Wang, Lin; Maranas, Costas D. (2020) Metabolic flux analysis reaching genome wide coverage: lessons learned and future perspectives. In : *Current Opinion in Chemical Engineering*, vol. 30, p. 17–25. DOI: 10.1016/j.coche.2020.05.008.

Hendry, John I.; Prasannan, Charulata B.; Joshi, Aditi; Dasgupta, Santanu; Wangikar, Pramod P. (2016) Metabolic model of *Synechococcus* sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. In : *Bioresource technology*, vol. 213, p. 190–197. DOI: 10.1016/j.biortech.2016.02.128.

Henkanatte-Gedera, S. M.; Selvaratnam, T.; Karbakhshravari, M.; Myint, M.; Nirmalakhandan, N.; van Voorhies, W.; Lammers, Peter J. (2017) Removal of dissolved organic carbon and nutrients from urban wastewaters by *Galdieria sulphuraria*: Laboratory to field scale demonstration. In : *Algal Research*, vol. 24, p. 450–456. DOI: 10.1016/j.algal.2016.08.001.

Hibberd, David J. (1981) Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). In : *Botanical Journal of the Linnean Society*, vol. 82, p. 93–119.

Ho, Shih-Hsin; Huang, Shu-Wen; Chen, Chun-Yen; Hasunuma, Tomohisa; Kondo, Akihiko; Chang, Jo-Shu (2013) Bioethanol production using carbohydrate-rich microalgae biomass as feedstock. In : *Bioresource technology*, vol. 135, p. 191–198. DOI: 10.1016/j.biortech.2012.10.015.

Hoffmann, Danielle Yvonne; Shachar-Hill, Yair (2023) Do betaine lipids replace phosphatidylcholine as fatty acid editing hubs in microalgae? In : *Frontiers in plant science*, vol. 14, p. 1077347. DOI: 10.3389/fpls.2023.1077347.

Hosseini, Zhaleh (2020) Optimization methods and objective functions for analysis of metabolic network models. In : PhD Thesis.

Imam, Saheed; Schäuble, Sascha; Valenzuela, Jacob; López García de Lomana, Adrián; Carter, Warren; Price, Nathan D.; Baliga, Nitin S. (2015a) A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. In : *The Plant journal : for cell and molecular biology*, vol. 84, n° 6, p. 1239–1256. DOI: 10.1111/tpj.13059.

Imam, Saheed; Schäuble, Sascha; Valenzuela, Jacob; López García de Lomana, Adrián; Carter, Warren; Price, Nathan D.; Baliga, Nitin S. (2015b) A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. In : *The Plant journal : for cell and molecular biology*, vol. 84, n° 6, p. 1239–1256. DOI: 10.1111/tpj.13059.

InChI Trust Downloads. [<http://www.inchi-trust.org/downloads/>].

IUPAC (1978) The nomenclature of lipids (recommendations 1976). IUPAC-IUB Commission on Biochemical Nomenclature. In : *Journal of lipid research*, vol. 19, n° 1, p. 114–128.

Janssen, Jorijn H.; Driessen, Jasper L.S.P.; Lamers, Packo P.; Wijffels, René H.; Barbosa, Maria J. (2018a) Effect of initial biomass-specific photon supply rate on fatty acid accumulation in nitrogen depleted *Nannochloropsis gaditana* under simulated outdoor light conditions. In : *Algal Research*, vol. 35, p. 595–601. DOI: 10.1016/j.algal.2018.10.002.

Janssen, Jorijn H.; Driessen, Jasper L.S.P.; Lamers, Packo P.; Wijffels, René H.; Barbosa, Maria J. (2018b) Effect of initial biomass-specific photon supply rate on fatty acid accumulation in nitrogen depleted *Nannochloropsis gaditana* under simulated outdoor light conditions. In : *Algal Research*, vol. 35, p. 595–601. DOI: 10.1016/j.algal.2018.10.002.

Janssen, Jorijn H.; Kastenhofer, Jens; Hoop, Jacob A. de; Lamers, Packo P.; Wijffels, René H.; Barbosa, Maria J. (2018) Effect of nitrogen addition on lipid productivity of nitrogen starved *Nannochloropsis gaditana*. In : *Algal Research*, vol. 33, p. 125–132. DOI: 10.1016/j.algal.2018.05.009.

Jensen, Kristian; Broeken, Valentijn; Hansen, Anne Sofie Lærke; Sonnenschein, Nikolaus; Herrgård, Markus J. (2019) OptCouple: Joint simulation of gene knockouts, insertions and medium modifications

for prediction of growth-coupled strain designs. In : *Metabolic Engineering Communications*, vol. 8, e00087. DOI: 10.1016/j.mec.2019.e00087.

John J. Sheehan; Terri G. Dunahay; John R. Benemann; Paul Roessler (1998) Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae; Close-Out Report. En ligne : <https://api.semanticscholar.org/CorpusID:128446188>.

Jouhet, Juliette; Lupette, Josselin; Clerc, Olivier; Magneschi, Leonardo; Bedhomme, Mariette; Collin, Séverine et al. (2017) LC-MS/MS versus TLC plus GC methods: Consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation. In : *PloS one*, vol. 12, n° 8, e0182423. DOI: 10.1371/journal.pone.0182423.

Joyce, Andrew R.; Palsson, Bernhard Ø. (2008) Predicting gene essentiality using genome-scale in silico models. In : *Methods in molecular biology* (Clifton, N.J.), vol. 416, p. 433–457. DOI: 10.1007/978-1-59745-321-9\_30.

Juneja, Ankita; Chaplen, Frank W. R.; Murthy, Ganti S. (2016) Genome scale metabolic reconstruction of *Chlorella variabilis* for exploring its metabolic potential for biofuels. In : *Bioresource technology*, vol. 213, p. 103–110. DOI: 10.1016/j.biortech.2016.02.118.

Kanehisa, Minoru; Goto, Susumu (2000a) KEGG: Kyoto Encyclopedia of Genes and Genomes. In : *Nucleic acids research*, vol. 28, n° 1, p. 27.

Kanehisa, Minoru; Goto, Susumu (2000b) KEGG: Kyoto Encyclopedia of Genes and Genomes, vol. 28.

Kareya, Mukul Suresh; Mariam, Iqra; Shaikh, Kashif Mohd; Nesamma, Asha Arumugam; Jutur, Pannaga Pavan (2020) Photosynthetic Carbon Partitioning and Metabolic Regulation in Response to Very-Low and High CO<sub>2</sub> in *Microchloropsis gaditana* NIES 2587. In : *Frontiers in plant science*, vol. 11, p. 981. DOI: 10.3389/fpls.2020.00981.

Karp, Peter D.; Billington, Richard; Caspi, Ron; Fulcher, Carol A.; Latendresse, Mario; Kothari, Anamika et al. (2019) The BioCyc collection of microbial genomes and metabolic pathways. In : *Briefings in bioinformatics*, vol. 20, n° 4, p. 1085–1093. DOI: 10.1093/bib/bbx085.

Karp, Peter D.; Midford, Peter E.; Billington, Richard; Kothari, Anamika; Kruppenacker, Markus; Latendresse, Mario et al. (2021) Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. In : *Briefings in bioinformatics*, vol. 22, n° 1, p. 109–126. DOI: 10.1093/bib/bbz104.

Kashyap, Mrinal; Chakraborty, Sukanya; Kumari, Anamika; Rai, Anshuman; Varjani, Sunita; Vinayak, Vandana (2023) Strategies and challenges to enhance commercial viability of algal biorefineries for biofuel production. In : *Bioresource technology*, vol. 387, p. 129551. DOI: 10.1016/j.biortech.2023.129551.

Kavšček, Martin; Bhutada, Govindprasad; Madl, Tobias; Natter, Klaus (2015) Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. In : *BMC systems biology*, vol. 9, p. 72. DOI: 10.1186/s12918-015-0217-4.

Keeling, Patrick J. (2004) Diversity and evolutionary history of plastids and their hosts. In : *American journal of botany*, vol. 91, n° 10, p. 1481–1493. DOI: 10.3732/ajb.91.10.1481.

Keeling, Patrick J. (2010) The endosymbiotic origin, diversification and fate of plastids. In : *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, n° 1541, p. 729–748. DOI: 10.1098/rstb.2009.0103.

Keeling, Patrick J. (2013) The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. In : *Annual review of plant biology*, vol. 64, p. 583–607. DOI: 10.1146/annurev-arplant-050312-120144.

Kelder, Thomas; van Iersel, Martijn P.; Hanspers, Kristina; Kutmon, Martina; Conklin, Bruce R.; Evelo, Chris T.; Pico, Alexander R. (2012) WikiPathways: building research communities on biological pathways. In : *Nucleic acids research*, vol. 40, n° D1, D1301-D1307. DOI: 10.1093/nar/gkr1074.

Kilian, Oliver; Benemann, Christina S. E.; Niyogi, Krishna K.; Vick, Bertrand (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, n° 52, p. 21265–21269. DOI: 10.1073/pnas.1105861108.

Kim, Chul Woong; Sung, Min-Gyu; Nam, Kibok; Moon, Myounghoon; Kwon, Jong-Hee; Yang, Ji-Won (2014) Effect of monochromatic illumination on lipid accumulation of *Nannochloropsis gaditana* under continuous cultivation. In : *Bioresource technology*, vol. 159, p. 30–35. DOI: 10.1016/j.biortech.2014.02.024.

Kim, Joonhoon; Reed, Jennifer L.; Maravelias, Christos T. (2011) Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. In : *PLoS one*, vol. 6, n° 9, e24162. DOI: 10.1371/journal.pone.0024162.

Kim, Minsuk; Park, Beom Gi; Kim, Eun-Jung; Kim, Joonwon; Kim, Byung-Gee (2019) In silico identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. In : *Biotechnology for biofuels*, vol. 12, p. 187. DOI: 10.1186/s13068-019-1518-4.

King, Zachary A.; Dräger, Andreas; Ebrahim, Ali; Sonnenschein, Nikolaus; Lewis, Nathan E.; Palsson, Bernhard O. (2015) Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. In : *PLoS computational biology*, vol. 11, n° 8, e1004321. DOI: 10.1371/journal.pcbi.1004321.

King, Zachary A.; Feist, Adam M. (2013) Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. In : *Industrial Biotechnology*, vol. 9, n° 4, p. 236–246. DOI: 10.1089/ind.2013.0005.

King, Zachary A.; Lu, Justin; Dräger, Andreas; Miller, Philip; Federowicz, Stephen; Lerman, Joshua A. et al. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. In : *Nucleic acids research*, vol. 44, n° D1, D515-22. DOI: 10.1093/nar/gkv1049.

Klitgord, Niels; Segrè, Daniel (2010) The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. In : *Genome informatics. International Conference on Genome Informatics*, vol. 22, p. 41–55.

Kluyver, Thomas; Benjamin, Ragan-Kelley; Fernando, Pérez; Brian E., Granger; Matthias, Bussonnier; Jonathan, Frederic et al. (2016) Positioning and power in academic publishing. Players, agents and agendas proceedings of the 20th International Conference on Electronic Publishing.

Knorr, Andrea L.; Jain, Rishi; Srivastava, Ranjan (2007) Bayesian-based selection of metabolic objective functions. In : *Bioinformatics* (Oxford, England), vol. 23, n° 3, p. 351–357. DOI: 10.1093/bioinformatics/btl619.

Kong, Fantao; Romero, Ismael Torres; Warakanont, Jaruswan; Li-Beisson, Yonghua (2018) Lipid catabolism in microalgae. In : *The New phytologist*, vol. 218, n° 4, p. 1340–1348. DOI: 10.1111/nph.15047.

Krumholz, Elias W.; Yang, Hong; Weisenhorn, Pamela; Henry, Christopher S.; Libourel, Igor G. L. (2012) Genome-wide metabolic network reconstruction of the picoalga *Ostreococcus*. In : *Journal of experimental botany*, vol. 63, n° 6, p. 2353–2362. DOI: 10.1093/jxb/err407.

Kurita, Tomokazu; Iwai, Masako; Ohta, Hiroyuki; Sakuma, Tetsushi; Yamamoto, Takashi (2023) Genome editing for biodiesel production in oleaginous microalga, *Nannochloropsis* species. In : *Gene and Genome Editing*, vol. 6, p. 100027. DOI: 10.1016/j.ggedit.2023.100027.

Kutmon, Martina; Riutta, Anders; Nunes, Nuno; Hanspers, Kristina; Willighagen, Egon L.; Bohler, Anwesha et al. (2016) WikiPathways: capturing the full diversity of pathway knowledge. In : *Nucleic acids research*, vol. 44, n° D1, D488–D494. DOI: 10.1093/nar/gkv1024.

Lachance, Jean-Christophe; Lloyd, Colton J.; Monk, Jonathan M.; Yang, Laurence; Sastry, Anand V.; Seif, Yara et al. (2019) BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. In : *PLoS computational biology*, vol. 15, n° 4, e1006971. DOI: 10.1371/journal.pcbi.1006971.

Lamminen, M.; Halmemies-Beauchet-Filleau, A.; Kokkonen, T.; Jaakkola, S.; Vanhatalo, A. (2019) Different microalgae species as a substitutive protein feed for soya bean meal in grass silage based dairy cow diets. In : *Animal Feed Science and Technology*, vol. 247, p. 112–126. DOI: 10.1016/j.anifeedsci.2018.11.005.

Lavoie, Michel; Saint-Béat, Blanche; Strauss, Jan; Guérin, Sébastien; Allard, Antoine; V. Hardy, Simon et al. (2020) Genome-Scale Metabolic Reconstruction and in Silico Perturbation Analysis of the Polar Diatom *Fragilariopsis cylindrus* Predicts High Metabolic Robustness. In : *Biology*, vol. 9, n° 2, p. 30. DOI: 10.3390/biology9020030.

Lee, Eunyong; Jalalizadeh, Mehregan; Zhang, Qiong (2015) Growth kinetic models for microalgae cultivation: A review. In : *Algal Research*, vol. 12, p. 497–512. DOI: 10.1016/j.algal.2015.10.004.

Legrand, Jack; Artu, Arnaud; Pruvost, Jérémy (2021) A review on photobioreactor design and modelling for microalgae production. In : *Reaction Chemistry & Engineering*, vol. 6, n° 7, p. 1134–1151. DOI: 10.1039/D0RE00450B.

Levering, Jennifer; Broddrick, Jared; Dupont, Christopher L.; Peers, Graham; Beerli, Karen; Mayers, Joshua et al. (2016) Genome-Scale Model Reveals Metabolic Basis of Biomass Partitioning in a Model Diatom. In : *PloS one*, vol. 11, n° 5, e0155038. DOI: 10.1371/journal.pone.0155038.

Lewis, Nathan E.; Hixson, Kim K.; Conrad, Tom M.; Lerman, Joshua A.; Charusanti, Pep; Polpitiya, Ashoka D. et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. In : *Molecular systems biology*, vol. 6, p. 390. DOI: 10.1038/msb.2010.47.

Lewis, Nathan E.; Nagarajan, Harish; Palsson, Bernhard O. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. In : Nature reviews. Microbiology, vol. 10, n° 4, p. 291–305. DOI: 10.1038/nrmicro2737.

Li-Beisson, Yonghua; Thelen, Jay J.; Fedosejevs, Eric; Harwood, John L. (2019) The lipid biochemistry of eukaryotic algae. In : Progress in lipid research, vol. 74, p. 31–68. DOI: 10.1016/j.plipres.2019.01.003.

Liebisch, Gerhard; Fahy, Eoin; Aoki, Junken; Dennis, Edward A.; Durand, Thierry; Ejsing, Christer S. et al. (2020) Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. In : Journal of lipid research, vol. 61, n° 12, p. 1539–1555. DOI: 10.1194/jlr.S120001025.

Lieven, Christian; Beber, Moritz E.; Olivier, Brett G.; Bergmann, Frank T.; Ataman, Meric; Babaei, Parizad et al. (2020) MEMOTE for standardized genome-scale metabolic model testing. In : Nature biotechnology, vol. 38, n° 3, p. 272–276. DOI: 10.1038/s41587-020-0446-y.

Lima, Serena; Brucato, Alberto; Caputo, Giuseppe; Schembri, Luca; Scargiali, Francesca (2022) Modelling *Nannochloropsis gaditana* Growth in Reactors with Different Geometries, Determination of Kinetic Parameters and Biochemical Analysis in Response to Light Intensity. In : Applied Sciences, vol. 12, n° 12, p. 5776. DOI: 10.3390/app12125776.

Loira, Nicolás; Mendoza, Sebastian; Paz Cortés, María; Rojas, Natalia; Travisany, Dante; Di Genova, Alex et al. (2017) Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production. In : BMC systems biology, vol. 11, n° 1, p. 66. DOI: 10.1186/s12918-017-0441-1.

Lubián, Luis M.; Montero, Olimpio; Moreno-Garrido, Ignacio; Huertas, I. Emma; Sobrino, Cristina; González-del Valle, Manuel; Parés, Griselda (2000) *Nannochloropsis* (Eustigmatophyceae) as source of commercially valuable pigments. In : Journal of Applied Phycology, vol. 12, n° 3, p. 249–255. DOI: 10.1023/A:1008170915932.

Lupette, Josselin (2018) Development, architecture and dynamics of lipid droplets in the diatom *Phaeodactylum tricornutum*. Développement, architecture et dynamique des gouttelettes lipidiques de la diatomée *Phaeodactylum tricornutum*. Université Grenoble Alpes. En ligne : <https://theses.hal.science/tel-02116000>.

Lupette, Josselin; Maréchal, Éric (2018) [Chapitre 13] Le potentiel des micro-algues pour la chimie verte et les bioénergies: Microbiodiversité. Paris: Éditions Matériologiques, p. 321–337. En ligne : <https://www.cairn-sciences.info/microbiodiversite--9782373611625-page-321.htm>.

Ma, Xiangmeng; Mi, Yuwei; Zhao, Chen; Wei, Qun (2022) A comprehensive review on carbon source effect of microalgae lipid accumulation for biofuel production. In : The Science of the total environment, vol. 806, n° Pt 3, p. 151387. DOI: 10.1016/j.scitotenv.2021.151387.

Ma, Xiao-Nian; Chen, Tian-Peng; Yang, Bo; Liu, Jin; Chen, Feng (2016) Lipid Production from *Nannochloropsis*. In : Marine Drugs, vol. 14, n° 4. DOI: 10.3390/md14040061.

Ma, Yubin; Wang, Zhiyao; Yu, Changjiang; Yin, Yehu; Zhou, Gongke (2014) Evaluation of the potential of 9 *Nannochloropsis* strains for biodiesel production. In : Bioresource technology, vol. 167, p. 503–509. DOI: 10.1016/j.biortech.2014.06.047.

Machado, Daniel; Andrejev, Sergej; Tramontano, Melanie; Patil, Kiran Raosaheb (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. In : *Nucleic acids research*, vol. 46, n° 15, p. 7542–7553. DOI: 10.1093/nar/gky537.

Machado, Daniel; Herrgård, Markus (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. In : *PLoS computational biology*, vol. 10, n° 4, e1003580. DOI: 10.1371/journal.pcbi.1003580.

Machado, Daniel; Herrgård, Markus J. (2015) Co-evolution of strain design methods based on flux balance and elementary mode analysis. In : *Metabolic Engineering Communications*, vol. 2, p. 85–92. DOI: 10.1016/j.meteno.2015.04.001.

Mahadevan, Radhakrishnan; Edwards, Jeremy S.; Doyle, Francis J. [III] (2002) Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. In : *Biophysical journal*, vol. 83, p. 1331–1340. DOI: 10.1016/S0006-3495(02)73903-9.

Maia, Paulo; Rocha, Miguel; Rocha, Isabel (2016) In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories. In : *Microbiology and molecular biology reviews : MMBR*, vol. 80, n° 1, p. 45–67. DOI: 10.1128/MMBR.00014-15.

Mannan, Ahmad A.; Toya, Yoshihiro; Shimizu, Kazuyuki; McFadden, Johnjoe; Kierzek, Andrzej M.; Rocco, Andrea (2015) Integrating Kinetic Model of *E. coli* with Genome Scale Metabolic Fluxes Overcomes Its Open System Problem and Reveals Bistability in Central Metabolism. In : *PLoS one*, vol. 10, n° 10, e0139507. DOI: 10.1371/journal.pone.0139507.

Marechal, Eric (2021) Carburants à base d'algues oléagineuses Principes, filières, verrous. In : *Techniques de l'Ingenieur*, IN186 v2.

Martin, Fergal J.; Amode, M. Ridwan; Aneja, Alisha; Austine-Orimoloye, Olanrewaju; Azov, Andrey G.; Barnes, If et al. (2023) Ensembl 2023. In : *Nucleic acids research*, vol. 51, n° D1, D933-D941. DOI: 10.1093/nar/gkac958.

Marudhupandi, Thangapandi; Sathishkumar, Ramamoorthy; Kumar, Thipramalai Thankappan Ajith (2016) Heterotrophic cultivation of *Nannochloropsis salina* for enhancing biomass and lipid production. In : *Biotechnology reports (Amsterdam, Netherlands)*, vol. 10, p. 8–16. DOI: 10.1016/j.btre.2016.02.001.

Mata, Teresa M.; Martins, António A.; Caetano, Nidia. S. (2010) Microalgae for biodiesel production and other applications: A review. In : *Renewable and Sustainable Energy Reviews*, vol. 14, n° 1, p. 217–232. DOI: 10.1016/j.rser.2009.07.020.

Matsui, Hideaki; Shiozaki, Kazuhiro; Okumura, Yutaka; Ishikawa, Manabu; Waqalevu, Viliame; Hayasaka, Oki et al. (2020) Effects of phosphorous deficiency of a microalga *Nannochloropsis oculata* on its fatty acid profiles and intracellular structure and the effectiveness in rotifer nutrition. In : *Algal Research*, vol. 49, p. 101905. DOI: 10.1016/j.algal.2020.101905.

May, Patrick; Wienkoop, Stefanie; Kempa, Stefan; Usadel, Björn; Christian, Nils; Rupperecht, Jens et al. (2008) Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. In : *Genetics*, vol. 179, n° 1, p. 157–166. DOI: 10.1534/genetics.108.088336.

Medina, Camila; Rubilar, Mónica; Shene, Carolina; Torres, Simonet; Verdugo, Marcela (2015) Protein Fractions with Techno-Functional and Antioxidant Properties from *Nannochloropsis gaditana* Microalgal Biomass. In : Journal of Biobased Materials and Bioenergy, vol. 9, n° 4, p. 417–425. DOI: 10.1166/jbmb.2015.1534.

Mendoza, Sebastián N.; Olivier, Brett G.; Molenaar, Douwe; Teusink, Bas (2019) A systematic assessment of current genome-scale metabolic reconstruction tools. In : Genome biology, vol. 20, n° 1, p. 158. DOI: 10.1186/s13059-019-1769-1.

Metcalf, Alex J.; Nagygyor, Anthony; Boyle, Nanette R. (2020) Rapid Annotation of Photosynthetic Systems (RAPS): automated algorithm to generate genome-scale metabolic networks from algal genomes. In : Algal Research, vol. 50, p. 101967. DOI: 10.1016/j.algal.2020.101967.

Metting, F. B. (1996) Biodiversity and application of microalgae. In : Journal of Industrial Microbiology, vol. 17, n° 5, p. 477–489. DOI: 10.1007/BF01574779.

Min, K. T.; Benzer, S. (1999) Preventing neurodegeneration in the *Drosophila* mutant bubblegum. In : Science (New York, N.Y.), vol. 284, n° 5422, p. 1985–1988. DOI: 10.1126/science.284.5422.1985.

Ministère de la Transition écologique et de la Cohésion des territoires, Ministère de la Transition énergétique (2023) Biocarburants. En ligne : <https://www.ecologie.gouv.fr/biocarburants>.

Monk, Jonathan M.; Lloyd, Colton J.; Brunk, Elizabeth; Mih, Nathan; Sastry, Anand; King, Zachary et al. (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. In : Nature biotechnology, vol. 35, n° 10, p. 904–908. DOI: 10.1038/nbt.3956.

Montagud, Arnau; Navarro, Emilio; Fernández de Córdoba, Pedro; Urchueguía, Javier F.; Patil, Kiran Raosaheb (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. In : BMC Systems Biology, vol. 4, n° 1, p. 156. DOI: 10.1186/1752-0509-4-156.

Mora Salguero, Daniela Alejandra; Fernández-Niño, Miguel; Serrano-Bermúdez, Luis Miguel; Páez Melo, David O.; Winck, Flavia V.; Caldana, Camila; González Barrios, Andrés Fernando (2018) Development of a *Chlamydomonas reinhardtii* metabolic network dynamic model to describe distinct phenotypes occurring at different CO<sub>2</sub> levels. In : PeerJ, vol. 6, e5528. DOI: 10.7717/peerj.5528.

Moretti, Sébastien; van Tran, T.; Mehl, Florence; Ibberson, Mark; Pagni, Marco (2021) MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. In : Nucleic acids research, vol. 49, n° D1, D570-D574. DOI: 10.1093/nar/gkaa992.

Mühlroth, Alice; Winge, Per; El Assimi, Aimen; Jouhet, Juliette; Maréchal, Eric; Hohmann-Marriott, Martin F. et al. (2017) Mechanisms of Phosphorus Acquisition and Lipid Class Remodeling under P Limitation in a Marine Microalga. In : Plant physiology, vol. 175, n° 4, p. 1543–1559. DOI: 10.1104/pp.17.00621.

Muñoz, Camilo F.; Südfeld, Christian; Naduthodi, Mihris I. S.; Weusthuis, Ruud A.; Barbosa, Maria J.; Wijffels, René H.; D'Adamo, Sarah (2021) Genetic engineering of microalgae for enhanced lipid production. In : Biotechnology advances, vol. 52, p. 107836. DOI: 10.1016/j.biotechadv.2021.107836.

Murakami, Hiroki; Nobusawa, Takashi; Hori, Koichi; Shimojima, Mie; Ohta, Hiroyuki (2018) Betaine Lipid Is Crucial for Adapting to Low Temperature and Phosphate Deficiency in *Nannochloropsis*. In : Plant physiology, vol. 177, n° 1, p. 181–193. DOI: 10.1104/pp.17.01573.

Ng, Rachel H.; Lee, Jihoon W.; Baloni, Priyanka; Diener, Christian; Heath, James R.; Su, Yapeng (2022) Constraint-Based Reconstruction and Analyses of Metabolic Models: Open-Source Python Tools and Applications to Cancer. In : *Frontiers in oncology*, vol. 12, p. 914594. DOI: 10.3389/fonc.2022.914594.

Niebel, Bastian; Leupold, Simeon; Heinemann, Matthias (2019) An upper limit on Gibbs energy dissipation governs cellular metabolism. In : *Nature metabolism*, vol. 1, n° 1, p. 125–132. DOI: 10.1038/s42255-018-0006-7.

Nielsen, Jens (2009) Systems biology of lipid metabolism: from yeast to human. In : *FEBS letters*, vol. 583, n° 24, p. 3905–3913. DOI: 10.1016/j.febslet.2009.10.054.

Nogales, Juan; Gudmundsson, Steinn; Knight, Eric M.; Palsson, Bernhard O.; Thiele, Ines (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, n° 7, p. 2678–2683. DOI: 10.1073/pnas.1117907109.

Nookaew, Intawat; Jewett, Michael C.; Meechai, Asawin; Thammarongtham, Chinae; Laoteng, Kobkul; Cheevadhanarak, Supapon et al. (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. In : *BMC systems biology*, vol. 2, p. 71. DOI: 10.1186/1752-0509-2-71.

Noor, Elad; Haraldsdóttir, Hulda S.; Milo, Ron; Fleming, Ronan M. T. (2013) Consistent estimation of Gibbs energy using component contributions. In : *PLoS computational biology*, vol. 9, n° 7, e1003098. DOI: 10.1371/journal.pcbi.1003098.

Noorman, Henk J.; Heijnen, Joseph J. (2017) Biochemical engineering's grand adventure. In : *Chemical Engineering Science*, vol. 170, p. 677–693. DOI: 10.1016/j.ces.2016.12.065.

Norsigian, Charles J.; Pusarla, Neha; McConn, John Luke; Yurkovich, James T.; Dräger, Andreas; Palsson, Bernhard O.; King, Zachary (2020) BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. In : *Nucleic acids research*, vol. 48, n° D1, D402-D406. DOI: 10.1093/nar/gkz1054.

O'Brien, Edward J.; Monk, Jonathan M.; Palsson, Bernhard O. (2015) Using Genome-scale Models to Predict Biological Capabilities. In : *Cell*, vol. 161, n° 5, p. 971–987. DOI: 10.1016/j.cell.2015.05.019.

Orth, Jeffrey D.; Thiele, Ines; Palsson, Bernhard Ø. (2010) What is flux balance analysis? In : *Nature biotechnology*, vol. 28, n° 3, p. 245–248. DOI: 10.1038/nbt.1614.

Pan, Shu; Reed, Jennifer L. (2018) Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. In : *Current opinion in biotechnology*, vol. 51, p. 103–108. DOI: 10.1016/j.copbio.2017.12.012.

Park, Seunghye; Nguyen, Thu Ha Thi; Jin, EonSeon (2019) Improving lipid production by strain development in microalgae: Strategies, challenges and perspectives. In : *Bioresource technology*, vol. 292, p. 121953. DOI: 10.1016/j.biortech.2019.121953.

Passi, Anurag; Tibocho-Bonilla, Juan D.; Kumar, Manish; Tec-Campos, Diego; Zengler, Karsten; Zuniga, Cristal (2021) Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. In : *Metabolites*, vol. 12, n° 1. DOI: 10.3390/metabo12010014.

Patané, Andrea; Jansen, Giorgio; Conca, Piero; Carapezza, Giovanni; Costanza, Jole; Nicosia, Giuseppe (2019) Multi-objective optimization of genome-scale metabolic models: the case of ethanol production. In : *Annals of Operations Research*, vol. 276, n° 1-2, p. 211–227. DOI: 10.1007/s10479-018-2865-4.

Patil, Kiran Raosaheb; Rocha, Isabel; Förster, Jochen; Nielsen, Jens (2005) Evolutionary programming as a platform for in silico metabolic engineering. In : *BMC bioinformatics*, vol. 6, p. 308. DOI: 10.1186/1471-2105-6-308.

Pereira, Bruno; Miguel, Joana; Vilaça, Paulo; Soares, Simão; Rocha, Isabel; Carneiro, Sónia (2018) Reconstruction of a genome-scale metabolic model for *Actinobacillus succinogenes* 130Z. In : *BMC systems biology*, vol. 12, n° 1, p. 61. DOI: 10.1186/s12918-018-0585-7.

Petroutsos, Dimitris; Amiar, Souad; Abida, Heni; Dolch, Lina-Juana; Bastien, Olivier; Rébeillé, Fabrice et al. (2014) Evolution of galactoglycerolipid biosynthetic pathways - From cyanobacteria to primary plastids and from primary to secondary plastids. In : *Progress in lipid research*, vol. 54, p. 68–85. DOI: 10.1016/j.plipres.2014.02.001.

Petrovs, Rudolfs; Stalidzans, Egils; Pentjuss, Agris (2021) IMFLer: A Web Application for Interactive Metabolic Flux Analysis and Visualization. In : *Journal of Computational Biology*, vol. 28, n° 10, p. 1021–1032. DOI: 10.1089/cmb.2021.0056.

Pham, Nhung (2016) Genome-scale constraint-based metabolic modeling and analysis of *Nannochloropsis* Sp.

Pham, Nhung; Reijnders, Maarten; Suarez-Diez, Maria; Nijssse, Bart; Springer, Jan; Eggink, Gerrit; Schaap, Peter J. (2021) Genome-scale metabolic modeling underscores the potential of *Cutaneotrichosporon oleaginosus* ATCC 20509 as a cell factory for biofuel production. In : *Biotechnology for biofuels*, vol. 14, n° 1, p. 2. DOI: 10.1186/s13068-020-01838-1.

Pharkya, Priti; Burgard, Anthony P.; Maranas, Costas D. (2004) OptStrain: a computational framework for redesign of microbial production systems. In : *Genome research*, vol. 14, n° 11, p. 2367–2376. DOI: 10.1101/gr.2872004.

Poliner, Eric; Farré, Eva M.; Benning, Christoph (2018) Advanced genetic tools enable synthetic biology in the oleaginous microalgae *Nannochloropsis* sp. In : *Plant cell reports*, vol. 37, n° 10, p. 1383–1399. DOI: 10.1007/s00299-018-2270-0.

Poliner, Eric; Panchy, Nicholas; Newton, Linsey; Wu, Guangxi; Lapinsky, Andrew; Bullard, Blair et al. (2015) Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles. In : *The Plant journal : for cell and molecular biology*, vol. 83, n° 6, p. 1097–1113. DOI: 10.1111/tpj.12944.

Poupin, Nathalie; Vinson, Florence; Moreau, Arthur; Batut, Aurélie; Chazalviel, Maxime; Colsch, Benoit et al. (2020) Improving lipid mapping in Genome Scale Metabolic Networks using ontologies. In : *Metabolomics : Official journal of the Metabolomic Society*, vol. 16, n° 4, p. 44. DOI: 10.1007/s11306-020-01663-5.

Pramanik, J.; Keasling, J. D. (1998) Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. In : *Biotechnology and Bioengineering*, vol. 60, n° 2, p. 230–238. DOI: 10.1002/(SICI)1097-0290(19981020)60:2<230::AID-BIT10>3.0.CO;2-Q.

Pruvost, J.; Cornet, J. F. (2012) Microalgal biotechnology. DOI: 10.1515/9783110225020.

Qian, Xiao; Kim, Min Kyung; Kumaraswamy, G. Kenchappa; Agarwal, Ananya; Lun, Desmond S.; Dismukes, G. Charles (2017) Flux balance analysis of photoautotrophic metabolism: Uncovering new biological details of subsystems involved in cyanobacterial photosynthesis. In : *Biochimica et biophysica acta. Bioenergetics*, vol. 1858, n° 4, p. 276–287. DOI: 10.1016/j.bbabi.2016.12.007.

Radakovits, Randor; Jinkerson, Robert E.; Fuerstenberg, Susan I.; Tae, Hongseok; Settlage, Robert E.; Boore, Jeffrey L.; Posewitz, Matthew C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. In : *Nature communications*, vol. 3, p. 686. DOI: 10.1038/ncomms1688.

Rafay, Ramis; Uratani, Joao M.; Hernandez, Hector H.; Rodríguez, Jorge (2020) Growth and Nitrate Uptake in *Nannochloropsis gaditana* and *Tetraselmis chuii* Cultures Grown in Sequential Batch Reactors. In : *Frontiers in Marine Science*, vol. 7. DOI: 10.3389/fmars.2020.00077.

Ranganathan, Sridhar; Suthers, Patrick F.; Maranas, Costas D. (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. In : *PLoS computational biology*, vol. 6, n° 4, e1000744. DOI: 10.1371/journal.pcbi.1000744.

Ranjbar, Saeed; Malcata, F. Xavier (2022) Challenges and prospects for sustainable microalga-based oil: A comprehensive review, with a focus on metabolic and genetic engineering. In : *Fuel*, vol. 324, p. 124567. DOI: 10.1016/j.fuel.2022.124567.

Ravikrishnan, Aarthi; Raman, Karthik (2015) Critical assessment of genome-scale metabolic networks: the need for a unified standard. In : *Briefings in bioinformatics*, vol. 16, n° 6, p. 1057–1068. DOI: 10.1093/bib/bbv003.

Ray, Ayusmita; Kundu, Pritam; Ghosh, Amit (2023) Reconstruction of a Genome-Scale Metabolic Model of *Scenedesmus obliquus* and Its Application for Lipid Production under Three Trophic Modes. In : *ACS synthetic biology*. DOI: 10.1021/acssynbio.3c00516.

Razzak, Shaikh A.; Hossain, Mohammad M.; Lucky, Rahima A.; Bassi, Amarjeet S.; Lasa, Hugo de (2013) Integrated CO<sub>2</sub> capture, wastewater treatment and biofuel production by microalgae culturing—A review. In : *Renewable and Sustainable Energy Reviews*, vol. 27, p. 622–653. DOI: 10.1016/j.rser.2013.05.063.

Reimers, Alexandra-M; Knoop, Henning; Bockmayr, Alexander; Steuer, Ralf (2017) Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. In : *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, n° 31, E6457-E6465. DOI: 10.1073/pnas.1617508114.

Ren, Ming; Ogden, Kimberly (2014) Cultivation of *Nannochloropsis gaditana* on mixtures of nitrogen sources. In : *Environmental Progress & Sustainable Energy*, vol. 33, n° 2, p. 551–555. DOI: 10.1002/ep.11818.

Resat, Haluk; Petzold, Linda; Pettigrew, Michel F. (2009) Kinetic modeling of biological systems. In : *Methods in molecular biology (Clifton, N.J.)*, vol. 541, p. 311–335. DOI: 10.1007/978-1-59745-243-4\_14.

Rizwan, Muhammad; Mujtaba, Ghulam; Memon, Sheraz Ahmed; Lee, Kisay; Rashid, Naim (2018) Exploring the potential of microalgae for new biotechnology applications and beyond: A review. In : *Renewable and Sustainable Energy Reviews*, vol. 92, p. 394–404. DOI: 10.1016/j.rser.2018.04.034.

Rock, Alberto; Novoveská, Lucie; Green, David (2021) Synthetic biology is essential to unlock commercial biofuel production through hyper lipid-producing microalgae: a review. In : *Applied Phycology*, vol. 2, n° 1, p. 41–59. DOI: 10.1080/26388081.2021.1886872.

Røkke, Gunvor Bjerkelund; Hohmann-Marriott, Martin Frank; Almaas, Eivind (2020) An adjustable algal chloroplast plug-and-play model for genome-scale metabolic models. In : *PloS one*, vol. 15, n° 2, e0229408. DOI: 10.1371/journal.pone.0229408.

Rowe, Elliot; Palsson, Bernhard O.; King, Zachary A. (2018) Escher-FBA: a web application for interactive flux balance analysis. In : *BMC systems biology*, vol. 12, n° 1, p. 84. DOI: 10.1186/s12918-018-0607-5.

Sahu, Ankur; Blätke, Mary-Ann; Szymański, Jędrzej Jakub; Töpfer, Nadine (2021) Advances in flux balance analysis by integrating machine learning and mechanism-based models. In : *Computational and structural biotechnology journal*, vol. 19, p. 4626–4640. DOI: 10.1016/j.csbj.2021.08.004.

Sajjadi, Baharak; Chen, Wei-Yin; Raman, Abdul. Aziz. Abdul; Ibrahim, Shaliza (2018) Microalgae lipid and biomass for biofuel production: A comprehensive review on lipid enhancement strategies and their effects on fatty acid composition. In : *Renewable and Sustainable Energy Reviews*, vol. 97, p. 200–232. DOI: 10.1016/j.rser.2018.07.050.

Sánchez, Benjamín J.; Li, Feiran; Kerkhoven, Eduard J.; Nielsen, Jens (2019) SLIMEr: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. In : *BMC systems biology*, vol. 13, n° 1, p. 4. DOI: 10.1186/s12918-018-0673-8.

Sánchez, Benjamín J.; Zhang, Cheng; Nilsson, Avlant; Lahtvee, Petri-Jaan; Kerkhoven, Eduard J.; Nielsen, Jens (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. In : *Molecular systems biology*, vol. 13, n° 8, p. 935. DOI: 10.15252/msb.20167411.

Savinell, J. M.; Palsson, B. O. (1992) Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. In : *Journal of theoretical biology*, vol. 154, n° 4, p. 421–454. DOI: 10.1016/s0022-5193(05)80161-4.

Savvidou, Maria G.; Boli, Elenitsa; Logothetis, Dimitrios; Lymperopoulou, Theopisti; Ferraro, Angelo; Louli, Vasiliki et al. (2020) A Study on the Effect of Macro- and Micro- Nutrients on *Nannochloropsis oceanica* Growth, Fatty Acid Composition and Magnetic Harvesting Efficiency. In : *Plants (Basel, Switzerland)*, vol. 9, n° 5. DOI: 10.3390/plants9050660.

Schaechter, M.; Maaløe, O.; Kjeldgaard, N. O. (1958) Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella typhimurium*. In : *Microbiology*, vol. 19, n° 3, p. 592–606. DOI: 10.1099/00221287-19-3-592.

Schneider, Philipp; Kamp, Axel von; Klamt, Steffen (2020) An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets. In : *PLoS computational biology*, vol. 16, n° 7, e1008110. DOI: 10.1371/journal.pcbi.1008110.

Schnitzer, Barbara; Österberg, Linnea; Cvijovic, Marija (2022) The choice of the objective function in flux balance analysis is crucial for predicting replicative lifespans in yeast. In : *PloS one*, vol. 17, n° 10, e0276112. DOI: 10.1371/journal.pone.0276112.

Scholz, Matthew J.; Weiss, Taylor L.; Jinkerson, Robert E.; Jing, Jia; Roth, Robyn; Goodenough, Ursula et al. (2014) Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall. In : *Eukaryotic cell*, vol. 13, n° 11, p. 1450–1464. DOI: 10.1128/EC.00183-14.

Schomburg, Ida; Hofmann, Oliver; Baensch, Claudia; Chang, Antje; Schomburg, Dietmar (2000) Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. In : *Gene Function & Disease*, vol. 1, n° 3-4, p. 109–118. DOI: 10.1002/1438-826X(200010)1:3/4<109::AID-GNFD109>3.0.CO;2-O.

Schuetz, Robert; Kuepfer, Lars; Sauer, Uwe (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. In : *Molecular systems biology*, vol. 3, p. 119. DOI: 10.1038/msb4100162.

Schulz, Christian; Kumelj Tjasa; Karlsen, Emil; Almaas, Eivind (2021) Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. In : *PLoS computational biology*. DOI: 10.1371/journal.pcbi.1008528.

Schuster, Stefan; Pfeiffer, Thomas; Fell, David A. (2008) Is maximization of molar yield in metabolic networks favoured by evolution? In : *Journal of theoretical biology*, vol. 252, n° 3, p. 497–504. DOI: 10.1016/j.jtbi.2007.12.008.

Seaver, Samuel M. D.; Lerma-Ortiz, Claudia; Conrad, Neal; Mikaili, Arman; Sreedasyam, Avinash; Hanson, Andrew D.; Henry, Christopher S. (2018) PlantSEED enables automated annotation and reconstruction of plant primary metabolism with improved compartmentalization and comparative consistency. In : *The Plant journal : for cell and molecular biology*, vol. 95, n° 6, p. 1102–1113. DOI: 10.1111/tpj.14003.

Seaver, Samuel M. D.; Liu, Filipe; Zhang, Qizhi; Jeffryes, James; Faria, José P.; Edirisinghe, Janaka N. et al. (2021) The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. In : *Nucleic acids research*, vol. 49, n° D1, D575-D588. DOI: 10.1093/nar/gkaa746.

Segrè, Daniel; Vitkup, Denis; Church George M. (2002) Analysis of optimality in natural and perturbed metabolic networks. In : *PNAS*, vol. 99, n° 23, p. 15112–15117.

Seif, Yara; Palsson, Bernhard Ørn (2021) Path to improving the life cycle and quality of genome-scale models of metabolism. In : *Cell systems*, vol. 12, n° 9, p. 842–859. DOI: 10.1016/j.cels.2021.06.005.

Serrano-Bermúdez, Luis Miguel; González Barrios, Andrés Fernando; Maranas, Costas D.; Montoya, Dolly (2017) *Clostridium butyricum* maximizes growth while minimizing enzyme usage and ATP production: metabolic flux distribution of a strain cultured in glycerol. In : *BMC systems biology*, vol. 11, n° 1, p. 58. DOI: 10.1186/s12918-017-0434-0.

Shah, Ab Rauf; Ahmad, Ahmad; Srivastava, Shireesh; Jaffar Ali, B. M. (2017) Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. In : *Algal Research*, vol. 26, p. 354–364. DOI: 10.1016/j.algal.2017.08.014.

Shene, Carolina; Asenjo, Juan A.; Chisti, Yusuf (2018) Metabolic modelling and simulation of the light and dark metabolism of *Chlamydomonas reinhardtii*. In : *The Plant journal : for cell and molecular biology*, vol. 96, n° 5, p. 1076–1088. DOI: 10.1111/tpj.14078.

Shlomi, Tomer; Berkman, Omer; Ruppin, Eytan (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. In : *PNAS*, vol. 102, n° 21, p. 7695–7700.

Simensen, Vetle; Schulz, Christian; Karlsen, Emil; Bråtelund, Signe; Burgos, Idun; Thorfinnsdottir, Lilja Brekke et al. (2022) Experimental determination of *Escherichia coli* biomass composition for constraint-based metabolic modeling. In : *PloS one*, vol. 17, n° 1, e0262450. DOI: 10.1371/journal.pone.0262450.

Simionato, Diana; Block, Maryse A.; La Rocca, Nicoletta; Jouhet, Juliette; Maréchal, Eric; Finazzi, Giovanni; Morosinotto, Tomas (2013) The response of *Nannochloropsis gaditana* to nitrogen starvation includes de novo biosynthesis of triacylglycerols, a decrease of chloroplast galactolipids, and reorganization of the photosynthetic apparatus. In : *Eukaryotic cell*, vol. 12, n° 5, p. 665–676. DOI: 10.1128/EC.00363-12.

Song, Si-Young; Kato, Chieko; Adachi, Eijiro; Moriya-Sato, Ayako; Inagawa-Ogashiwa, Masayo; Umeda, Rieko; Hashimoto, Naohiro (2007) Expression of an acyl-CoA synthetase, lipidosin, in astrocytes of the murine brain and its up-regulation during remyelination following cuprizone-induced demyelination. In : *Journal of neuroscience research*, vol. 85, n° 16, p. 3586–3597. DOI: 10.1002/jnr.21456.

Spolaore, Pauline; Joannis-Cassan, Claire; Duran, Elie; Isambert, Arsène (2006) Commercial applications of microalgae. In : *Journal of bioscience and bioengineering*, vol. 101, n° 2, p. 87–96. DOI: 10.1263/jbb.101.87.

Steinberg, S. J.; Morgenthaler, J.; Heinzer, A. K.; Smith, K. D.; Watkins, P. A. (2000) Very long-chain acyl-CoA synthetases. Human "bubblegum" represents a new family of proteins capable of activating very long-chain fatty acids. In : *The Journal of biological chemistry*, vol. 275, n° 45, p. 35162–35169. DOI: 10.1074/jbc.M006403200.

Straathof, Adrie J. J.; Wahl, S. Aljoscha; Benjamin, Kirsten R.; Takors, Ralf; Wierckx, Nick; Noorman, Henk J. (2019) Grand Research Challenges for Sustainable Industrial Biotechnology. In : *Trends in biotechnology*, vol. 37, n° 10, p. 1042–1050. DOI: 10.1016/j.tibtech.2019.04.002.

Sundaram, Thanigaivel; Rajendran, Saravanan; Gnanasekaran, Lalitha; Rachmadona, Nova; Jiang, Jheng-Jie; Khoo, Kuan Shiong; Show, Pau Loke (2023) Bioengineering strategies of microalgae biomass for biofuel production: recent advancement and insight. In : *Bioengineered*, vol. 14, n° 1, p. 2252228. DOI: 10.1080/21655979.2023.2252228.

Suthers, Patrick F.; Zomorodi, Alireza; Maranas, Costas D. (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. In : *Molecular systems biology*, vol. 5, p. 301. DOI: 10.1038/msb.2009.56.

Széliová, Diana; Ruckerbauer, David E.; Galleguillos, Sarah N.; Petersen, Lars B.; Natter, Klaus; Hanscho, Michael et al. (2020) What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. In : *Metabolic engineering*, vol. 61, p. 288–300. DOI: 10.1016/j.ymben.2020.06.002.

Taleb, A.; Kandilian, R.; Touchard, R.; Montalescot, V.; Rinaldi, T.; Taha, S. et al. (2016) Screening of freshwater and seawater microalgae strains in fully controlled photobioreactors for biodiesel production. In : *Bioresource technology*, vol. 218, p. 480–490. DOI: 10.1016/j.biortech.2016.06.086.

Thiele, Ines; Palsson, Bernhard Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. In : *Nature protocols*, vol. 5, n° 1, p. 93–121. DOI: 10.1038/nprot.2009.203.

Tibbetts, Sean M.; Bjornsson, William J.; McGinn, Patrick J. (2015) Biochemical composition and amino acid profiles of *Nannochloropsis granulata* algal biomass before and after supercritical fluid CO<sub>2</sub> extraction at two processing temperatures. In : *Animal Feed Science and Technology*, vol. 204, p. 62–71. DOI: 10.1016/j.anifeedsci.2015.04.006.

Tibocha-Bonilla, Juan D.; Zuniga, Cristal; Broddrick, Jared T.; Zengler, Karsten; Godoy-Silva, Rubén D. (2021) A multiscale model predicts the sensitivity of *Chlorella vulgaris* to light and nitrogen levels in photobioreactors. In : *bioRxiv*. DOI: 10.1101/2021.04.14.439858.

Tibocha-Bonilla, Juan D.; Zuñiga, Cristal; Godoy-Silva, Rubén D.; Zengler, Karsten (2018) Advances in metabolic modeling of oleaginous microalgae. In : *Biotechnology for biofuels*, vol. 11, p. 241.

Tolteer, Dimitri; Smith, Edward N.; Dupont-Thibert, Clémence; Uwizeye, Clarisse; Vile, Denis; Gloaguen, Pauline et al. (2023) An evolutive quantitative cellular/subcellular atlas of *Arabidopsis* reference leaf 6. In : *Submitted*.

van Tol, Helena M.; Armbrust, E. Virginia (2021) Genome-scale metabolic model of the diatom *Thalassiosira pseudonana* highlights the importance of nitrogen and sulfur metabolism in redox balance. In : *PloS one*, vol. 16, n° 3, e0241960. DOI: 10.1371/journal.pone.0241960.

Varma, A.; Palsson, B. O. (1994a) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. In : *Bio/Technology*, vol. 12, n° 10, p. 994–998. DOI: 10.1038/nbt1094-994.

Varma, A.; Palsson, B. O. (1994b) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. In : *Applied and environmental microbiology*, vol. 60, n° 10, p. 3724–3731. DOI: 10.1128/aem.60.10.3724-3731.1994.

Ventorim, Rafaela Zandonade; Ferreira, Maurício Alexander de Moura; Almeida, Eduardo Luís Menezes de; Kerkhoven, Eduard J.; Da Silveira, Wendel Batista (2022) Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. In : *Biochemical Engineering Journal*, vol. 180, p. 108353. DOI: 10.1016/j.bej.2022.108353.

Vieira, Vítor; Maia, Paulo; Rocha, Miguel; Rocha, Isabel (2019) Comparison of pathway analysis and constraint-based methods for cell factory design. In : *BMC bioinformatics*, vol. 20, n° 1, p. 350. DOI: 10.1186/s12859-019-2934-y.

Vieler, Astrid; Wu, Guangxi; Tsai, Chia-Hong; Bullard, Blair; Cornish, Adam J.; Harvey, Christopher et al. (2012) Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. In : *PLoS genetics*, vol. 8, n° 11, e1003064. DOI: 10.1371/journal.pgen.1003064.

VIZBI. Visualizing Biological Data. En ligne : <https://vizbi.org/>, consulté le 2023.

Volkman, John; Brown, Malcolm; Dunstan, Greame; Jeffrey S. (1993) The biochemical composition of marine microalgae from the class eustigmatophyceae. In : *Phycologia*, vol. 29, p. 69–78.

Vu, Trang T.; Hill, Eric A.; Kucek, Leo A.; Konopka, Allan E.; Beliaev, Alexander S.; Reed, Jennifer L. (2013) Computational evaluation of *Synechococcus* sp. PCC 7002 metabolism for chemical production. In : *Biotechnology journal*, vol. 8, n° 5, p. 619–630. DOI: 10.1002/biot.201200315.

Wang, Hao; Marcišauskas, Simonas; Sánchez, Benjamín J.; Domenzain, Iván; Hermansson, Daniel; Agren, Rasmus et al. (2018) RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. In : *PLOS Computational Biology*, vol. 14, n° 10, e1006541. DOI: 10.1371/journal.pcbi.1006541.

Wiechert, W. (2001) <sup>13</sup>C metabolic flux analysis. In : *Metabolic engineering*, vol. 3, n° 3, p. 195–206. DOI: 10.1006/mben.2001.0187.

Witting, Michael (2020) Suggestions for Standardized Identifiers for Fatty Acyl Compounds in Genome Scale Metabolic Models and Their Application to the WormJam *Caenorhabditis elegans* Model. In : *Metabolites*, vol. 10, n° 4. DOI: 10.3390/metabo10040130.

Xavier, Joana C.; Patil, Kiran Raosaheb; Rocha, Isabel (2017) Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. In : *Metabolic engineering*, vol. 39, p. 200–208. DOI: 10.1016/j.ymben.2016.12.002.

Xiong, Wei; Liu, Lixia; Wu, Chao; Yang, Chen; Wu, Qingyu (2010) <sup>13</sup>C-tracer and gas chromatography-mass spectrometry analyses reveal metabolic flux distribution in the oleaginous microalga *Chlorella protothecoides*. In : *Plant physiology*, vol. 154, n° 2, p. 1001–1011. DOI: 10.1104/pp.110.158956.

Xu, Fang; Cong, Wei; Cai, Zhao-Ling; Ouyang, Fan (2004) Effects of organic carbon sources on cell growth and eicosapentaenoic acid content of *Nannochloropsis* sp. In : *Journal of applied phycology*, vol. 16, p. 499–503.

Xu, Zixiang; Zheng, Ping; Sun, Jibin; Ma, Yanhe (2013) ReactKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. In : *PloS one*, vol. 8, n° 12, e72150. DOI: 10.1371/journal.pone.0072150.

Yasemi, Mohammadreza; Jolicoeur, Mario (2021) Modelling Cell Metabolism: A Review on Constraint-Based Steady-State and Kinetic Approaches. In : *Processes*, vol. 9, n° 2, p. 322. DOI: 10.3390/pr9020322.

Ye, Chao; Wei, Xinyu; Shi, Tianqiong; Sun, Xiaoman; Xu, Nan; Gao, Cong; Zou, Wei (2022) Genome-scale metabolic network models: from first-generation to next-generation. In : *Applied microbiology and biotechnology*, vol. 106, n° 13-16, p. 4907–4920. DOI: 10.1007/s00253-022-12066-y.

Yoshikawa, Katsunori; Kojima, Yuta; Nakajima, Tsubasa; Furusawa, Chikara; Hirasawa, Takashi; Shimizu, Hiroshi (2011) Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC6803. In : *Applied microbiology and biotechnology*, vol. 92, n° 2, p. 347–358. DOI: 10.1007/s00253-011-3559-x.

Zhang, Cheng; Hua, Qiang (2015) Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. In : *Frontiers in physiology*, vol. 6, p. 413. DOI: 10.3389/fphys.2015.00413.

Zhang, Dongmei; Yan, Fei; Sun, Zhongliang; Zhang, Qinghua; Xue, Shengzhang; Cong, Wei (2014) On-line modeling intracellular carbon and energy metabolism of *Nannochloropsis* sp. in nitrogen-repletion and nitrogen-limitation cultures. In : *Bioresource technology*, vol. 164, p. 86–92. DOI: 10.1016/j.biortech.2014.04.083.

Zhao, Qi; Stettner, Arion I.; Reznik, Ed; Paschalidis, Ioannis Ch; Segrè, Daniel (2016) Mapping the landscape of metabolic goals of a cell. In : *Genome biology*, vol. 17, n° 1, p. 109. DOI: 10.1186/s13059-016-0968-2.

Zhu, Shunni; Huang, Wei; Xu, Jin; Wang, Zhongming; Xu, Jingliang; Yuan, Zhenhong (2014) Metabolic changes of starch and lipid triggered by nitrogen starvation in the microalga *Chlorella zofingiensis*. In : *Bioresource technology*, vol. 152, p. 292–298. DOI: 10.1016/j.biortech.2013.10.092.

Zienkiewicz, Krzysztof; Du, Zhi-Yan; Ma, Wei; Vollheyde, Katharina; Benning, Christoph (2016) Stress-induced neutral lipid biosynthesis in microalgae - Molecular, cellular and physiological insights. In : *Biochimica et biophysica acta*, vol. 1861, n° 9 Pt B, p. 1269–1281. DOI: 10.1016/j.bbalip.2016.02.008.

Zimorski, Verena; Ku, Chuan; Martin, William F.; Gould, Sven B. (2014) Endosymbiotic theory for organelle origins. In : *Current opinion in microbiology*, vol. 22, p. 38–48. DOI: 10.1016/j.mib.2014.09.008.

Zuñiga, Cristal; Levering, Jennifer; Antoniewicz, Maciek R.; Guarnieri, Michael T.; Betenbaugh, Michael J.; Zengler, Karsten (2018) Predicting Dynamic Metabolic Demands in the Photosynthetic Eukaryote *Chlorella vulgaris*. In : *Plant physiology*, vol. 176, n° 1, p. 450–462. DOI: 10.1104/pp.17.00605.

Zuñiga, Cristal; Li, Chien-Ting; Huelsman, Tyler; Levering, Jennifer; Zielinski, Daniel C.; McConnell, Brian O. et al. (2016) Genome-Scale Metabolic Model for the Green Alga *Chlorella vulgaris* UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. In : *Plant physiology*, vol. 172, n° 1, p. 589–602. DOI: 10.1104/pp.16.00593.

Certains éléments des figures du manuscrit sont adaptés de Servier Medical Art (smart.servier.com).

# Modélisation du métabolisme de la microalgue oléagineuse d'intérêt industriel *Microchloropsis gaditana*.

Dans le cadre actuel des transitions écologiques et énergétiques, les microalgues suscitent un intérêt croissant en tant que ressources potentielles pour diverses applications industrielles. Parmi ces organismes, des souches oléagineuses telles que *Microchloropsis gaditana*, se positionnent comme candidates prometteuses pour la production de biodiesels en raison de leur capacité à accumuler d'importantes quantités de lipides, notamment des triacylglycérols. Néanmoins, de nombreux verrous technologiques doivent encore être levés pour permettre une exploitation à l'échelle industrielle de ces microalgues. Lever ces verrous requiert une compréhension approfondie du métabolisme de ces souches, en particulier du métabolisme lipidique.

C'est dans ce cadre de recherche sur la caractérisation et la compréhension du métabolisme lipidique de *M. gaditana* que s'inscrit ce travail de thèse.

Dans un premier temps, les voies de synthèse et de dégradation des lipides d'un modèle métabolique à l'échelle du génome de *M. gaditana*, disponible en amont de la thèse, ont été corrigées à partir des connaissances les plus avancées du métabolisme lipidique provenant des experts du LPCV et de la littérature. Ensuite, BioModTool, un package Python, a été développé pour faciliter l'intégration de fonctions objectives de biomasse dans les modèles métaboliques à partir de données expérimentales telles que des glycérolipidomes. BioModTool a été utilisé pour incorporer le profil lipidique détaillé d'un mutant généré au laboratoire, le mutant *acsbg*, au modèle métabolique de *M. gaditana*. Ce mutant présente un profil lipidique intéressant, notamment avec une accumulation de triacylglycérols huit fois plus élevée que celle de la souche sauvage. Des simulations de distributions de flux ont été réalisées pour ce mutant et comparées aux flux prédits pour la souche sauvage. Les résultats de cette analyse confortent les conclusions de l'analyse expérimentale du mutant *acsbg*.

Les méthodes développées dans cette thèse, ainsi que les résultats obtenus ont un potentiel important pour accroître notre compréhension du métabolisme lipidique de *M. gaditana*. Il s'agit d'un des premiers modèles de microalgue incorporant l'intégralité de nos connaissances du métabolisme lipidique à ce jour. Il pourra donc être utilisé pour toutes les applications traditionnellement effectuées avec les modèles métaboliques à l'échelle du génome.

Une meilleure connaissance du métabolisme des lipides et des approches *in silico* sont nécessaires à l'exploitation des microalgues pour des applications industrielles telles que la production de biocarburants. Le modèle métabolique développé au cours de cette thèse facilitera l'optimisation des procédés de culture et le développement de souches génétiquement modifiées dans le but d'accroître la production de lipides.

**Mots clés :** modélisation métabolique à l'échelle du génome, métabolisme lipidique, microalgue, *Microchloropsis gaditana*, biocarburant.

# Metabolic modeling of the industrial oleaginous microalga *Microchloropsis gaditana*.

With the ongoing ecological and energy transitions, microalgae have attracted increasing interest as potential resources for a wide range of industrial applications. Among these organisms, oleaginous strains such as *Microchloropsis gaditana* stand out as promising candidates for biodiesel production, due to their capacity to accumulate large amounts of lipids, particularly triglycerides. However, several technological challenges remain before achieving industrial scales. To overcome these challenges, an in-depth understanding of the metabolism, specifically lipid metabolism of these strains is required. It is within this framework of research into the characterization and understanding of lipid metabolism of *M. gaditana* that this PhD project is situated.

First, lipid biosynthesis and degradation pathways of a genome-scale metabolic model of *M. gaditana*, available before of the PhD, were meticulously corrected by incorporating the most advanced knowledge of lipid metabolism from LPCV specialists and the literature. Then, BioModTool, a Python package, was developed to facilitate addition of biomass objective functions into metabolic models from experimental data such as glycerolipidomes. BioModTool was used to integrate the detailed lipid profile of a laboratory-generated mutant, the *acsbg* mutant, into *M. gaditana* metabolic model. This mutant exhibits an interesting phenotype, notably with an eight-fold increase in triglycerides accumulation compared to the wild-type strain. Flux distributions were simulated for this mutant and compared with fluxes predicted for the wild-type strain. The results of this analysis confirm the conclusions of the experimental analysis of the *acsbg* mutant.

The approaches developed and the results obtained in this project hold significant promise for enhancing our understanding of lipid metabolism in *M. gaditana*. To date, the model developed in this study stands as one of the first microalgal models to incorporate such detailed lipid metabolism. As a result, it constitutes a valuable tool for a wide range of applications traditionally performed with genome-scale metabolic models. Deciphering lipid metabolism and *in silico* approaches are needed to industrial exploitation of microalgae for diverse applications, including biofuel production. The metabolic model developed during this PhD will enable the optimization of cultivation processes and the design of new strains with enhanced lipid production.

**Keywords:** genome-scale metabolic modeling, lipid metabolism, microalga, *Microchloropsis gaditana*, biofuel.

---

## CEA Grenoble

Laboratoire de Physiologie Cellulaire et Végétale –UMR 5168 CEA / CNRS / Université Grenoble  
Alpes –UMR 1417 INRA  
17 avenue des Martyrs  
38054 Grenoble cedex 9

## TotalEnergies

Programme Biofuels, OneTech  
Centre de Recherches de Solaize TotalEnergies,  
Chemin du canal  
69 360 Solaize