

INRAE

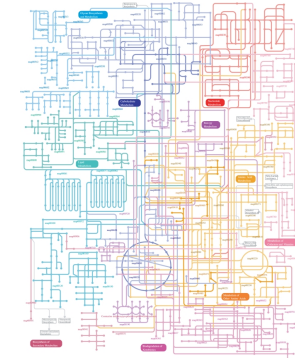
➤ WP1 : focus sur l'ontologie et l'encodage des données

Olivier & Anne

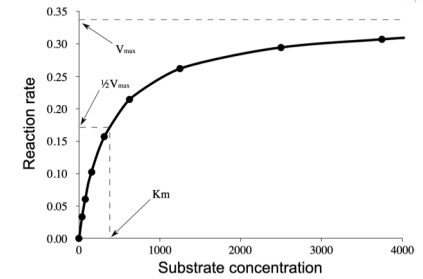
* Contexte du WP1

- Les données et la bio informatique : omics, big, fragmentées ...
- Quels sont les **objets formels** (onotologies) existants pour manipuler ces données ?
- Par ex : BiOPAX et GO à des fins d'annotation ou d'échange.
- Par contre pas d'objet formel pour organiser les données à l'échelle de l'organisme entier.
- Autre élément du contexte : il existe plusieurs **objets biologiques** avec des données associées.
- Objets biologiques : objets produits par une communauté scientifique.
- Ex : les enzymes (communauté des enzymologistes), le réseaux métaboliques (communauté des biologistes des systèmes), « molécules » (communautés des biochimistes)
- Les objets produits sont **isolés** les uns des autres et se situent à des **échelles différentes**.
- Le fonctionnement de ces objets biologiques peuvent être décrit pas un ou plusieurs « type » modèles mathématiques
- Ex : équation de Michaelis Menten pour décrire la cinétique des enzymes, modèles sous forme de contraintes pour décrire le fonctionnement les réseaux métaboliques
-

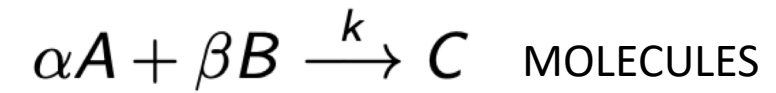
* Les objets biologiques



METABOLIC NETWORK



ENZYMES



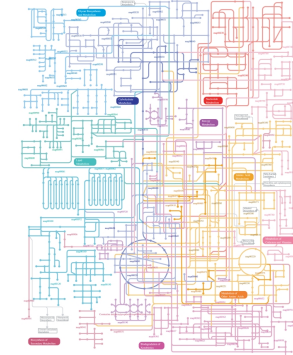
* Les modèles associés



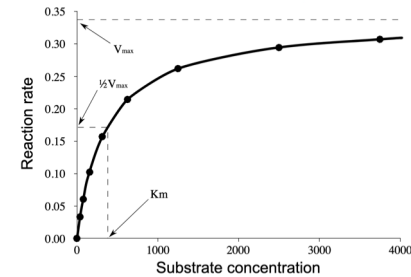
$$V_i = K_e[E]$$

$$V_i = \frac{K_{cat} [S]_0}{K_m [S]_0} [E]$$

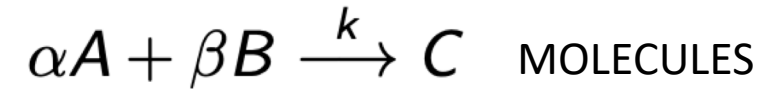
$$\begin{cases} \frac{da}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{db}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{dc}{dt} = k_1 a(t)b(t) - k_{-1} c(t) \\ a(0) = a_0, b(0) = b_0, c(0) = 0. \end{cases}$$



METABOLIC NETWORK



ENZYMES



* Les données associées aux modèles

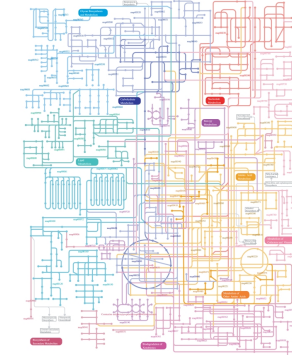


fluxomics

parameter value

proteomics

$$V_i = K_e[E]$$

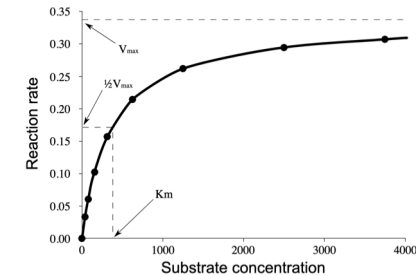


METABOLIC NETWORK

fluxomics

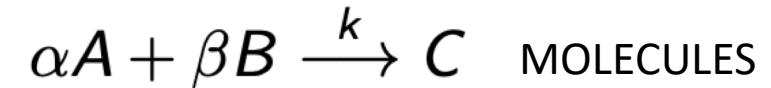
proteomics

$$V_i = \frac{K_{cat} [S]_0}{K_m [S]_0} [E]$$



ENZYMES

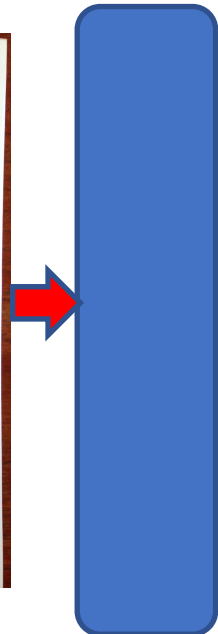
metabolomics



$$\left\{ \begin{array}{l} \frac{da}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{db}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{dc}{dt} = k_1 a(t)b(t) - k_{-1} c(t) \\ a(0) = a_0, b(0) = b_0, c(0) = 0. \end{array} \right.$$

* Un ontologie qui représente le fonctionnement de la bactérie

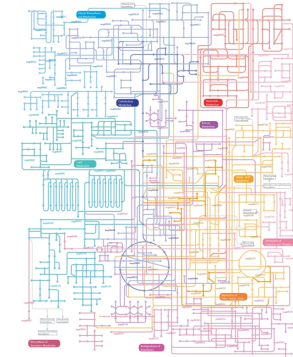
AN ONTOLOGY



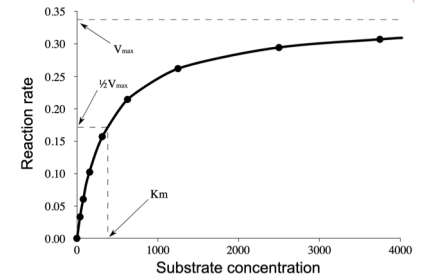
$$V_i = K_e[E]$$

$$V_i = \frac{K_{cat} [S]_0}{K_m [S]_0} [E]$$

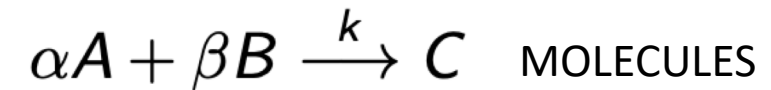
$$\left\{ \begin{array}{l} \frac{da}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{db}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{dc}{dt} = k_1 a(t)b(t) - k_{-1} c(t) \\ a(0) = a_0, b(0) = b_0, c(0) = 0. \end{array} \right.$$



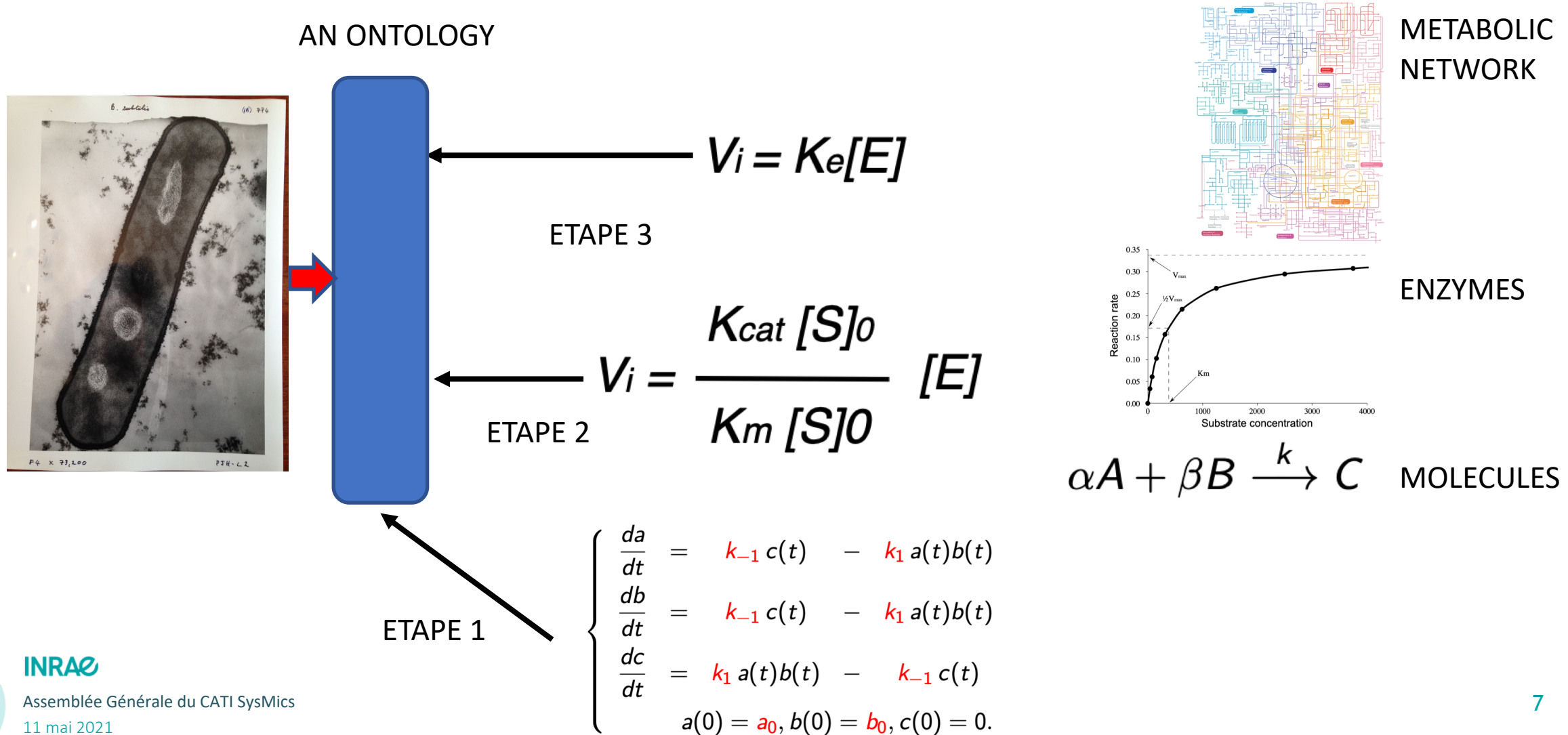
METABOLIC NETWORK



ENZYMES



* Lier les modèles et les données à l'ontologie



* Objectifs du WP1

- Créer un objet formel partageable : un ontologie
- Utiliser cet objet formel pour **relier** entre-eux les modèles des objets biologiques et les données qui leur sont associées.
- Faire ce travail sur une sous-partie de la cellule : le réseau métabolique.
- Donc à minima représenter les métabolites, les réactions et les enzymes.
- Intégrer au minimum au moins 3 types d'omiques : fluxomique, métabolomique et protéomique

* Rappel : construction d'une ontologie

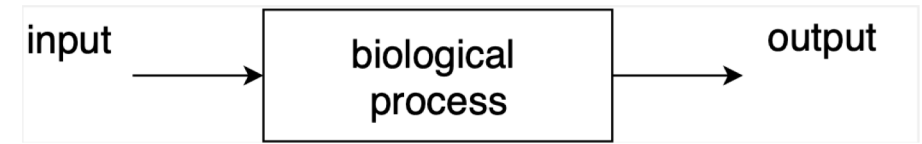
Ontologie construite selon les principes de la systémique (processus et modèles mathématiques associés).

Motivation pour la construction de l'ontologie:

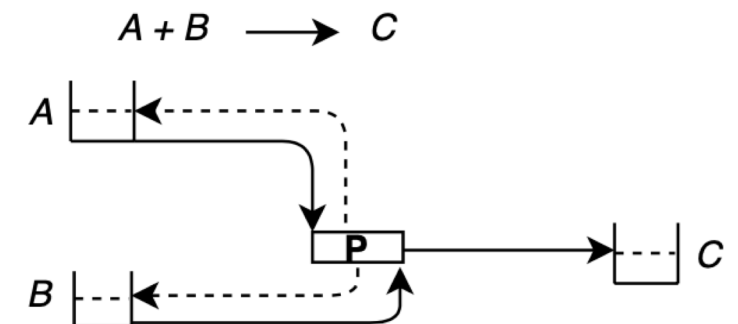
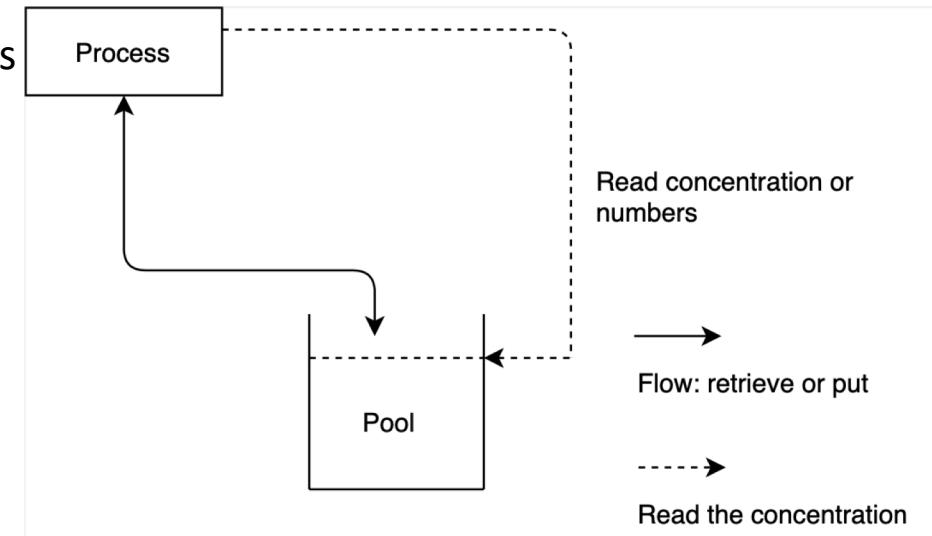
- la communauté des modélisateurs, à travers les principes systémiques a révélé un principe pour organiser les réseaux chez *B.subtilis*
- être en mesure d'exporter ce principe organisateur hors de la communauté des modélisateurs

Résultats :

- Construction d'une ontologie selon 2 contraintes et avec peu de concepts/rerelations (pools, flux, concentration)
- Peuplement de l'ontologie avec l'ensemble des réactions du réseau métabolique de *B.subtilis* : 990 métabolites, 1560 réactions, 1560 enzymes.



$$v_i = \frac{v_{max}[S]_0}{K_M + [S]_0}$$



* Détails de l'étape 1 : Lier les modèles type ODE et l'ontologie

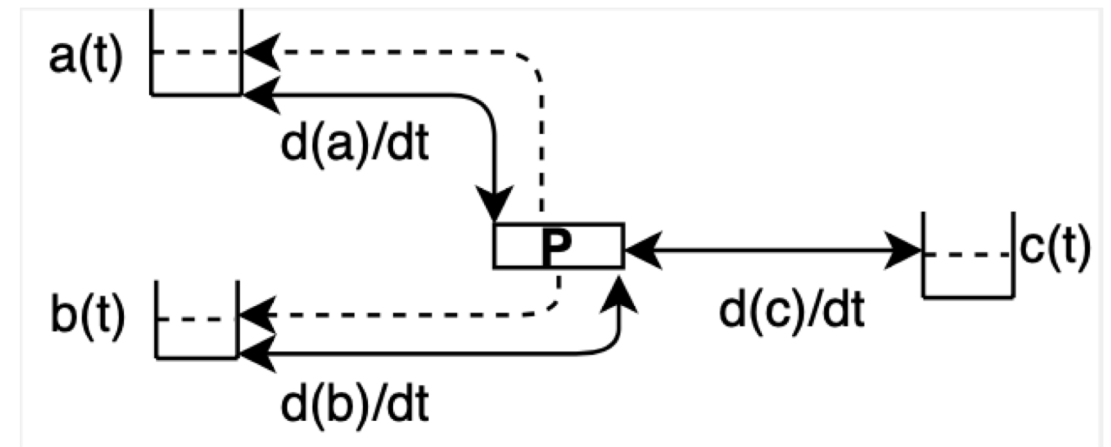
- Lier les modèles à l'ontologie : est-ce que dans le modèle et dans l'ontologie on parle de la même chose ?
- Lier les modèles à l'ontologie : faire un **lien** entre les concepts présents dans les modèles et ceux présents dans l'ontologie.

– Exemple, la réaction : $A + B \xrightleftharpoons[k_{-1}]{k_1} C$ est modélisée et représentée comme suit :

Modèle

$$\left\{ \begin{array}{l} \frac{da}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{db}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{dc}{dt} = k_1 a(t)b(t) - k_{-1} c(t) \\ a(0) = a_0, b(0) = b_0, c(0) = 0. \end{array} \right.$$

Ontologie



=> La concentration (resp. le flux) signifie la même chose dans le modèle et l'ontologie.

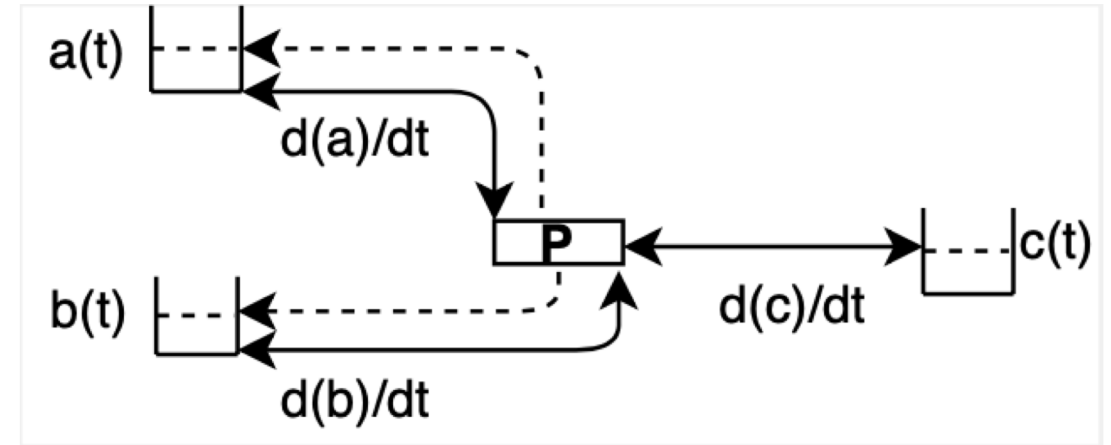
=> A priori oui, nous avons fait un lien entre le modèle et l'ontologie à travers les concentration et les flux.

* Faire le lien pour intégrer les données

Modèle

$$\left\{ \begin{array}{l} \frac{da}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{db}{dt} = k_{-1} c(t) - k_1 a(t)b(t) \\ \frac{dc}{dt} = k_1 a(t)b(t) - k_{-1} c(t) \\ a(0) = a_0, b(0) = b_0, c(0) = 0. \end{array} \right.$$

Ontologie



- les valeurs de paramètres k_1 et k_{-1} pourront être accrochées à l'instance du processus **P** dans l'ontologie.
- => sur cet exemple simple, on voit qu'il est possible d'intégrer les données associées au modèle à travers l'ontologie.

* Résultats

Etapes	Lien	Résultats / Commentaires
ETAPE 1 : liens équations différentielles-ontologie	Fait	Toutes les réaction biochimique peuvent être représentées par l'ontologie
ETAPE 2 : liens Michaelis Menten - ontologie	Fait	i) Un ensemble conséquent de réactions enzymatiques peuvent être représentées par l'ontologie. ii) Pour quelque réactions il faut adapter l'ontologie.
ETAPE 3 : liens RBA-ontologie	A faire	Le lien RBA-ontologie est très « proche » du lien Michaelis Menten-ontologie

- lien ontologie-RBA : un cas propice pour l'intégration de données car nous disposons de données multi-omiques (Projet Basyntec) et des valeurs de paramètre pour près de 600 enzymes qui ont servi à caler le modèle pour *B.subtilis*.

* Avancées sur la reconstruction des processus de la plante

Stratégie mise en place

1. Sélection d'un processus biologique à rajouter
2. Analyse de la littérature + bases de données (Uniprot, AraCyc, KEGG, TAIR, ChloroKB) pour déterminer :
 - la composition de la machine moléculaire (protéines, RNA, cofacteurs, ...), et sa localisation sub-cellulaire
 - les coûts métaboliques de fonctionnement de la machine
3. Modélisation du processus (-> formalisation du fonctionnement de la machine sous forme d'une contrainte égalité ou inégalité compatible avec un modèle allocation de ressources (RBA)).
 - Nécessite souvent d'introduire des paramètres liés au fonctionnement de la machine, ou liés aux caractéristiques des cibles de la machine (ex : durée de vie d'une protéine / efficacité de la protéase)
 - 2ème analyse de la littérature pour déterminer certains de ces paramètres, les autres étant estimés soit sur des données, soit assignés à des valeurs par défaut.
4. Vérifier que le processus ajouté reste compatible avec la croissance de la plante
 - objectif: garantir la consistance de l'information
 - Attention !!! cela ne veut pas dire que l'information est « vraie », uniquement qu'elle reste compatible avec la croissance.
 - pour cela on réalise une simulation du modèle, et on vérifie que les prédictions sont consistantes, c-à-d en accord avec une configuration réaliste de la plante (d'après les données de la littérature + expertise). Dans le cas contraire, on cherche à savoir ce qui manque.

* Processus actuellement reconstruits (et intégrés dans le modèle)

Compartiments cellulaires. cytoplasme, noyau, mitochondrie, chloroplaste, thylakoïde, vacuole, peroxisome, golgi, endoplasmique, matrice extracellulaire. Description fine comprenant les membranes des organelles.

Processus non-métaboliques (PNM)

- Replication : noyau, mitochondrie, chloroplaste
- Transcription : noyau(x3 : Pol1, Pol2, Pol3), mitochondrie, chloroplaste
- Traduction : cytoplasme, mitochondrie, chloroplaste
- Folding : cytoplasme (x2), mitochondrie, chloroplaste (x3)
- Protein, RNA translocation : noyau, mitochondrie (x6), chloroplaste (x2)
- RNA dégradation : cytoplasme
- Protein dégradation : cytoplasme

} A consolider



Première présentation aux partenaires principaux de l'IJPB en avril 2021 (retours très positifs)

Processus (ou réactions) métaboliques Photosynthesis

- Calvin cycle, photorespiration, gluconeogenesis/glycolysis, PPP
- TCA cycle, oxidative phosphorylation
- Starch synthesis/degradation, sucrose synthesis/degradation
- Amino acid, (deoxy)-nucleotide synthesis, nucleotide salvage pathways
- Chlorophyll A and B, Riboflavin, Pyridoxal-5P syntheses
- Transport of metabolites in/out of the cell and of organelles
- Fatty-acids : aggregated on Malonyl-coA
- Cell wall : cellulose, aggregated on dimers of glucose



* Perspectives du *workpackage* 1

- lien ontologie-RBA : un cas propice pour l'intégration de données car nous disposons de données multi-omiques (Projet Basyntec) et des valeurs de paramètres pour près de 600 enzymes qui ont servi à caler le modèle pour *B.subtilis*.
- Objectif : créer une première version de l'entrepôt à partir de ces données.
- Proposition animation CATI SYSMICS : peupler l'entrepôt à travers un **hackathon**.
- Planning prévisionnel :

