

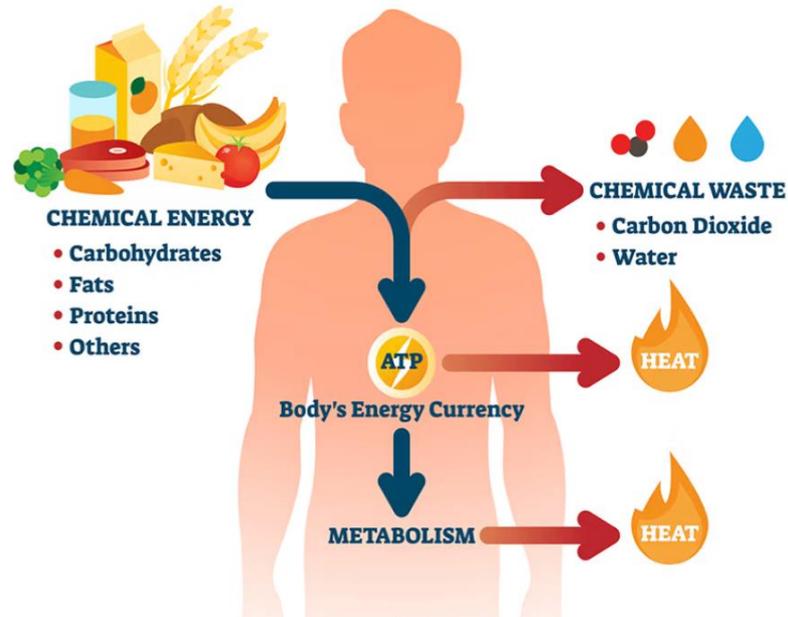


ML génératif pour les études du métabolisme

Ljubisa Miskovic

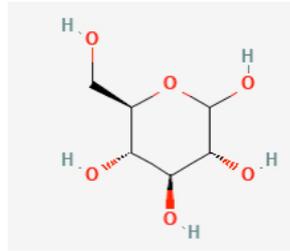
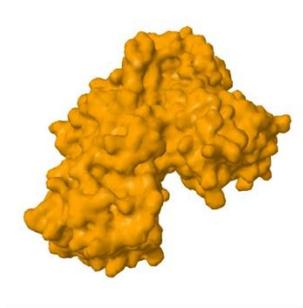
Qu'est-ce que le métabolisme ?

Le métabolisme est l'ensemble des réactions chimiques qui se produisent dans les organismes vivants afin de maintenir la vie.

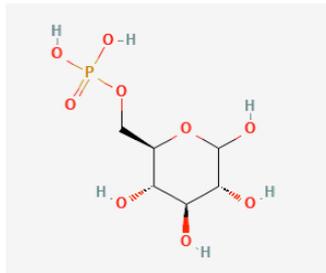
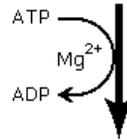


Enzymes

- Les réactions métaboliques sont catalysées par des protéines (**enzymes**).



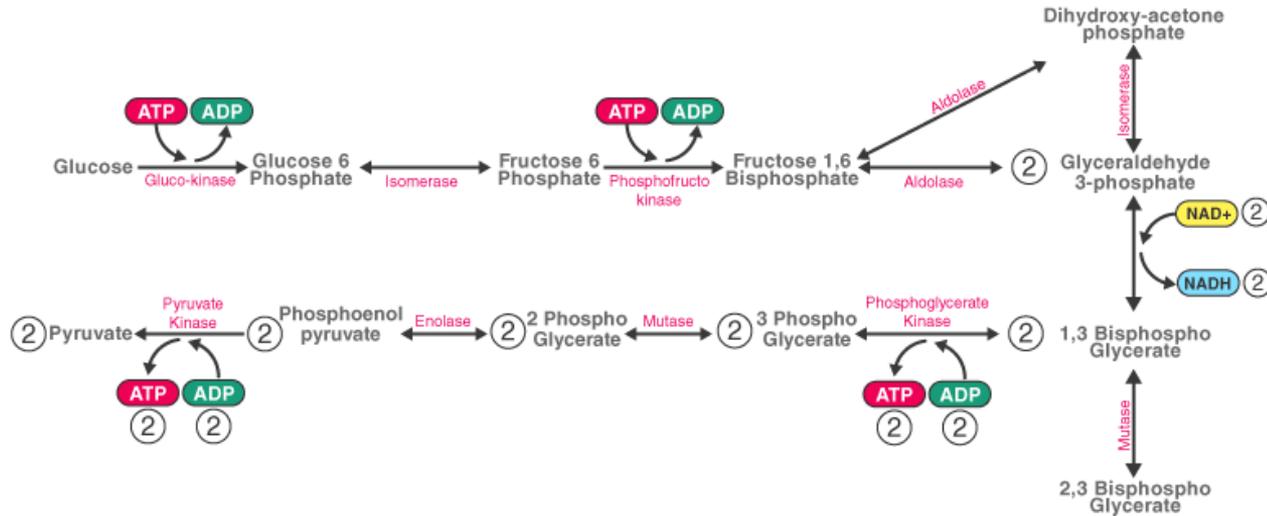
glucose



glucose 6-phosphate

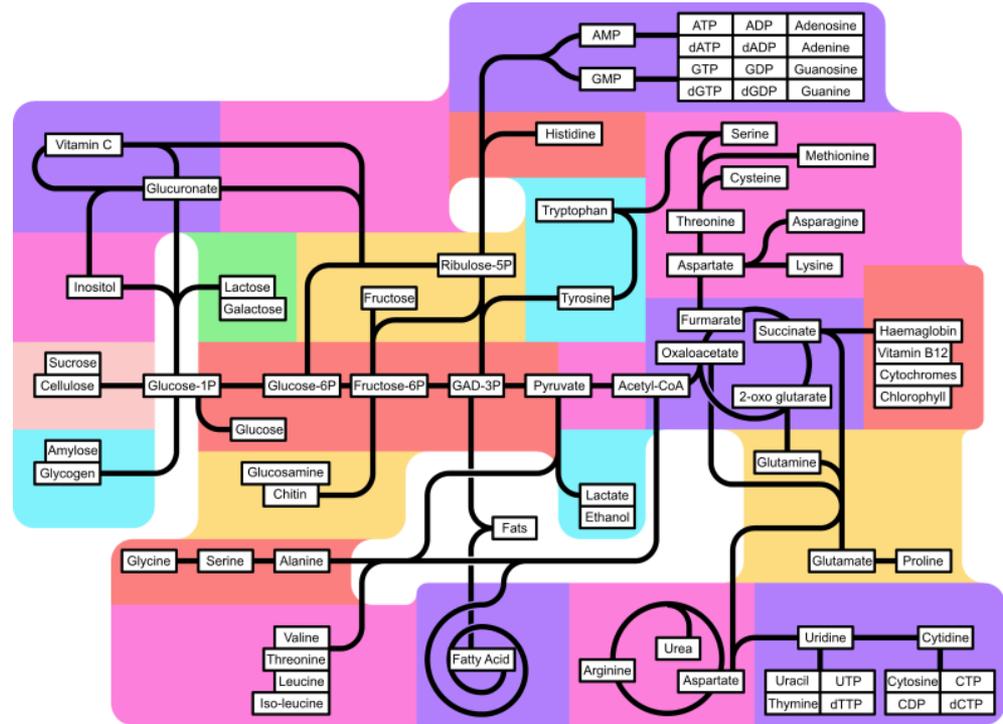
Voies métaboliques

- Traditionnellement, les biochimistes considèrent qu'une série de réactions métaboliques consécutives forme une **voie métaboliques**.

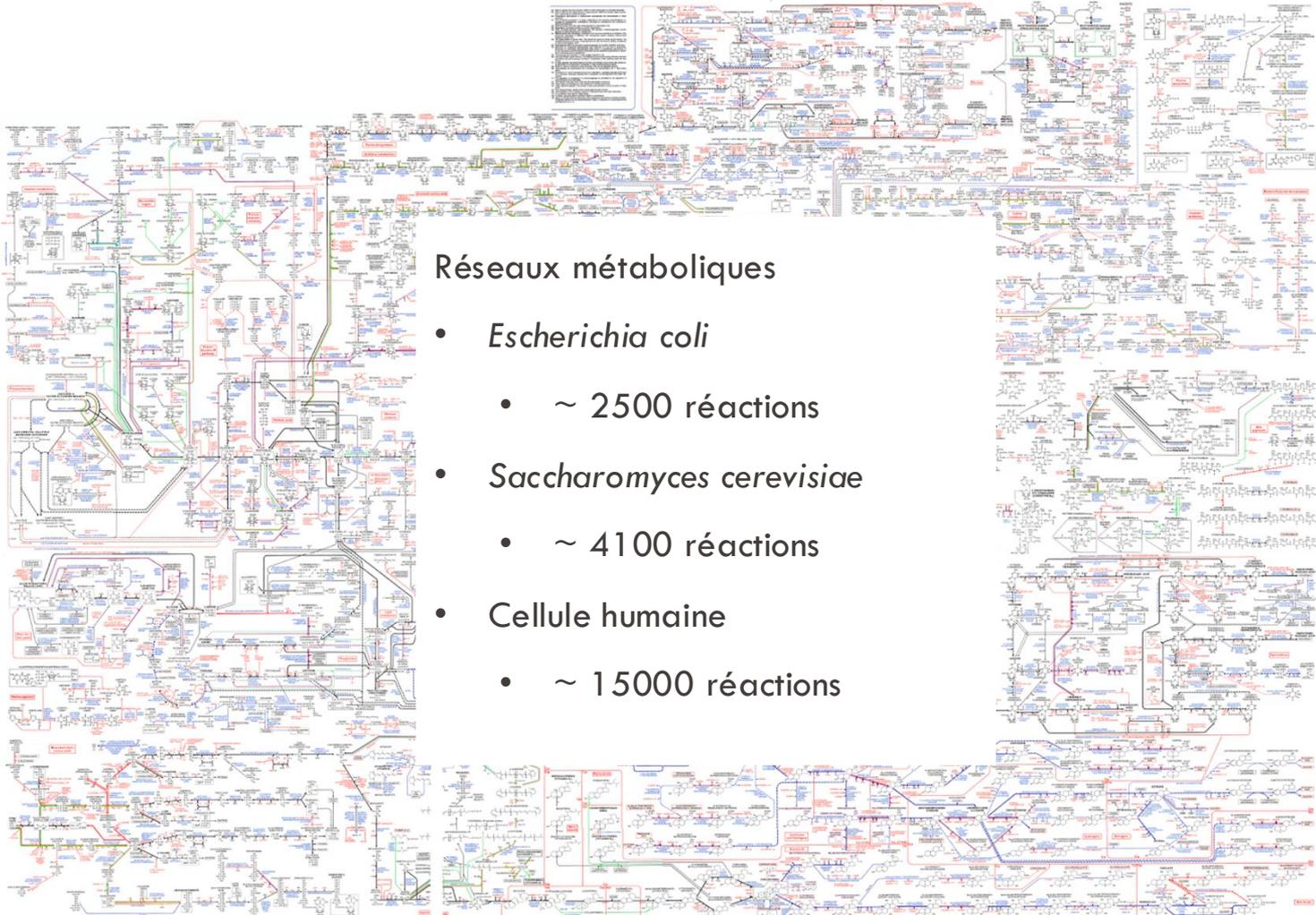


Réseaux métaboliques

- Les voies métaboliques se chevauchent tellement -> mieux vaut considérer un réseau métabolique complet.



Le réseau métabolique le plus complet



Pourquoi étudier les réseaux métaboliques ?

▪ *Biotechnologie*

- Une compréhension approfondie des réseaux métaboliques des bactéries est nécessaire si l'on veut les modifier génétiquement pour qu'elles produisent un produit souhaité avec des rendements maximaux.

▪ *Médecine*

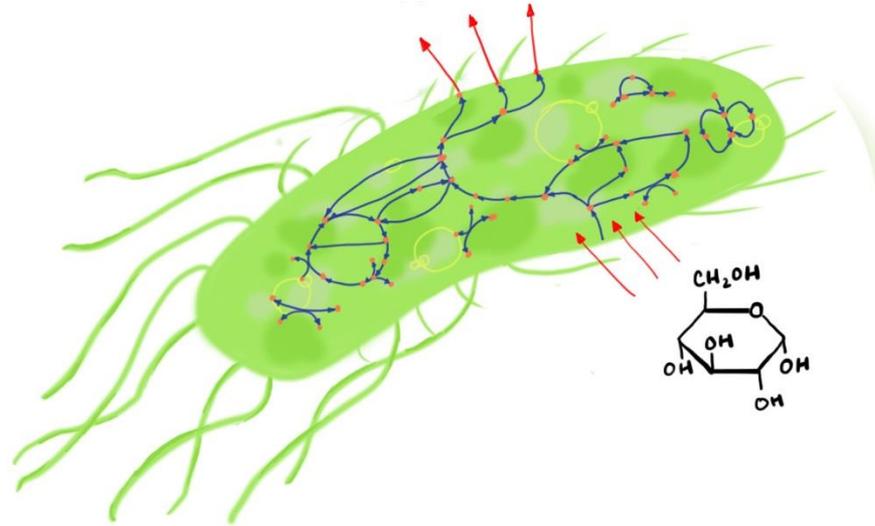
- Les aberrations du métabolisme humain sont à l'origine de maladies telles que le diabète et certains types de cancer.
- La connaissance des réseaux métaboliques des agents pathogènes et des parasites peut aider à sélectionner les cibles pharmacologiques (ou les combinaisons de cibles) qui seront les plus efficaces.

▪ *Fondamentale à la vie*

- Étant donné que les enzymes sont encodées dans le **génom**e, le métabolisme est l'un des mécanismes par lesquels le **génotype** d'un organisme (ensemble spécifique de gènes) est relié à son **phénotype** (comportement). De nombreux processus métaboliques sont communs à toutes les formes de vie.

Concevoir des usines cellulaires

À partir d'une souche microbienne



Découverte de médicaments



Bioproduction



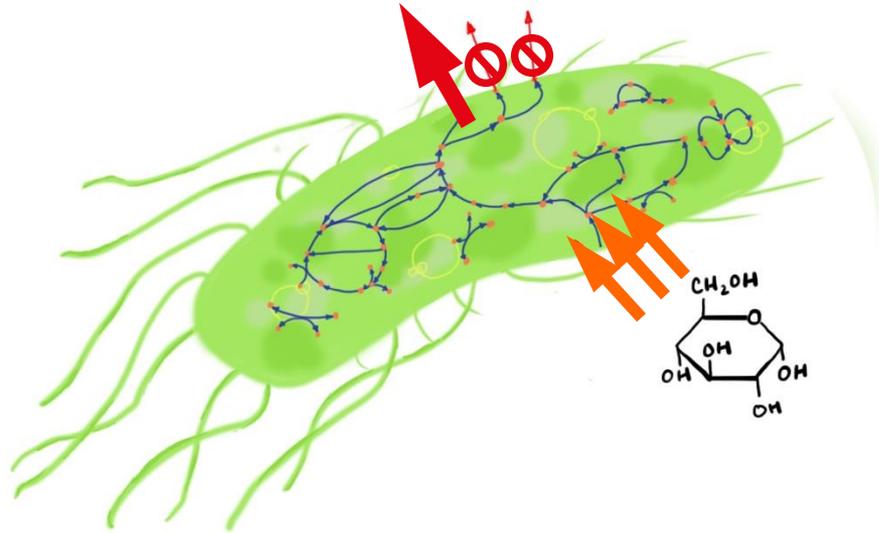
Biorémédiation
environnementale



Interactions
microbiennes

Concevoir des usines cellulaires

Comment créer une souche améliorée?



Découverte de médicaments



Bioproduction



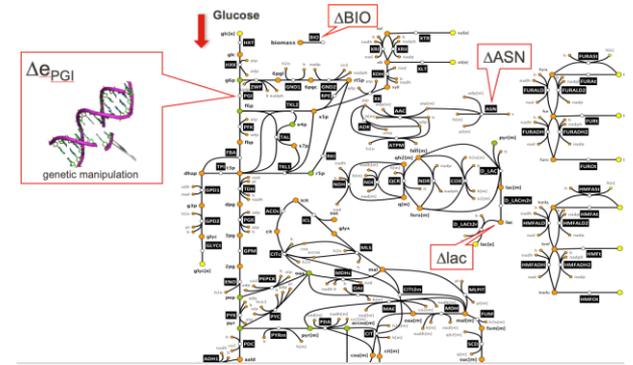
Biorémediation
environnementale



Interactions
microbiennes

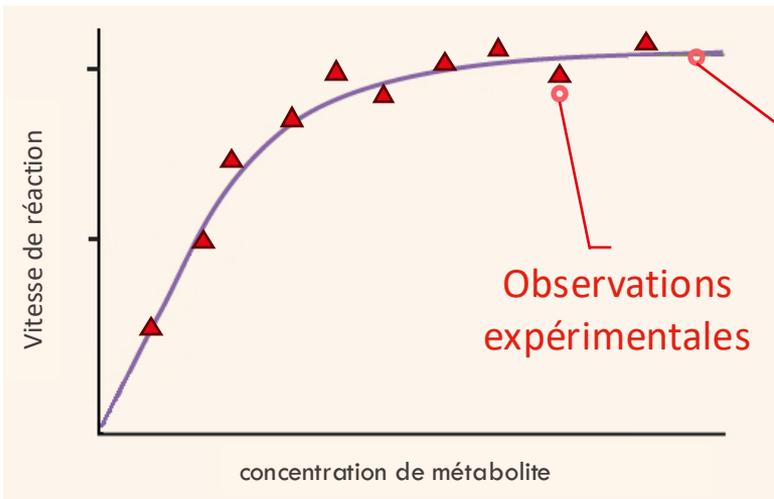
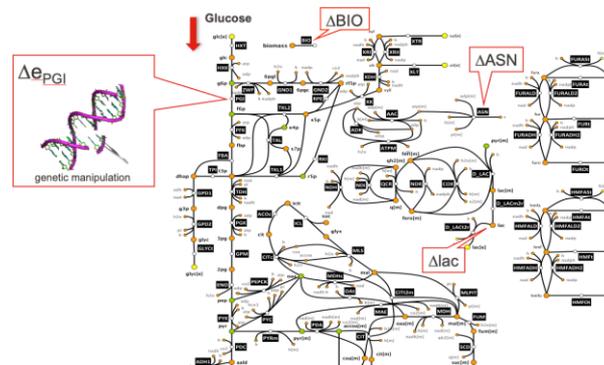
EPFL Modéliser les réactions pour identifier les gènes à cibler

Chaque réaction (bio)chimique doit être modélisée !



EPFL Modéliser les réactions pour identifier les gènes à cibler

Chaque réaction (bio)chimique doit être modélisée !



Modèle

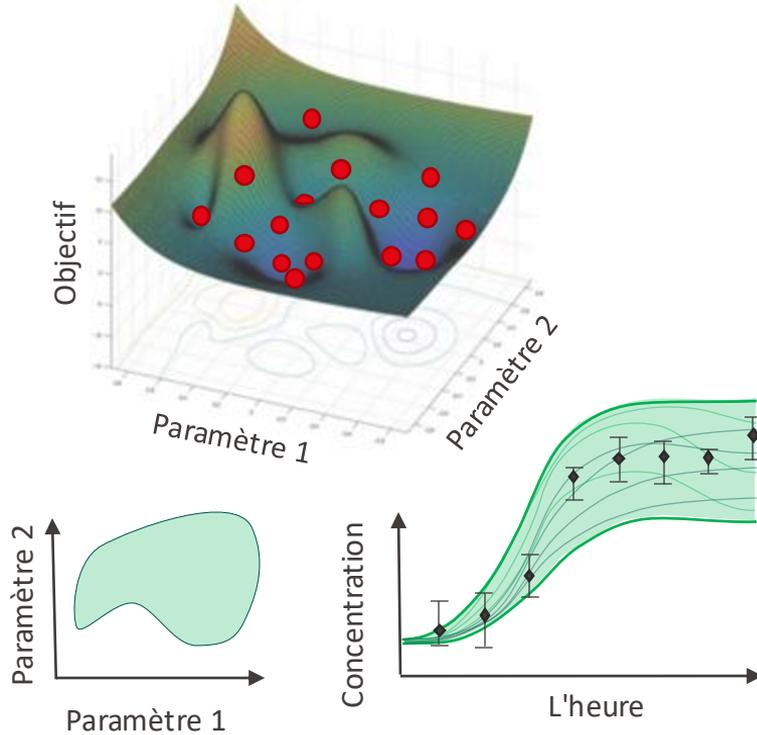
Vitesse de réaction

Concentrations
des métabolites

$$v_{PGI} = f(V_{max}^{PGI}, K_{M,PGI}^{g6p}, K_{M,PGI}^{f6p}, [g6p], [f6p])$$

Paramètres
cinétiques

Détermination des paramètres cinétiques



Estimation des paramètres

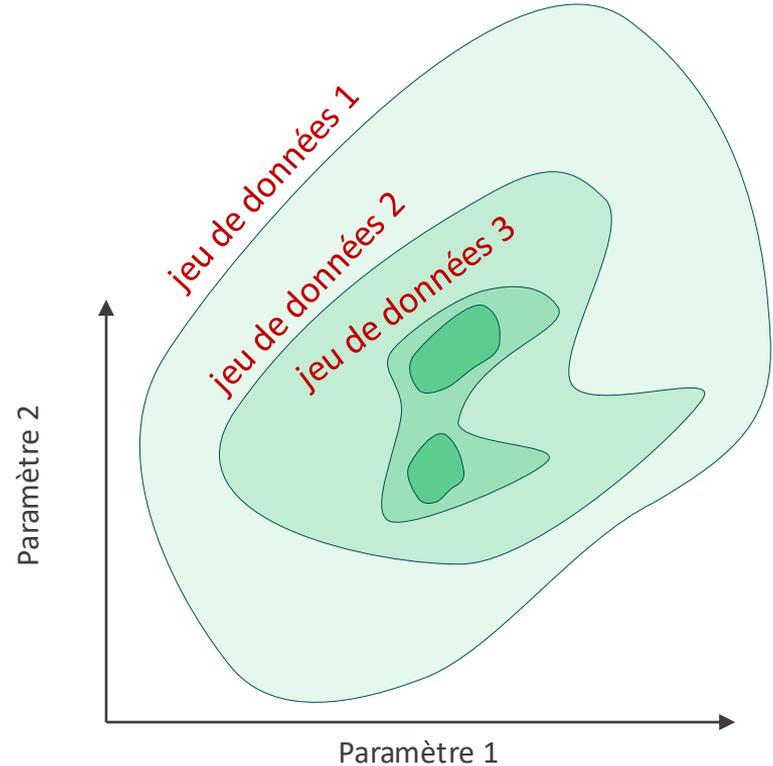
- Optimisation non linéaire dans un espace de haute dimension (minima locaux, sous-détermination)
- Les paramètres obtenus sont parfois dénués de sens

Échantillonnage de Monte Carlo

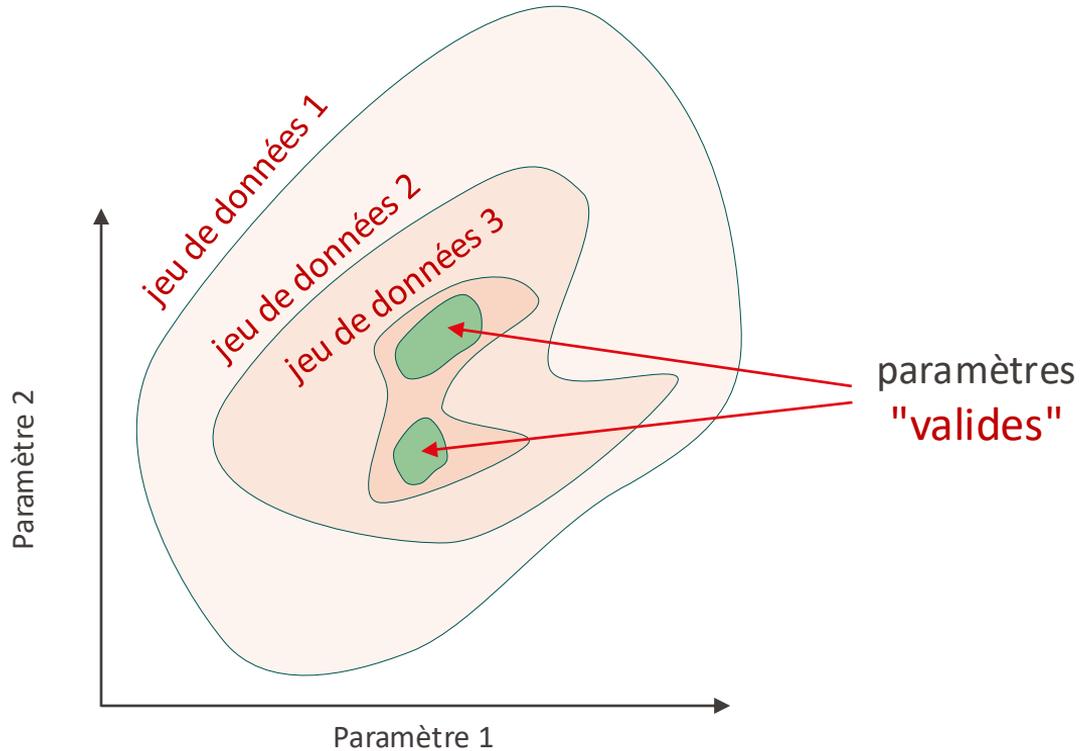
- Échantillonnage sans biais dans un espace bien contraint
- Nécessité d'une paramétrisation intelligente pour éviter des besoins de calcul élevés
- Favorable à l'apprentissage automatique (machine learning)

Exigences relatives aux paramètres cinétiques

Les paramètres cinétiques "valides" doivent être cohérents avec les multiples jeux de données expérimentales et les connaissances des experts



Peut-on générer uniquement des paramètres valides ?



Deux types de modèles de deep learning

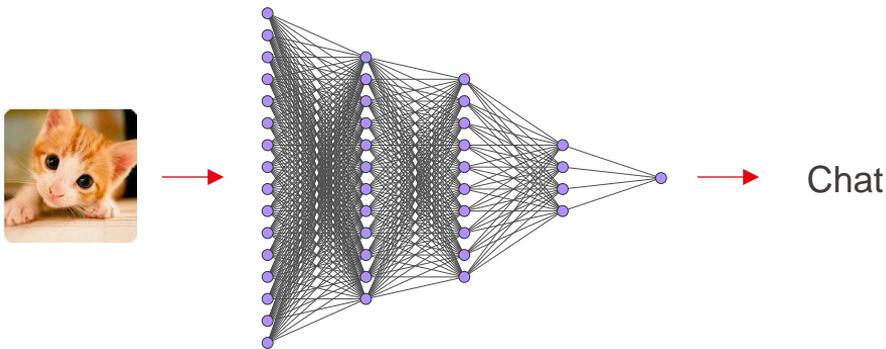
Modèles discriminants

Formé sur



Limite de décision

Classer



Deux types de modèles de deep learning

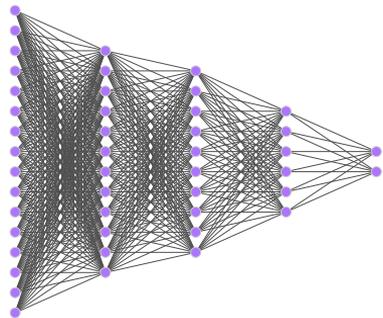
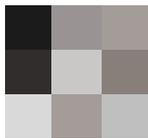
Modèles **génératifs**

Formé sur



Limite de décision

Générer une
image d'un
chien



Nouveau
contenu !

EPFL **Approches Generative Machine Learning**

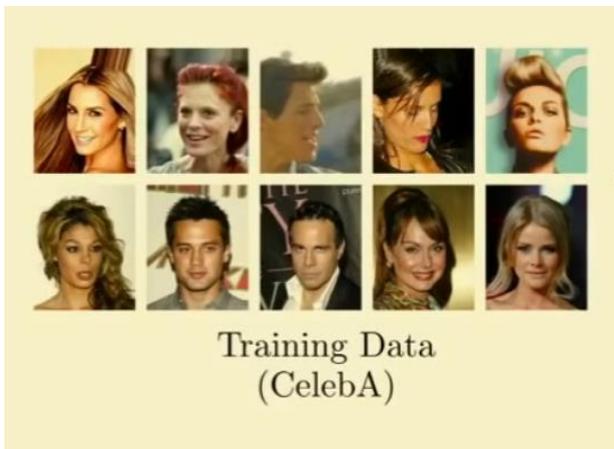
- **Modèles linguistiques génératifs (transformateurs) tels que ChatGPT (GPT - generative pre-trained transformers)**
- **Réseaux antagoniste génératifs (GAN)**
- **Autoencodeurs variationnels (VAE)**
- **Modèles de diffusion**
- **...**

Réseaux antagonistes génératifs

Generative Adversarial Networks (GAN) : Ils se composent de deux réseaux : un générateur qui crée de nouveaux échantillons de données et un discriminateur qui fait la distinction entre les échantillons réels et les échantillons générés. Les deux réseaux sont formés ensemble dans le cadre d'un processus qui permet d'améliorer la qualité des données générées.

Nombreuses applications, notamment dans la création d'images :

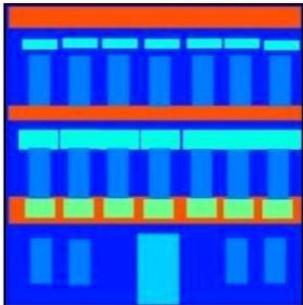
- **Créer de nouveaux visages**



Réseaux antagonistes génératifs

- Traduction d'image à image

Labels to Facade



input



output

BW to Color



input



output



zebra → horse



Day to Night



input



output

Edges to Photo



input



output

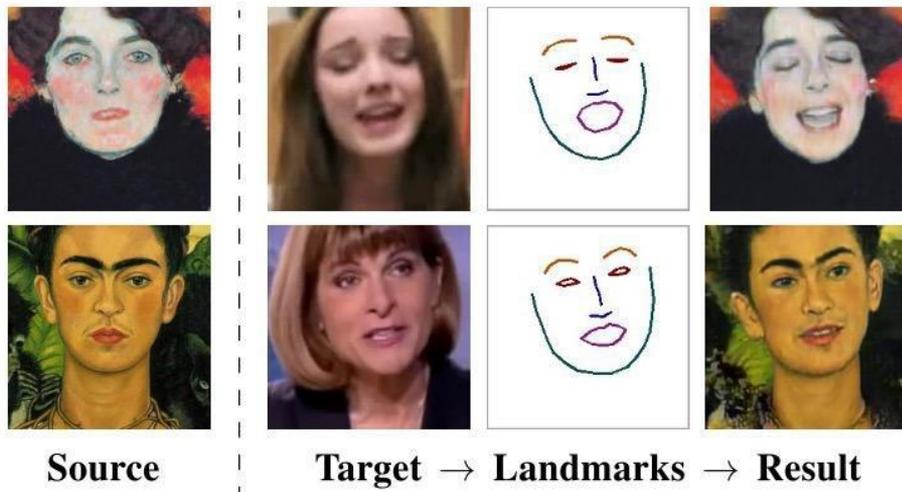


horse → zebra



Réseaux antagonistes génératifs

- Méta-apprentissage



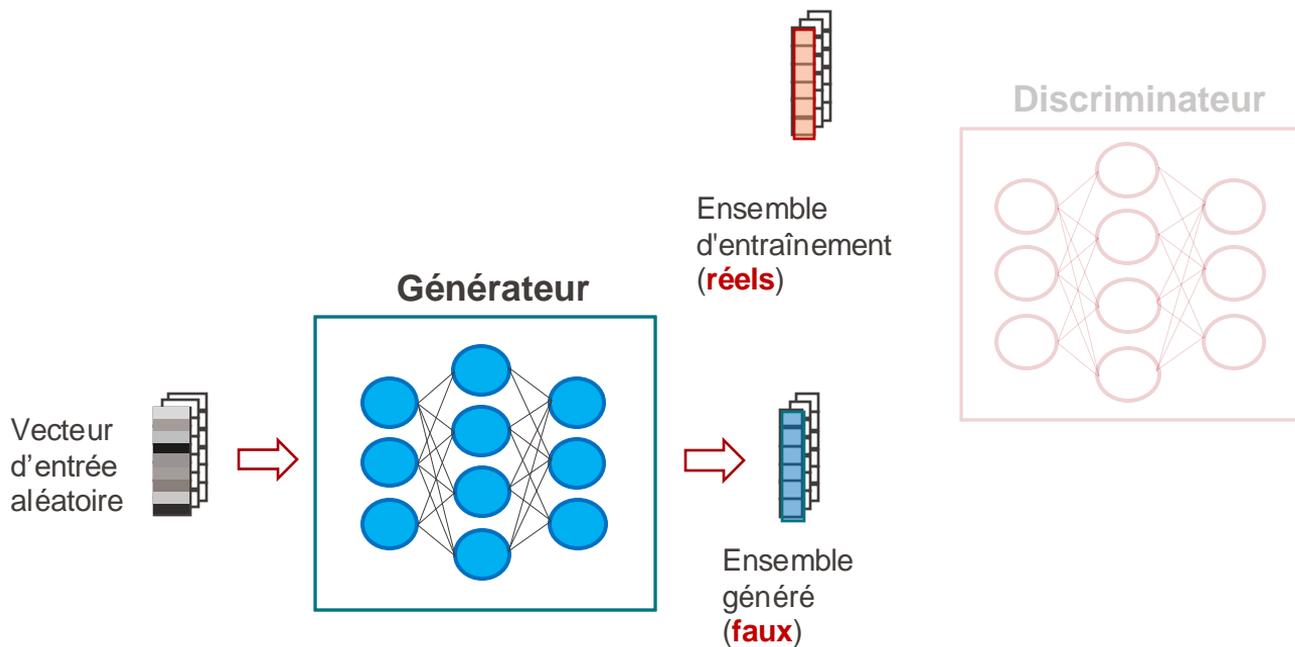
Les résultats sont conditionnés par les repères pris dans l'image cible, tandis que l'image source est un exemple tiré de l'ensemble d'apprentissage.

Réseaux antagonistes génératifs

- Deepfake

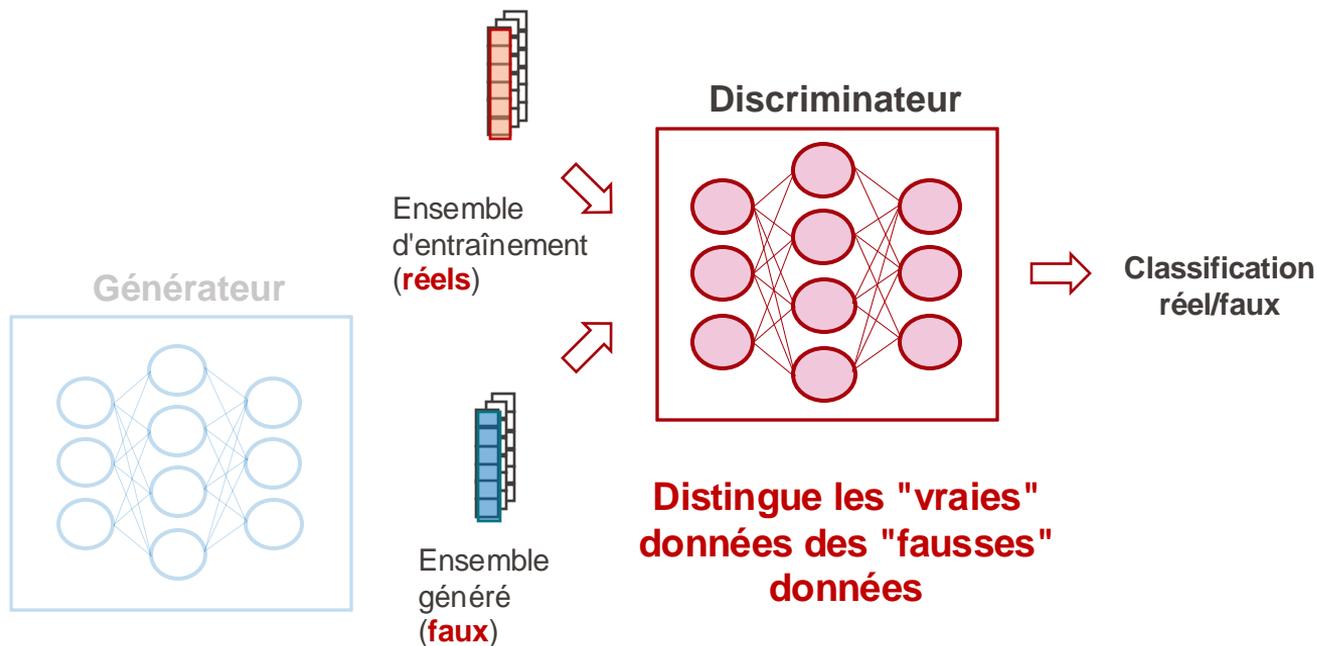


Generative adversarial networks (GAN) en quelques mots

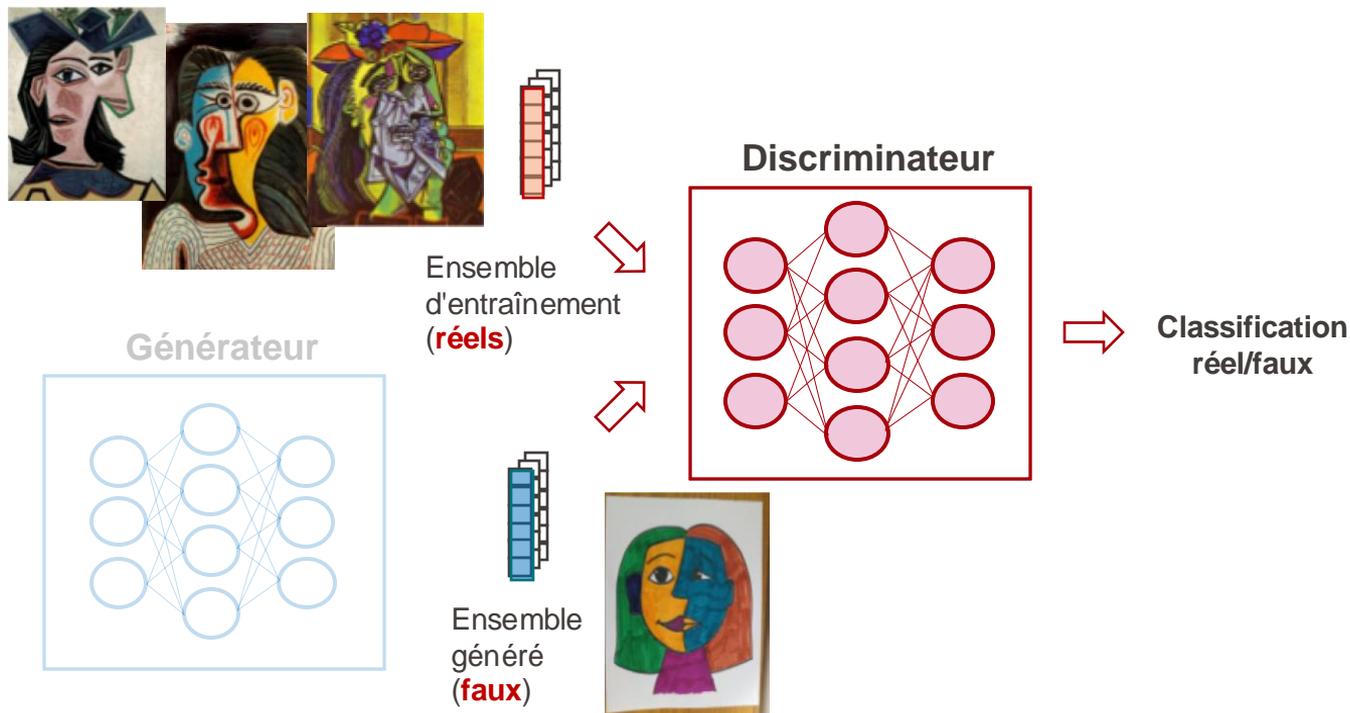


Génère des données à partir d'un vecteur aléatoire

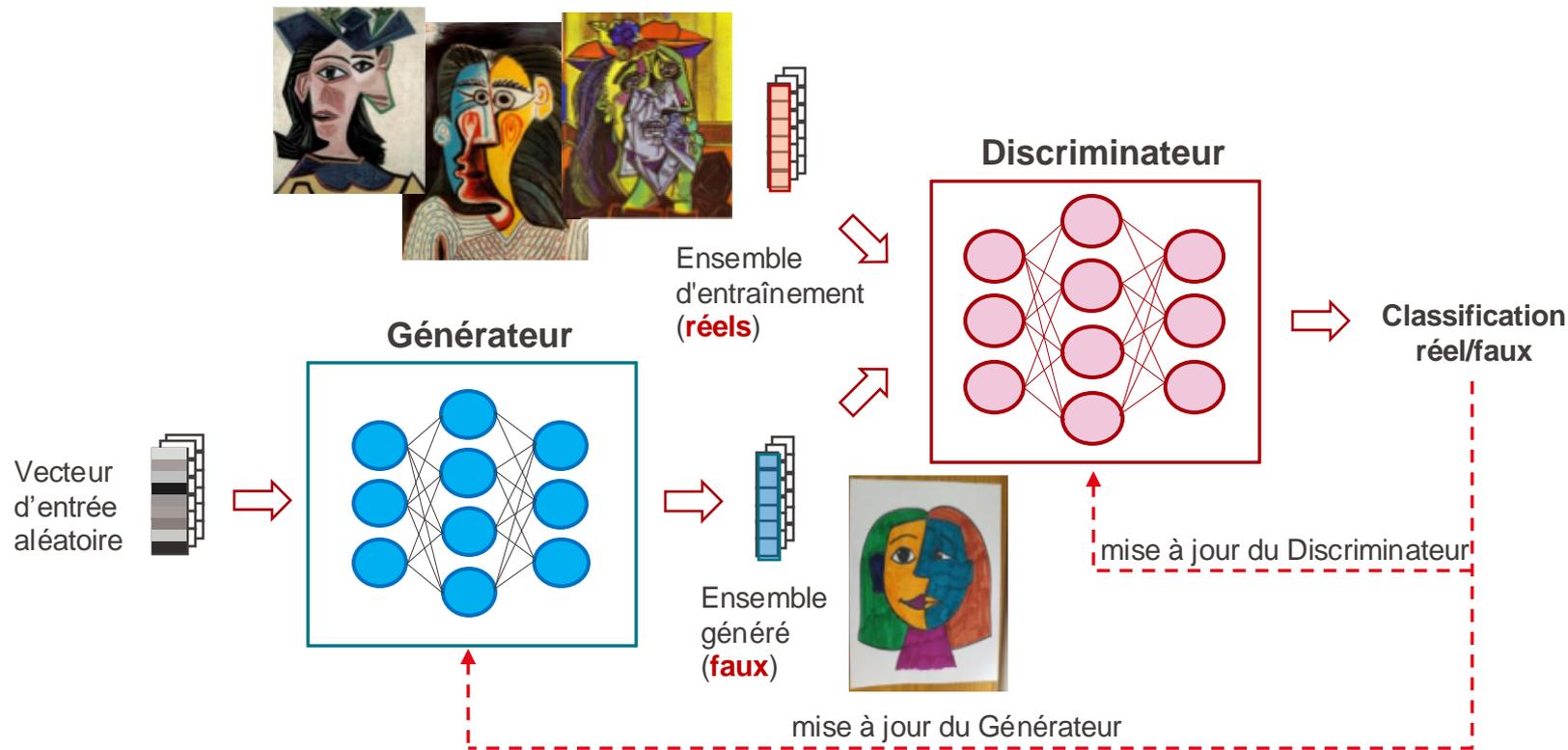
Generative adversarial networks (GAN) en quelques mots



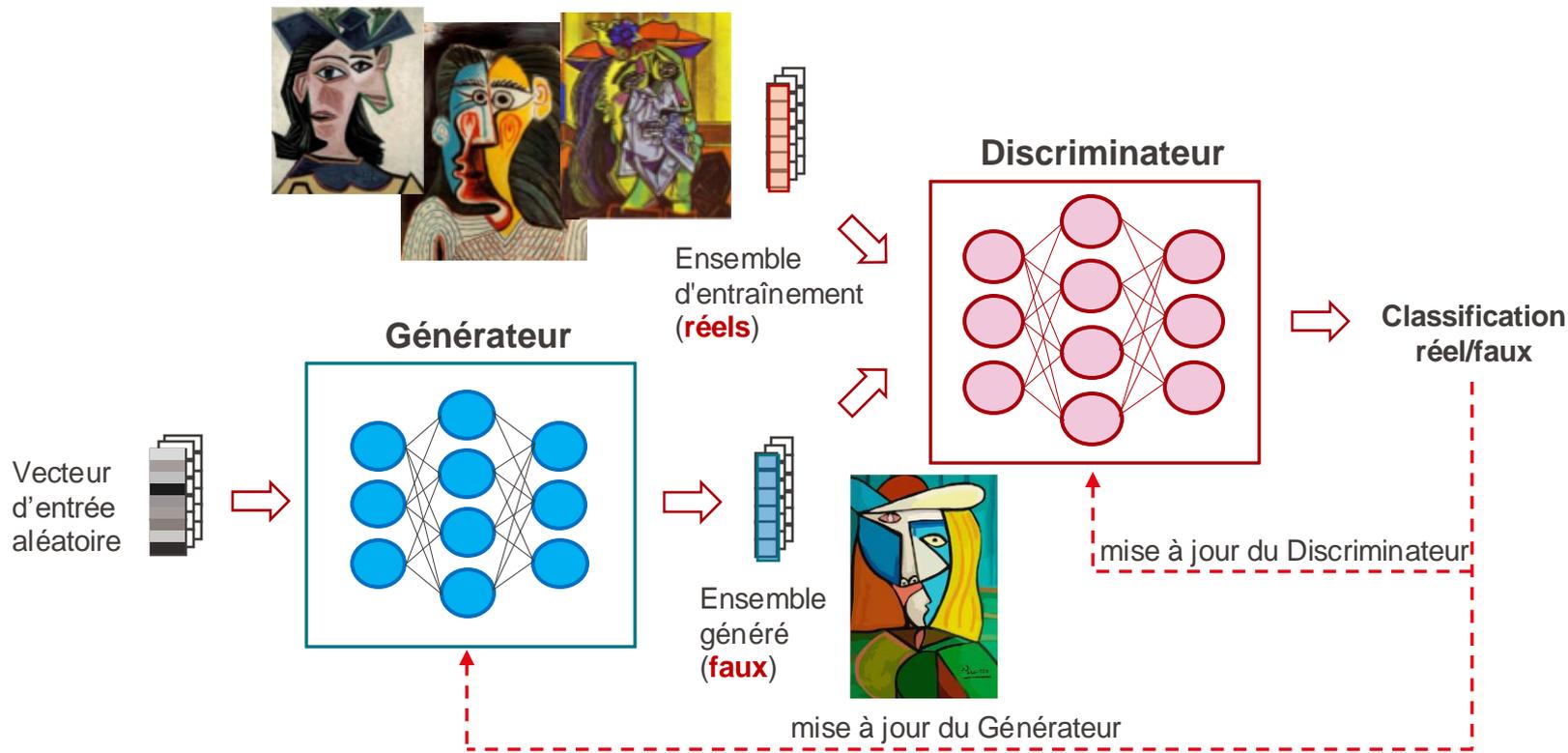
Generative adversarial networks (GAN) en quelques mots



Generative adversarial networks (GAN) en quelques mots



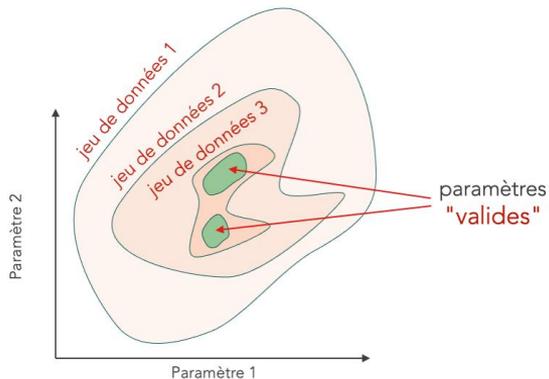
Generative adversarial networks (GAN) en quelques mots



À l'équilibre, les échantillons du générateur ne peuvent être distingués des données réelles

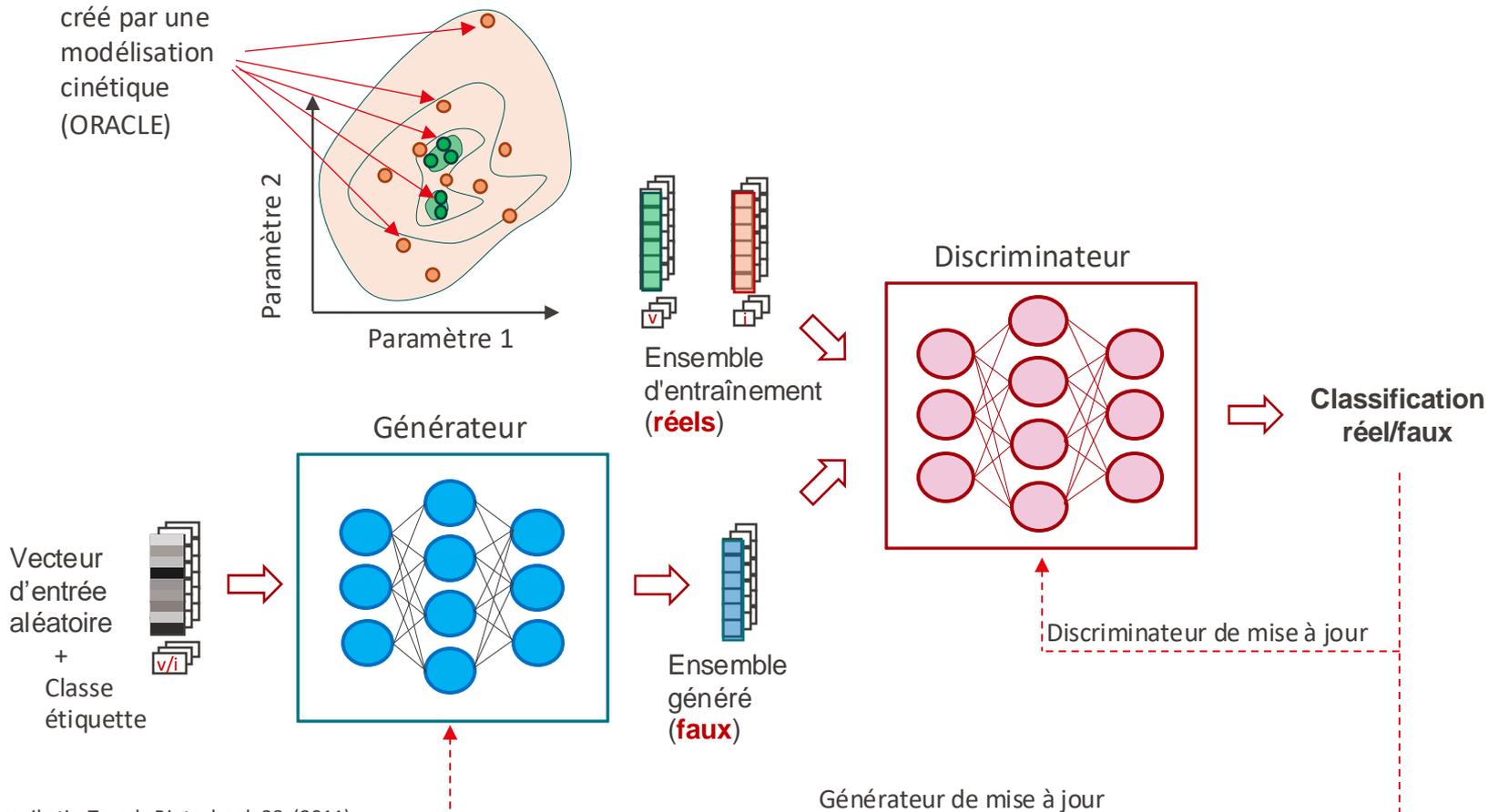


Utiliser des réseaux neuronaux **génératifs** pour reconstruire des modèles avec des paramètres cinétiques **valides**



Reconstruction of Kinetic Models using Deep Learning

REKINDLE utilise des GAN conditionnels



ORACLE :

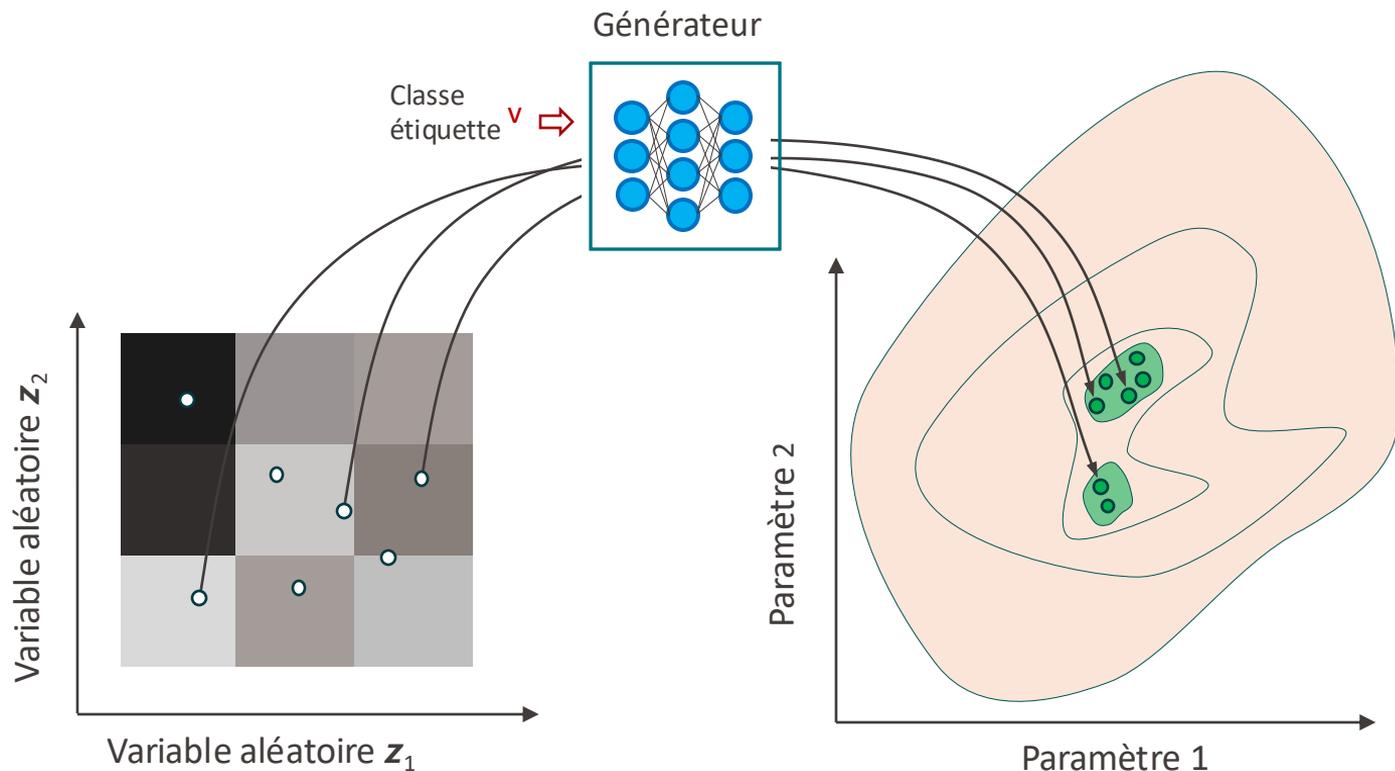
Miskovic & Hatzimanikatis, Trends Biotechnol, 28, (2011)

Tokic et al, Biotechnol Biofuels, 13:33, (2020)

Mirza & Osindero, Archiv.org, (2014)

Choudhury et al, Nat Mach Intell, 4, 710-719, (2022)

Un bon générateur reconstruit des modèles valides

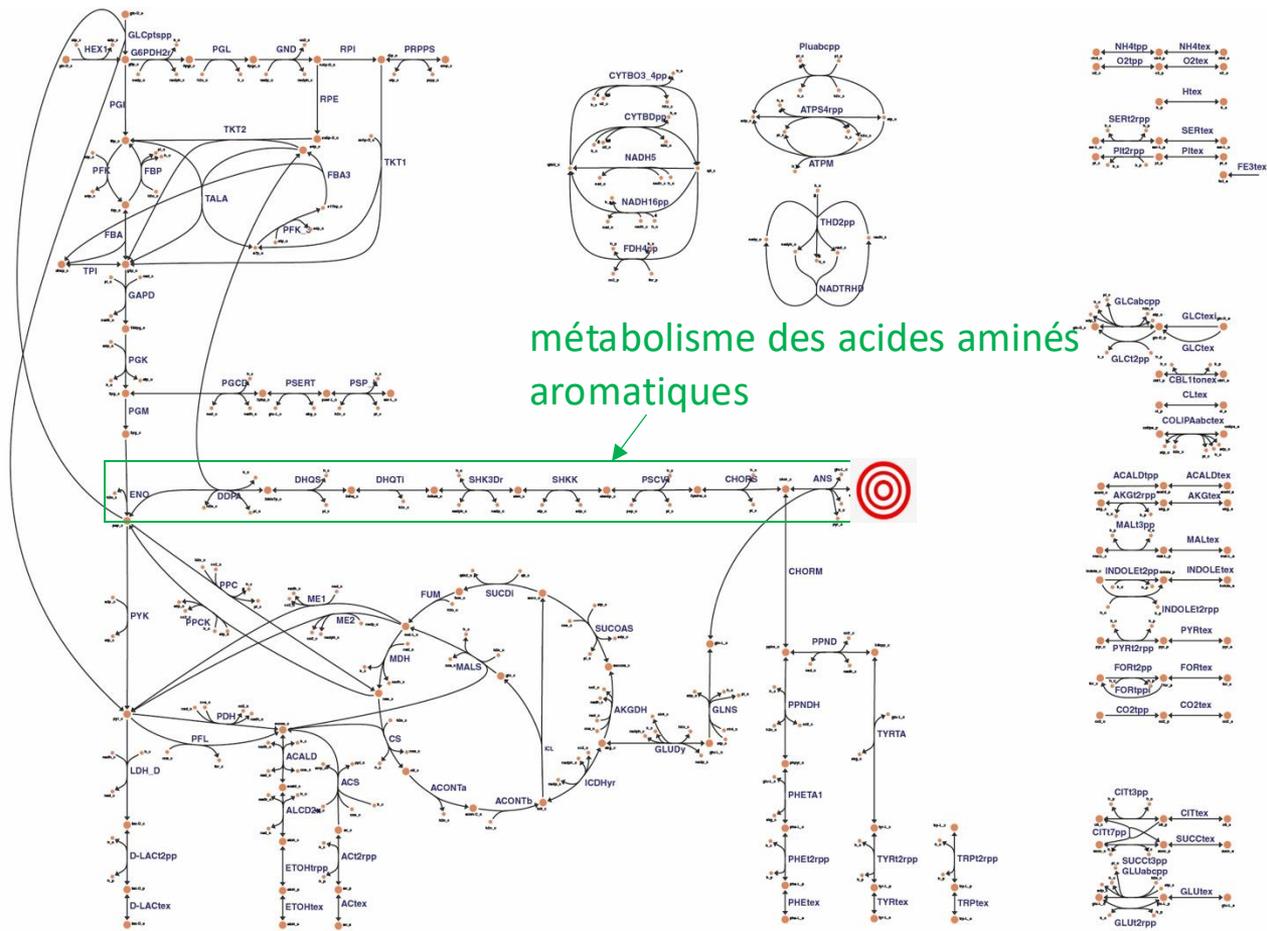


Étude de cas

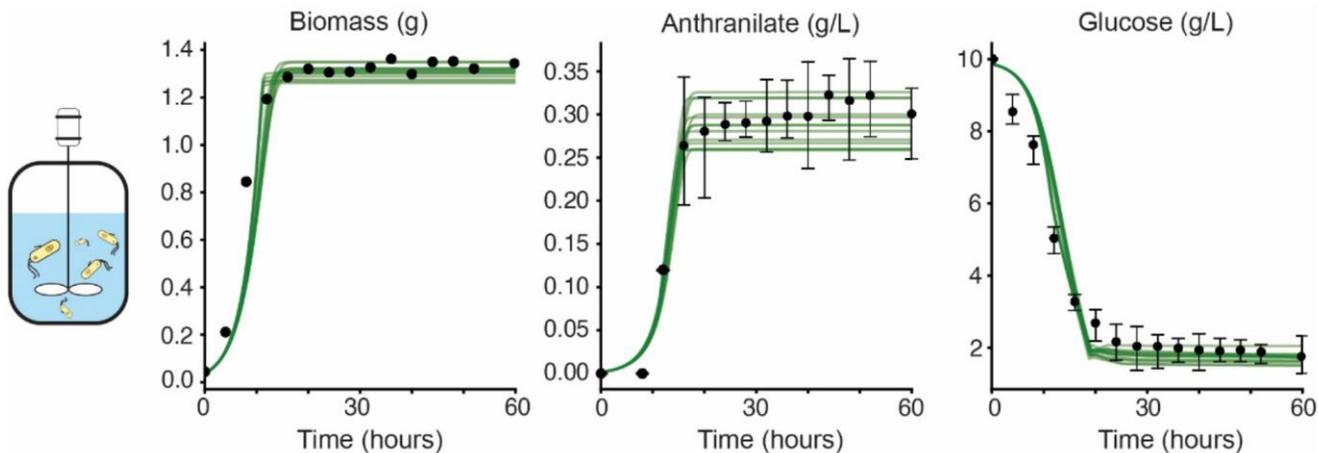
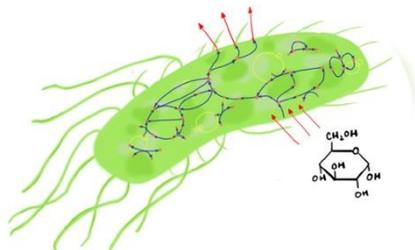
Modèles cinétiques du
métabolisme central du
carbone + acides aminés
aromatiques de *E.coli*

89 bilans de masse (ODE)
122 réactions

506 paramètres à estimer



Simulations de bioréacteur pour notre souche bactérienne de départ



Génération efficace de modèles valides

1. Temps de entraînement

10 modèles:	2-3 minutes
1000 modèles:	10 minutes
72000 modèles:	45-60 minutes

2 . Temps de génération (1 million de modèles valides)

REKINDLE (avec GPU) : 15-20 s

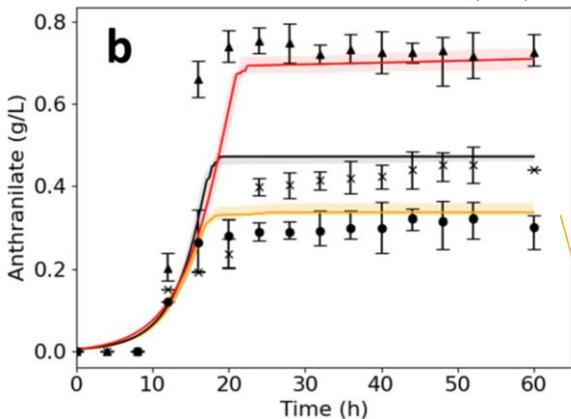
REKINDLE (sans GPU) : 1 minute

en comparaison

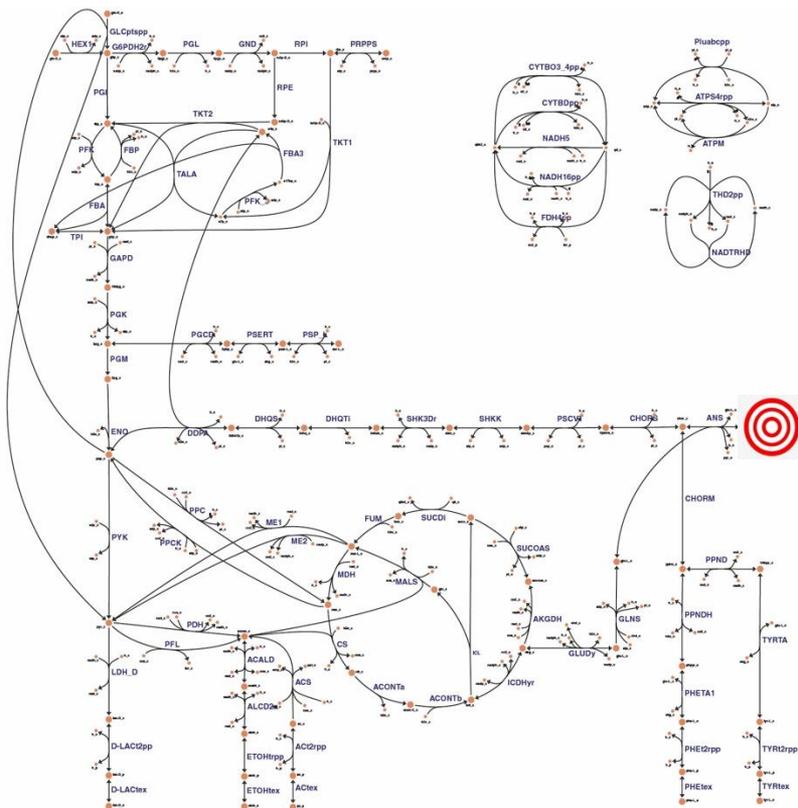
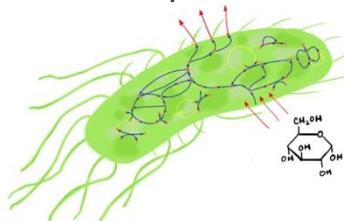
ORACLE : 36-48 heures

EPFL Améliorer la production de produits chimiques souhaités

Balderas-Hernández et al. *Microb Cell Fact*, (2009)



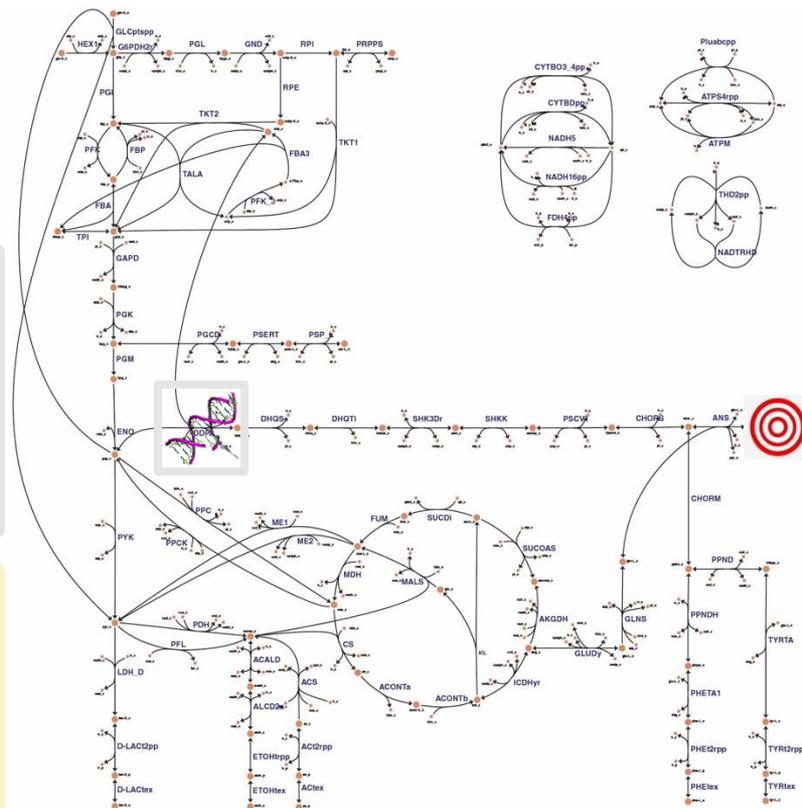
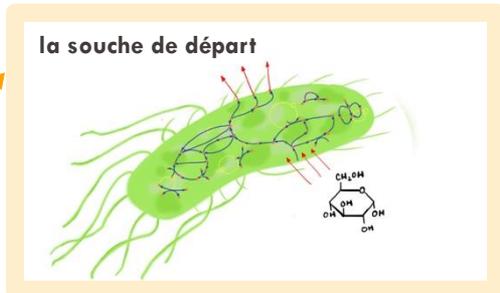
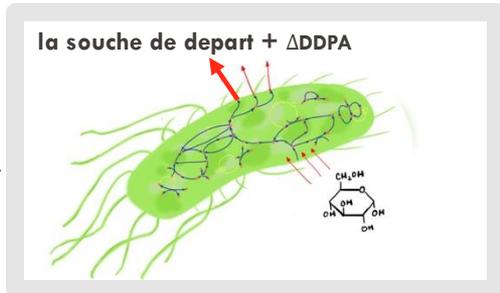
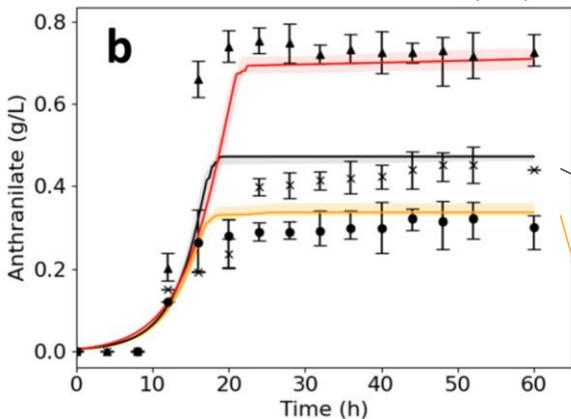
la souche de départ



Narayanan et al. *Nature Communications*, 15:723, (2024)

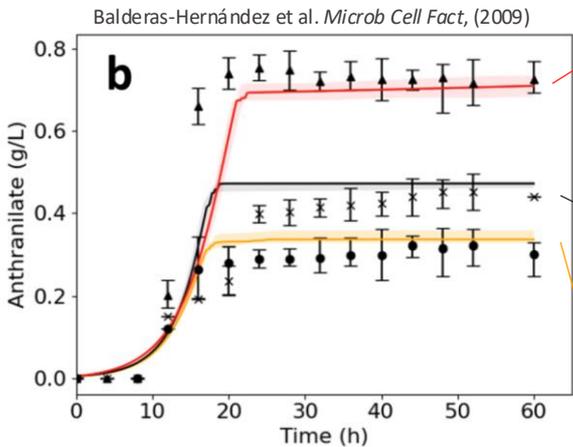
EPFL Améliorer la production de produits chimiques souhaités

Balderas-Hernández et al. *Microb Cell Fact*, (2009)

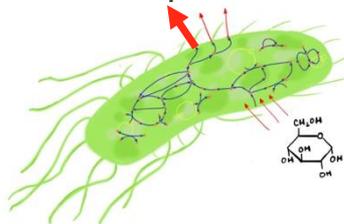


Narayanan et al. *Nature Communications*, 15:723, (2024)

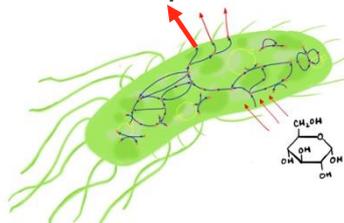
EPFL Améliorer la production de produits chimiques souhaités



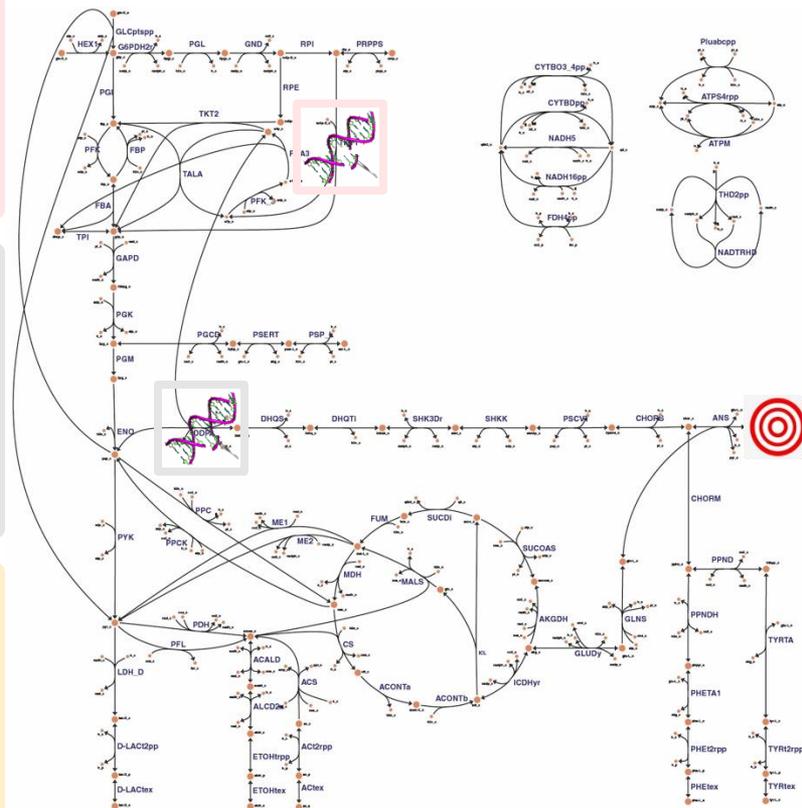
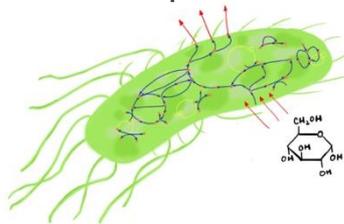
la souche de départ + $\Delta DDPA + \Delta TKT1$



la souche de départ + $\Delta DDPA$

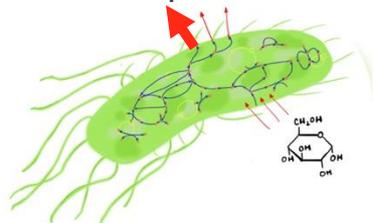


la souche de départ

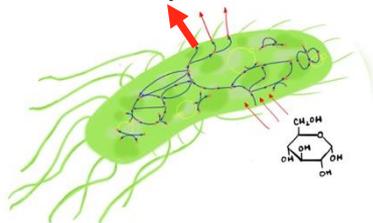


EPFL Améliorer la production de produits chimiques souhaités

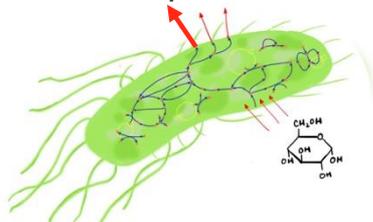
la souche de départ + Δ DDPA + Δ PGK



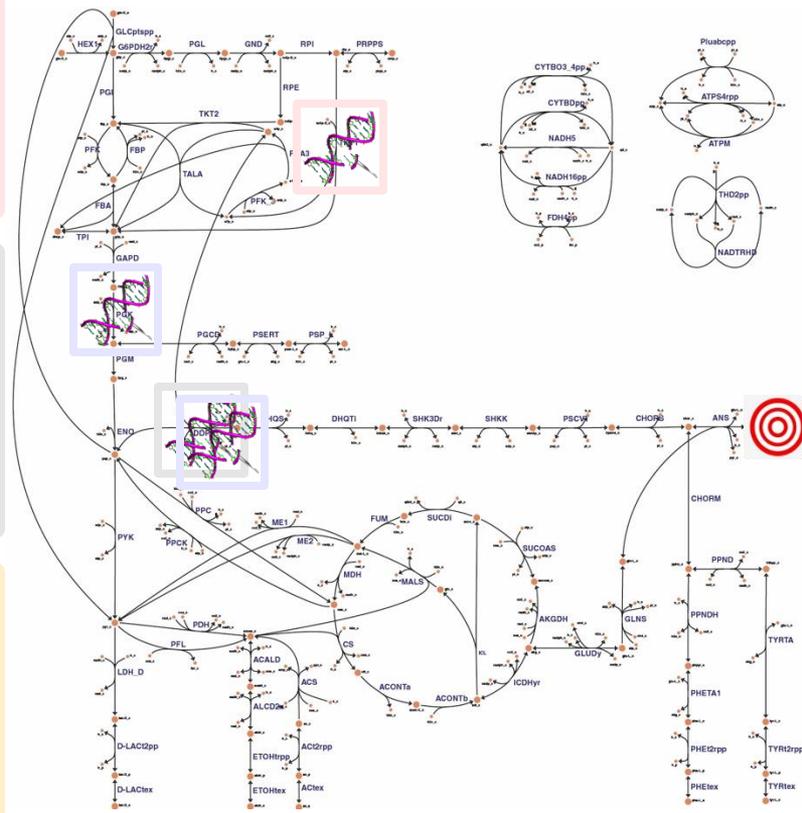
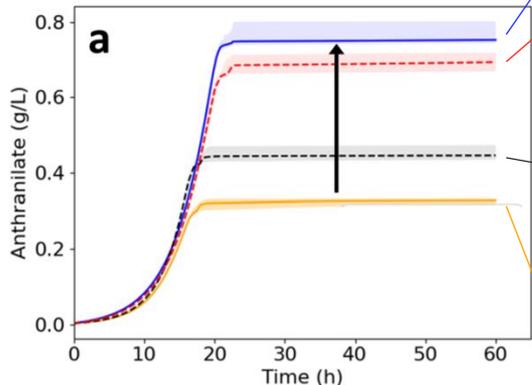
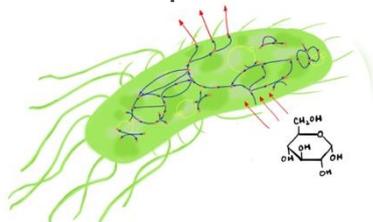
la souche de départ + Δ DDPA + Δ TKT1



la souche de départ + Δ DDPA



la souche de départ



Conclusions

L'apprentissage automatique offre de nombreuses améliorations pour l'étude de la dynamique du métabolisme

- **Conceptions et études à haut débit**
 - Reconstruction accélérée de modèles
 - Apprentissage par transfert
- **Facilité d'utilisation**
- **Augmentation des données et analyse statistique**
- **Interprétabilité**

Health



Biotechnology

