



Annotation fonctionnelle du protéome d'*Arabidopsis thaliana* via l'analyse et la prédiction de son interactome

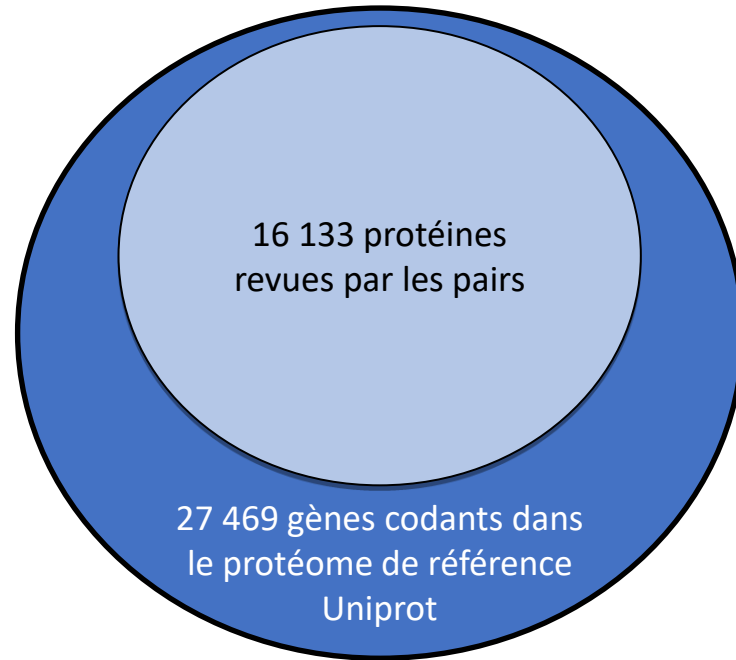
Comité de thèse - année 1

Simon Gosset – Equipe GNET

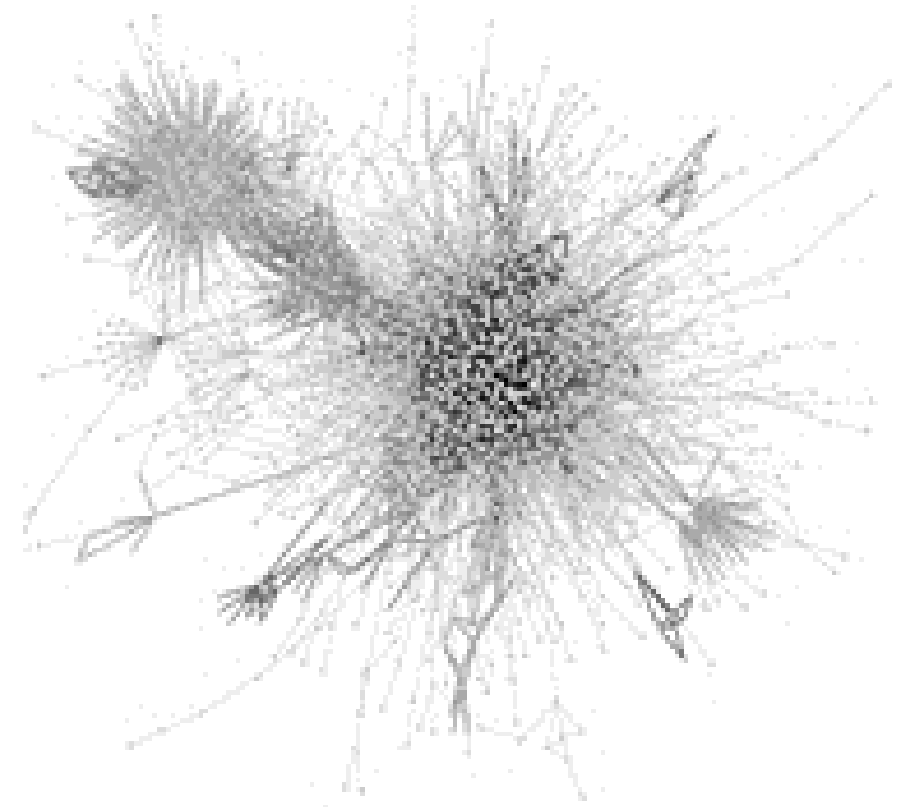
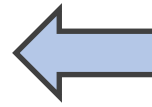
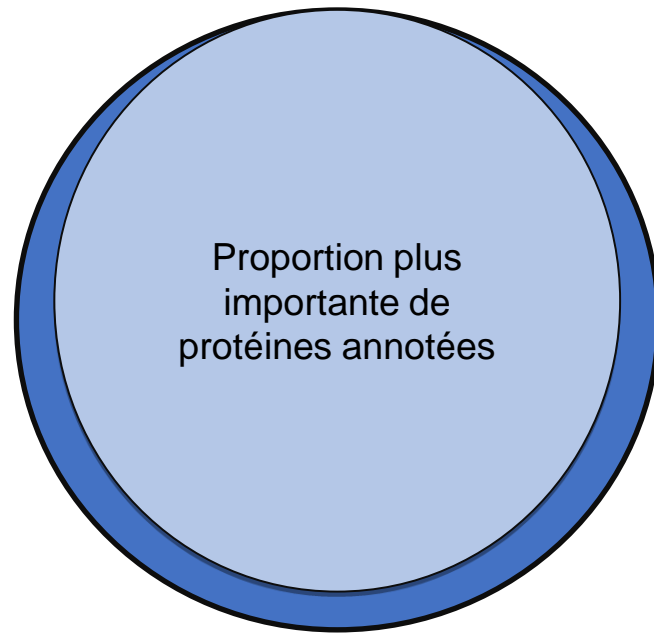
Sous la direction de Marie-Hélène Mucchielli Giorgi



Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome



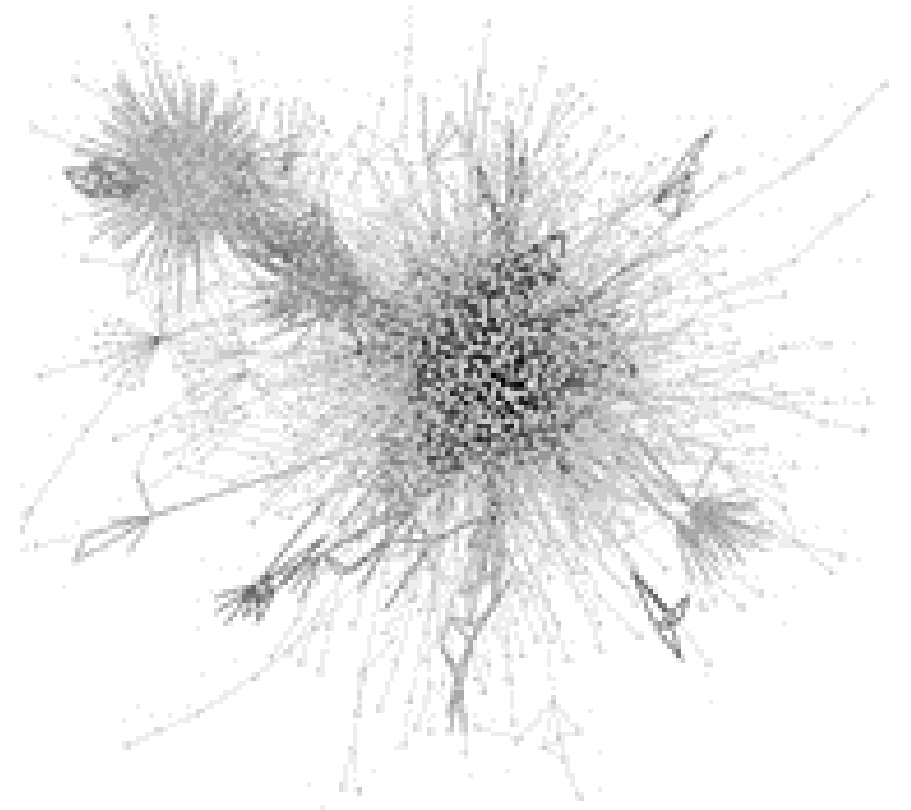
Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome



Interactome d'A. thaliana

Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome

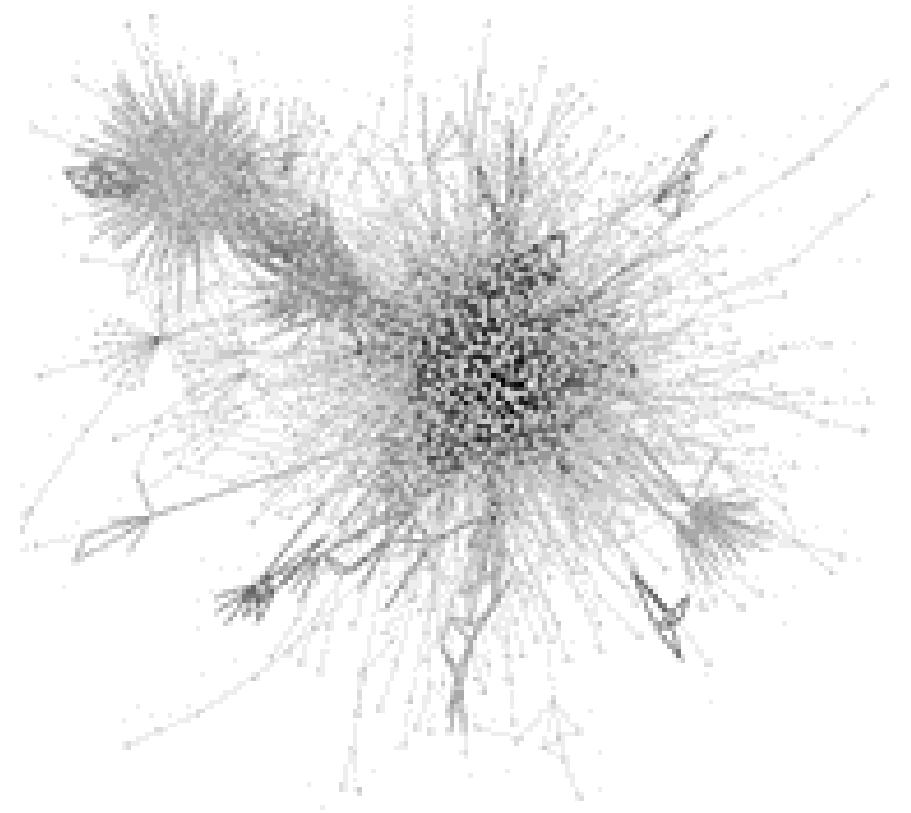
- Un réseau en apparence illisible



Interactome d'A. thaliana

Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome

- Un réseau en apparence illisible
- Ce réseau n'est pas construit de façon aléatoire



Interactome d'A. thaliana

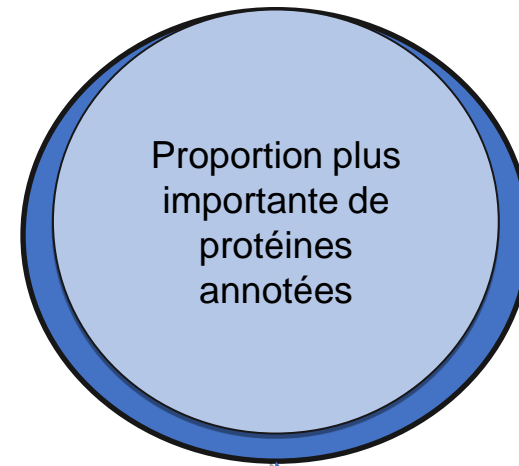
Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome

- Un réseau en apparence illisible
- Ce réseau n'est pas construit de façon aléatoire
- **Les protéines de même fonction regroupées et plus connectées entre elles qu'avec le reste du réseau**



Objectif de la thèse : Annoter le protéome d'*A. thaliana* en explorant son interactome

- Un réseau en apparence illisible
- Ce réseau n'est pas construit de façon aléatoire
- Les protéines de même fonction regroupées et plus connectées entre elles qu'avec le reste du réseau
- **Annotation fonctionnelle des protéines = recherche de sous-réseaux**

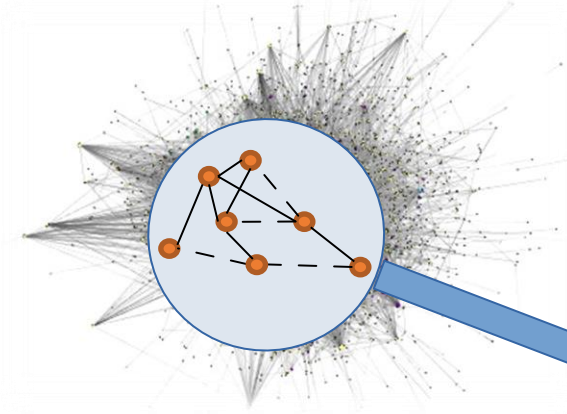
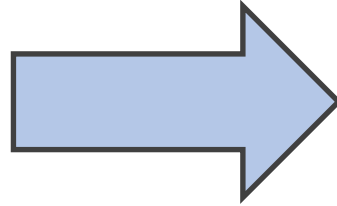


*Interactome d'*A. thaliana**
Les couleurs correspondent aux différentes conditions de stress

Un réseau incomplet

137 690 PPIs pour 27 469 gènes
codants chez *A. thaliana*

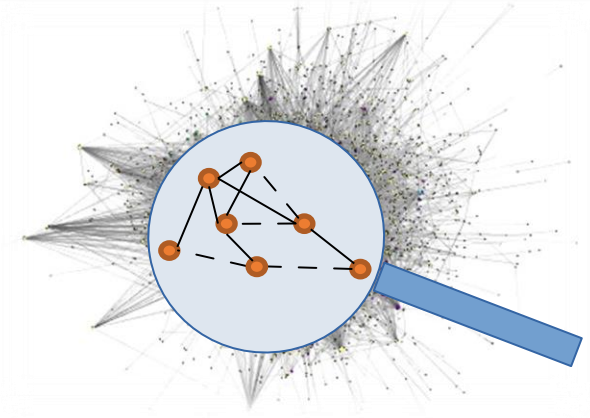
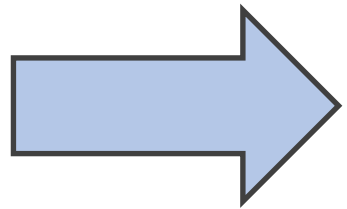
-> Beaucoup de **PPI manquantes**
dans l'interactome d' *A. thaliana*



**Ajouter des
PPIs**

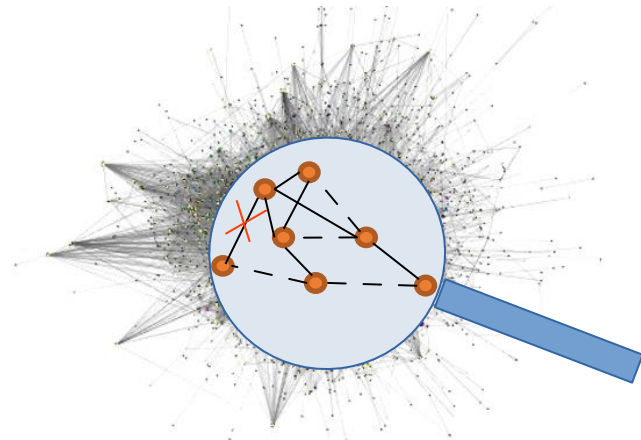
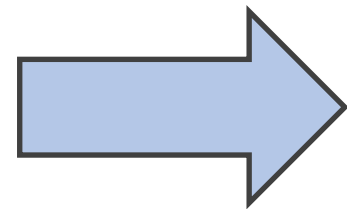
Un réseau incomplet et partiellement faux

137 690 PPIs pour 27 469 gènes codants chez *A. thaliana*
-> Beaucoup de **PPI manquantes** dans l'interactome d' *A. thaliana*



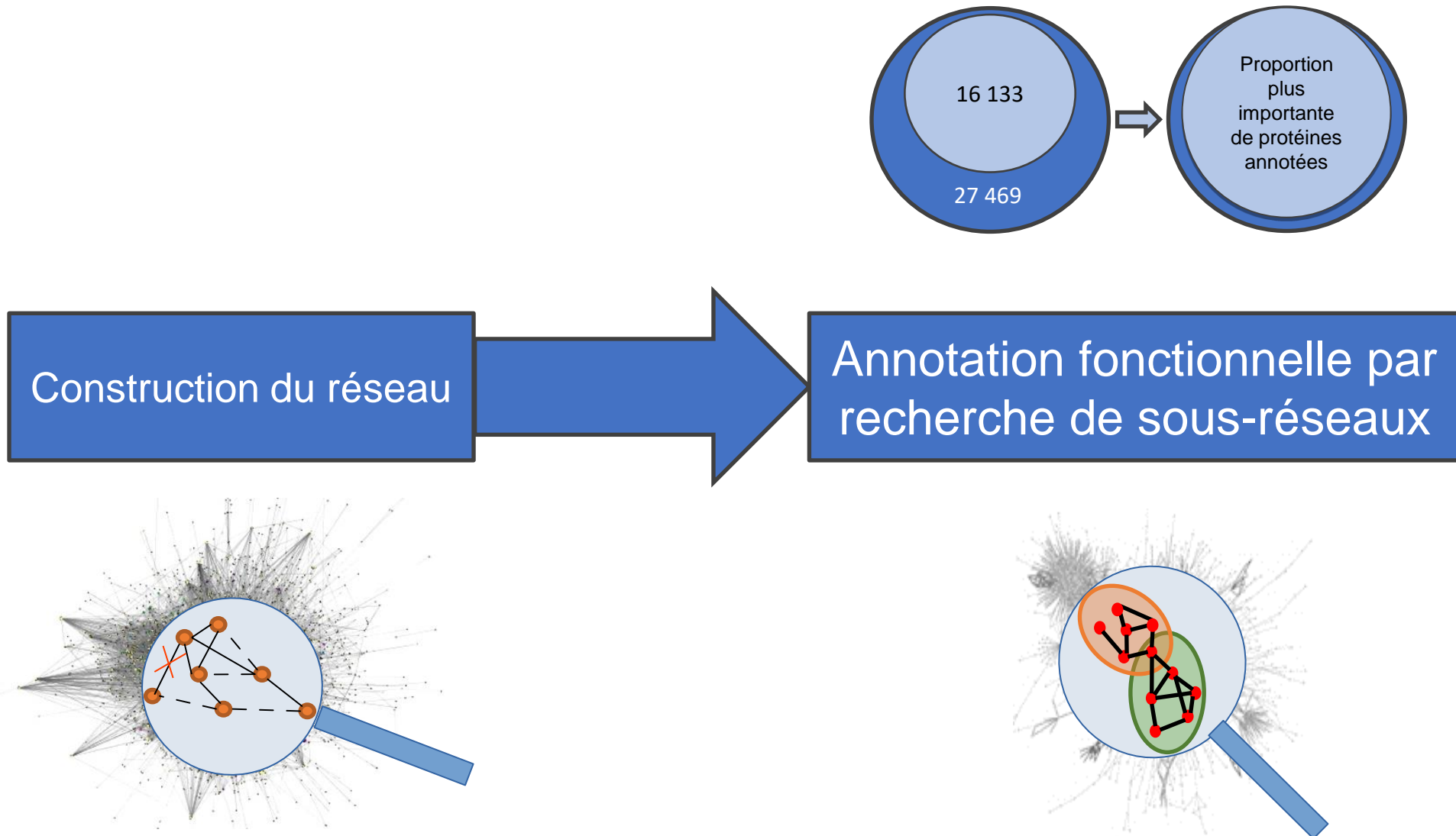
Ajouter des PPIs

Méthode haut débit : Beaucoup de faux positifs

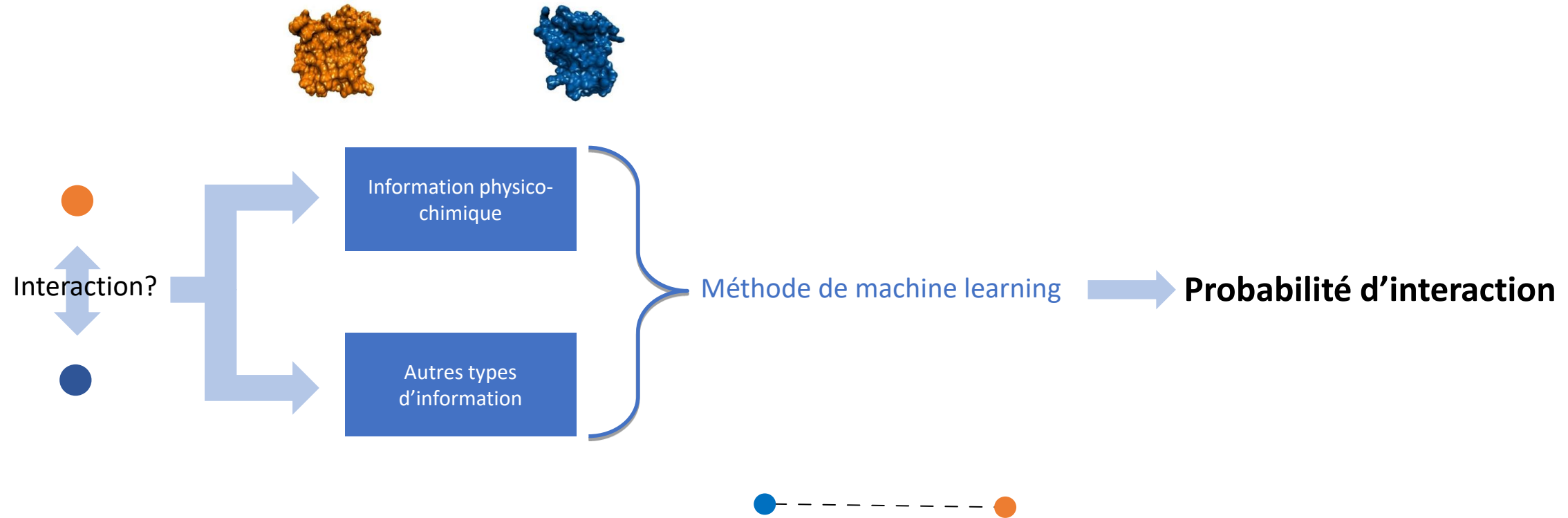


Supprimer les Faux positifs

Une thèse en 2 parties

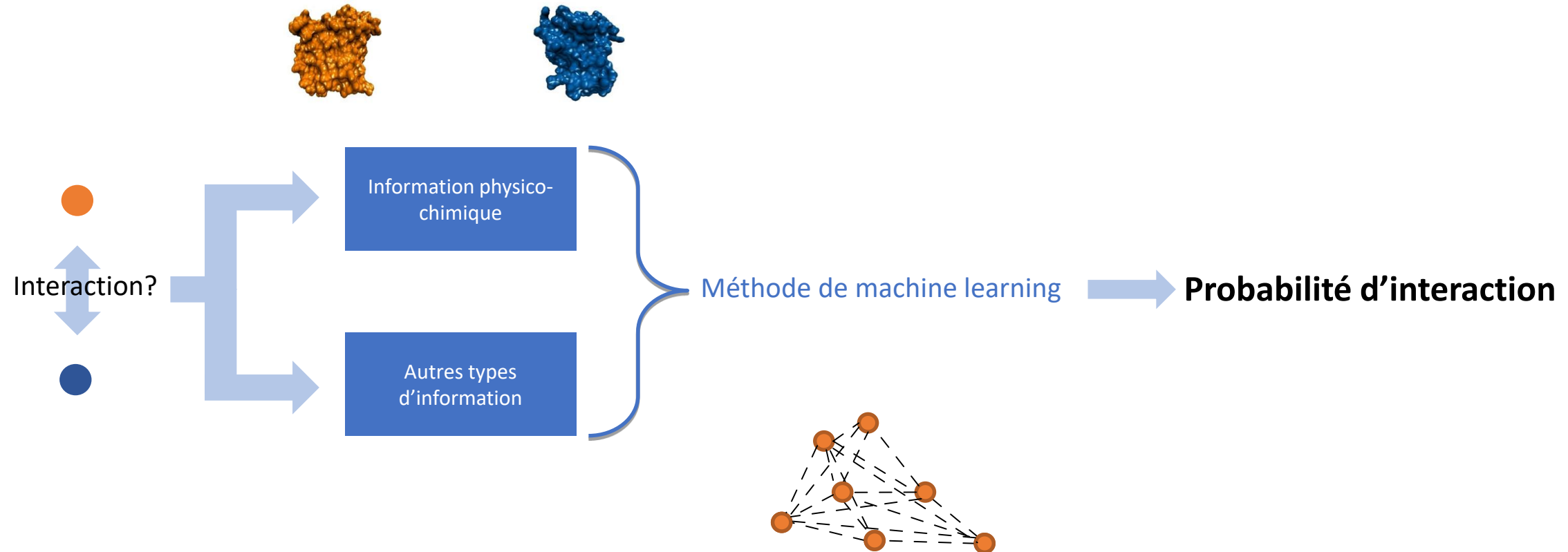


Des probabilités d'interaction pour compléter et corriger le réseau



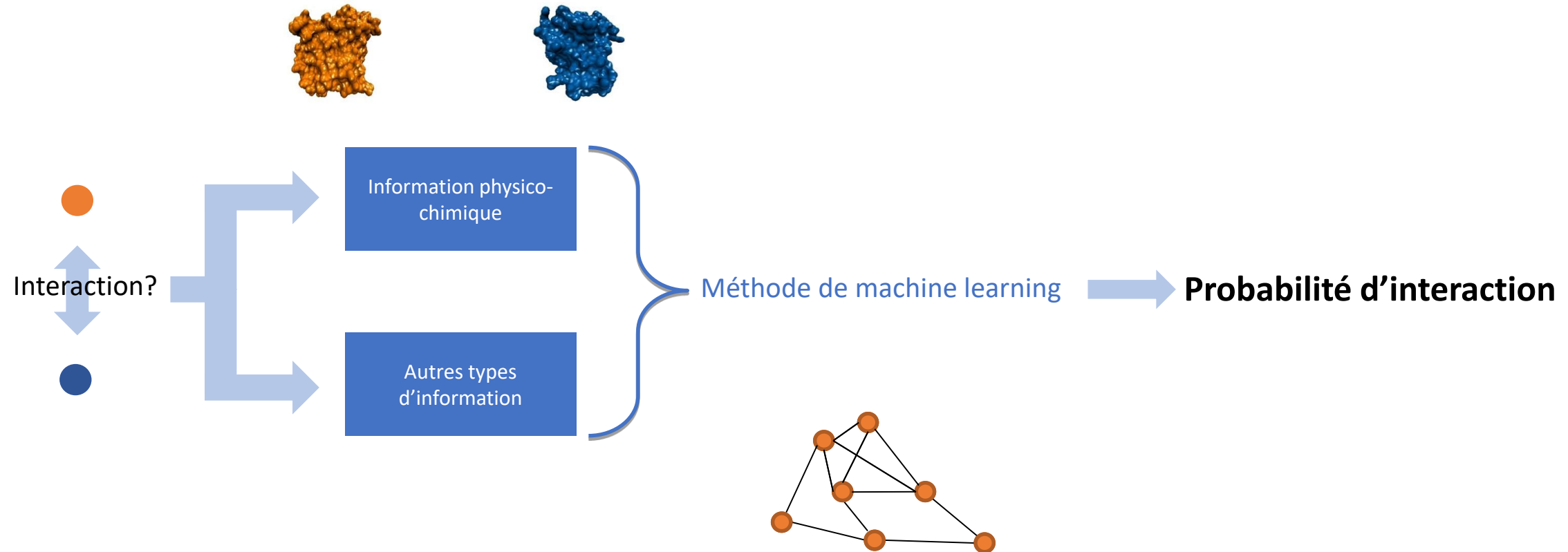
Calcul d'une probabilité d'interaction pour une paire de protéines

Des probabilités d'interaction pour compléter et corriger le réseau



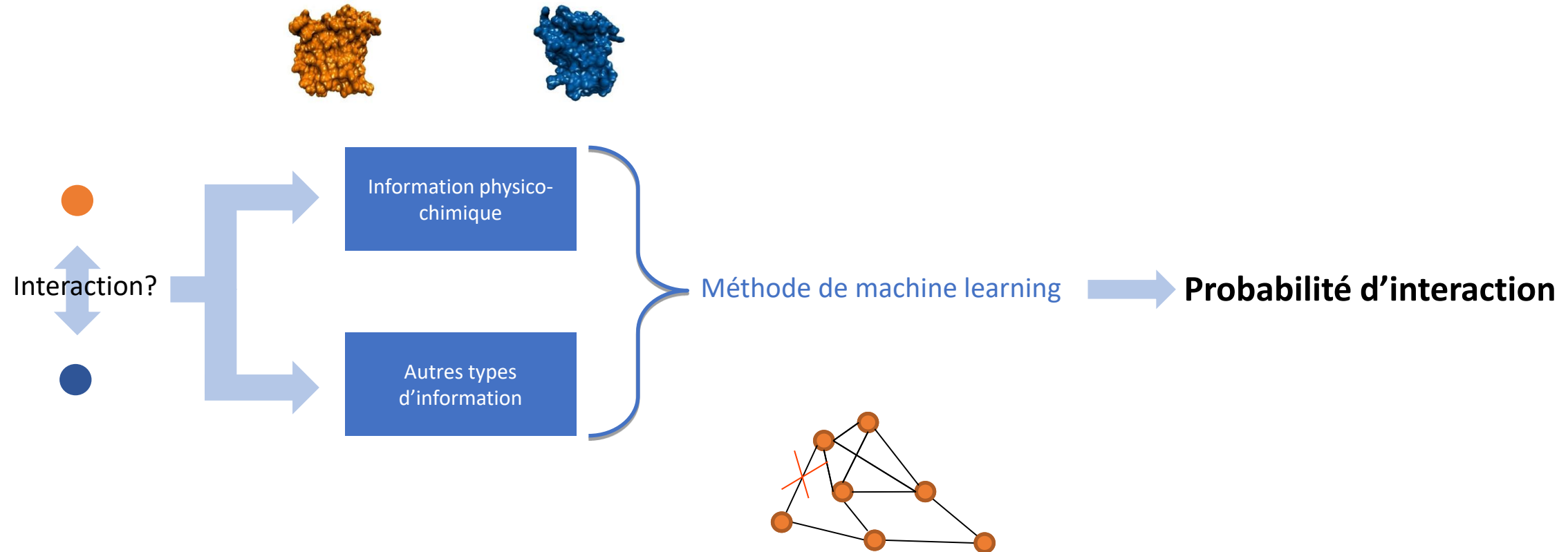
Calcul d'une probabilité d'interaction associée à **toutes les paires de protéines**

Des probabilités d'interaction pour compléter et corriger le réseau



1. Prédire les interactions

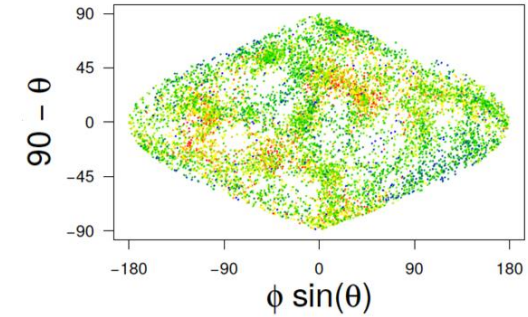
Des probabilités d'interaction pour compléter et corriger le réseau



1. Prédire les interactions
2. **Éliminer les faux positifs**

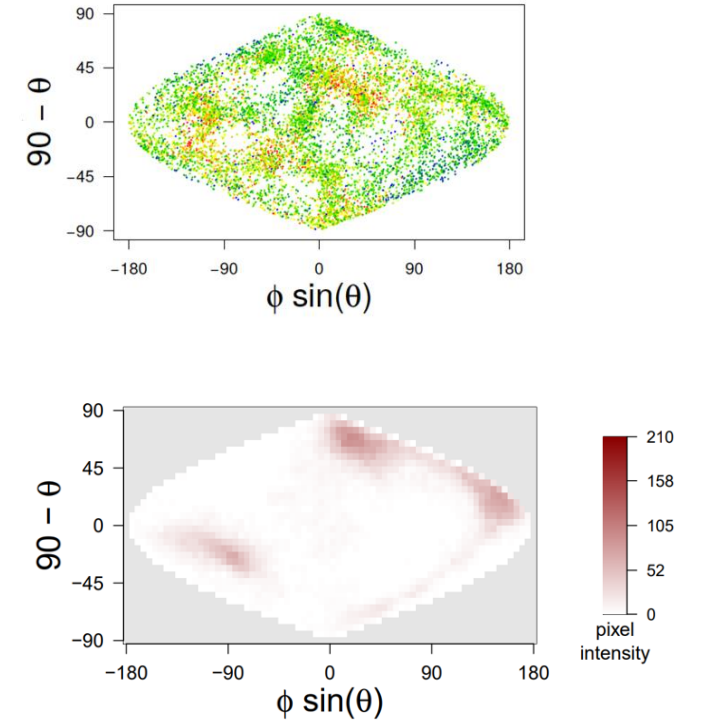
Préparation des données pour la prédiction de partenaires

- Les cartes d'énergies d'interaction



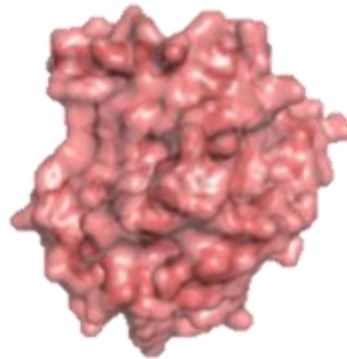
Préparation des données pour la prédiction de partenaires

- Les cartes d'énergies d'interaction
- Les cartes IPOPS (Interaction Propensity Of Protein Surface maps)

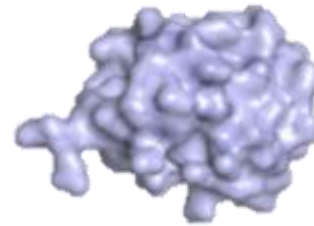


Attract, un algorithme de docking gros grains

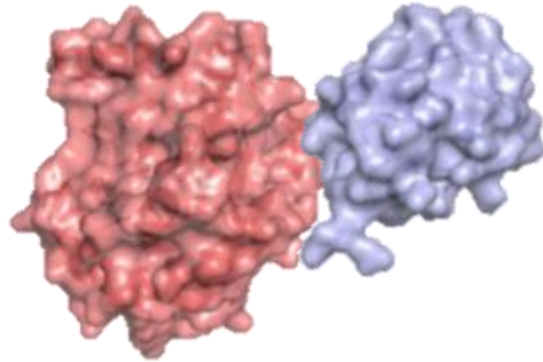
Récepteur



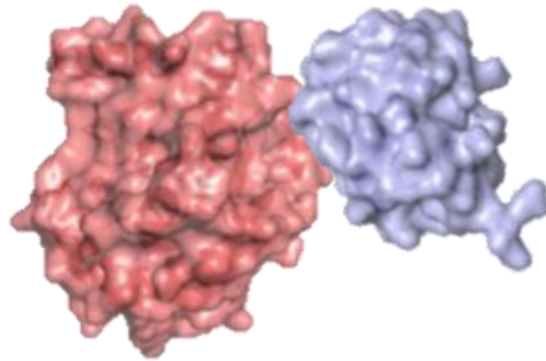
Ligand



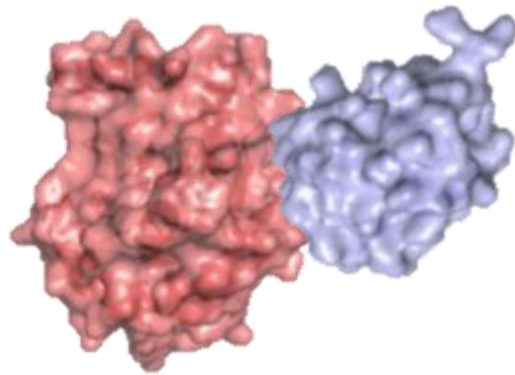
Attract, un algorithme de docking gros grains



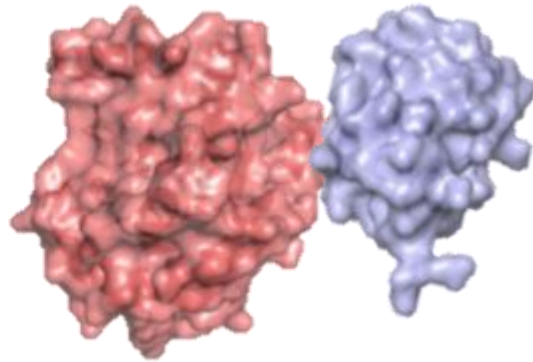
Attract, un algorithme de docking gros grains



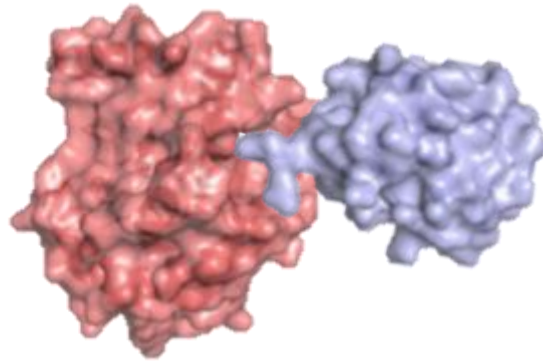
Attract, un algorithme de docking gros grains



Attract, un algorithme de docking gros grains

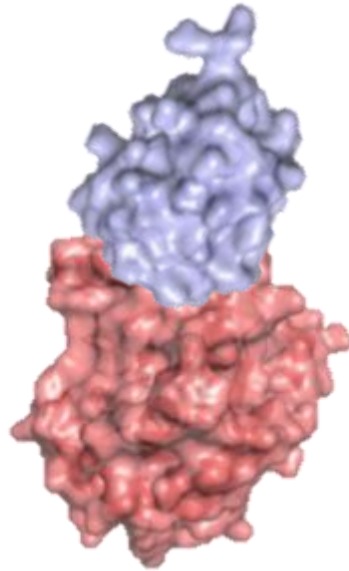


Attract, un algorithme de docking gros grains

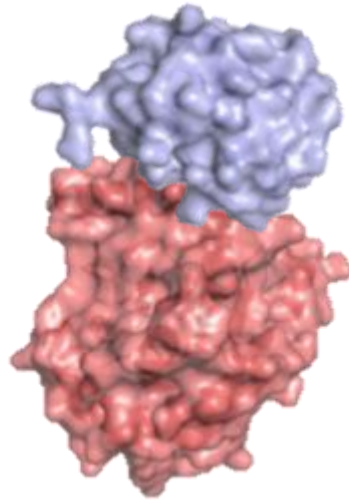


Conformation optimale

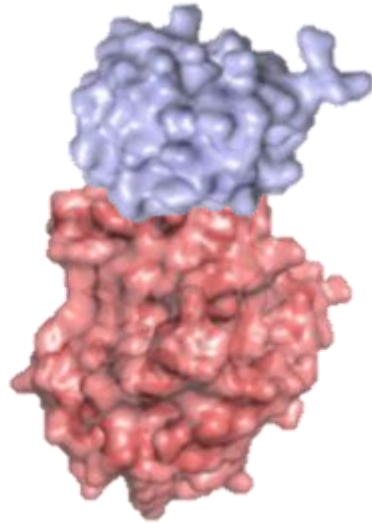
Attract, un algorithme de docking gros grains



Attract, un algorithme de docking gros grains

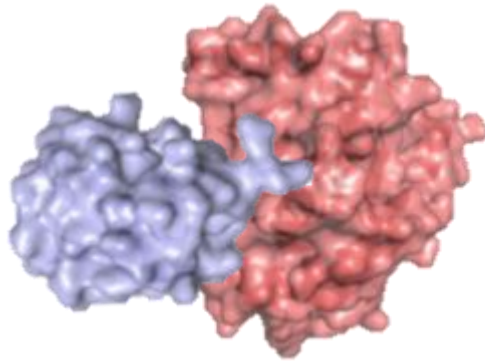


Attract, un algorithme de docking gros grains



Conformation optimale

Attract, un algorithme de docking gros grains



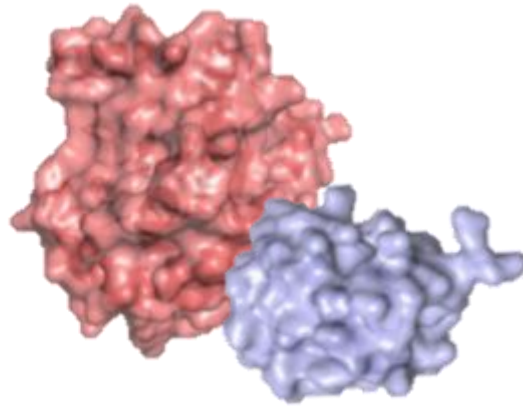
Conformation optimale

Attract, un algorithme de docking gros grains



Conformation optimale

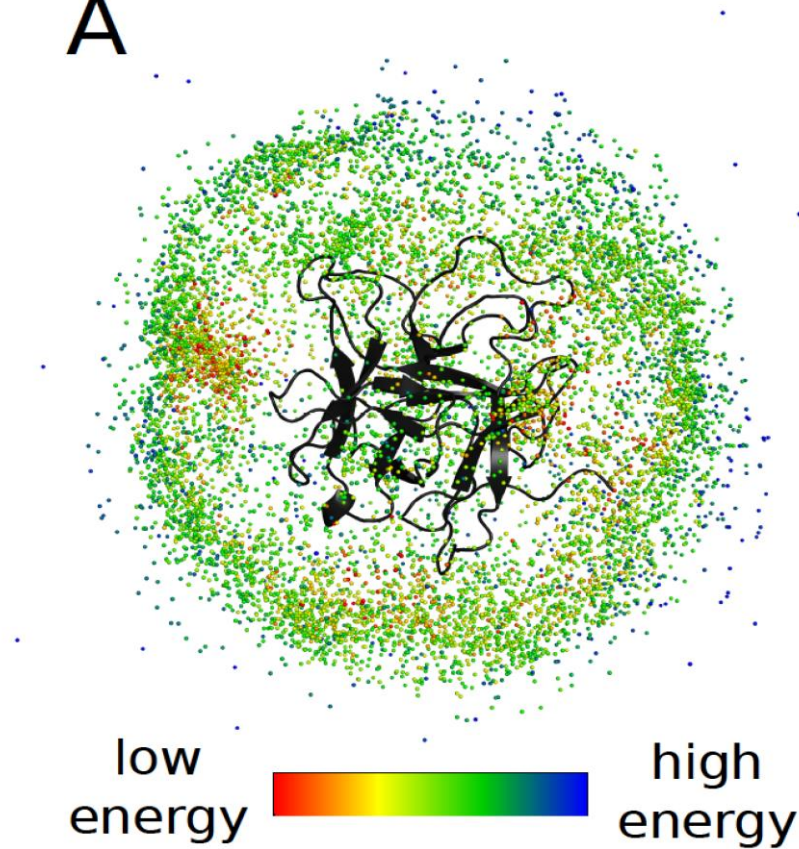
Attract, un algorithme de docking gros grains



Conformation optimale

Attract, un algorithme de docking gros grains

A

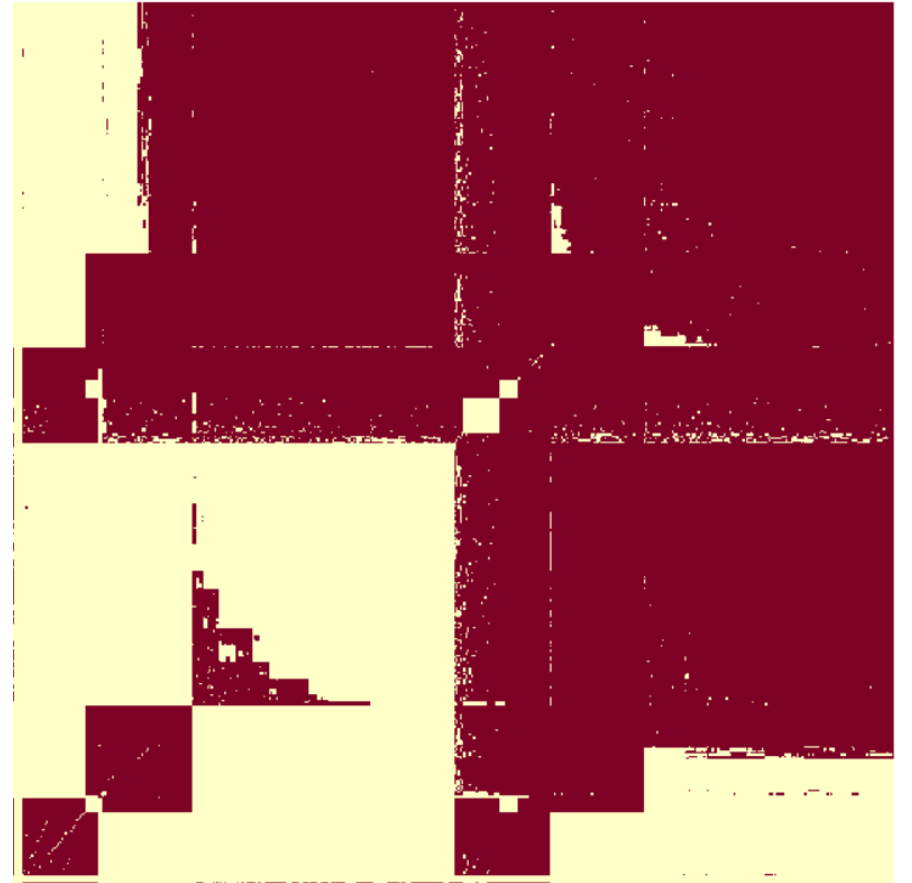


Les données à notre disposition chez *Cerevisiae*

- 12 To de Données de dockings générées par un précédent doctorant sur le protéome de *S. cerevisiae*
- Données désorganisées → Inventaire nécessaire (première année)

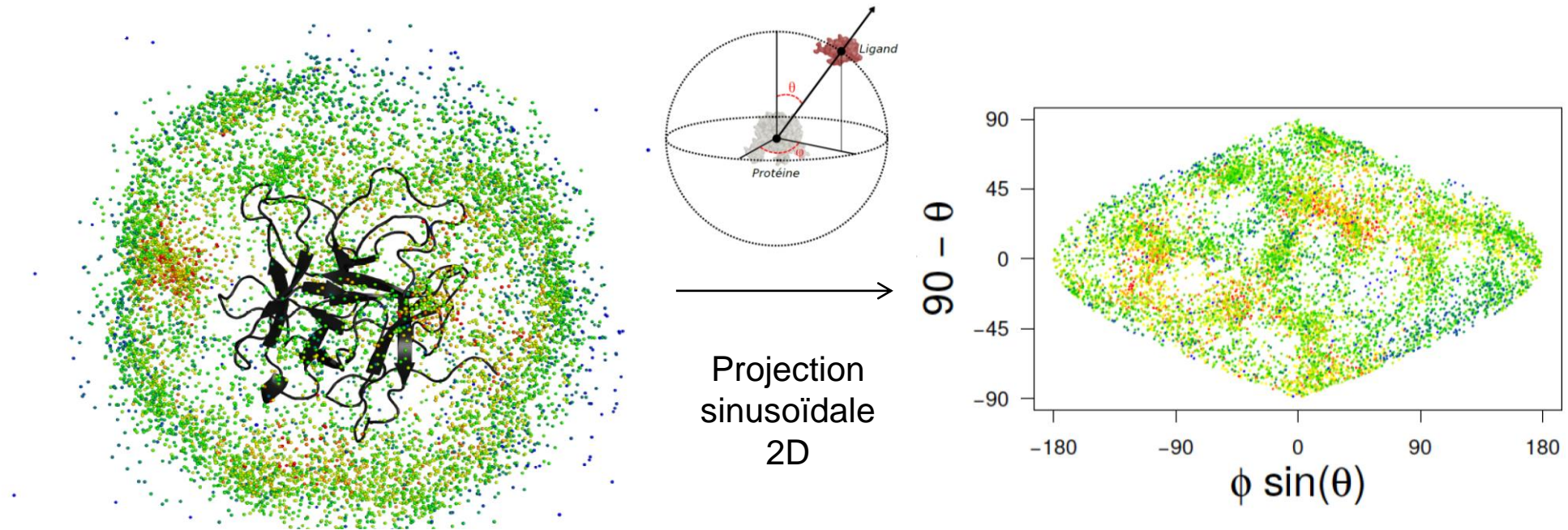
Les données à notre disposition chez *S. cerevisiae*

- Matrice de 471 par 471
 - Structures pdb complètes
 - Modèles de confiances
- Problèmes rencontrés : espace de stockage, fichiers vide, Na...



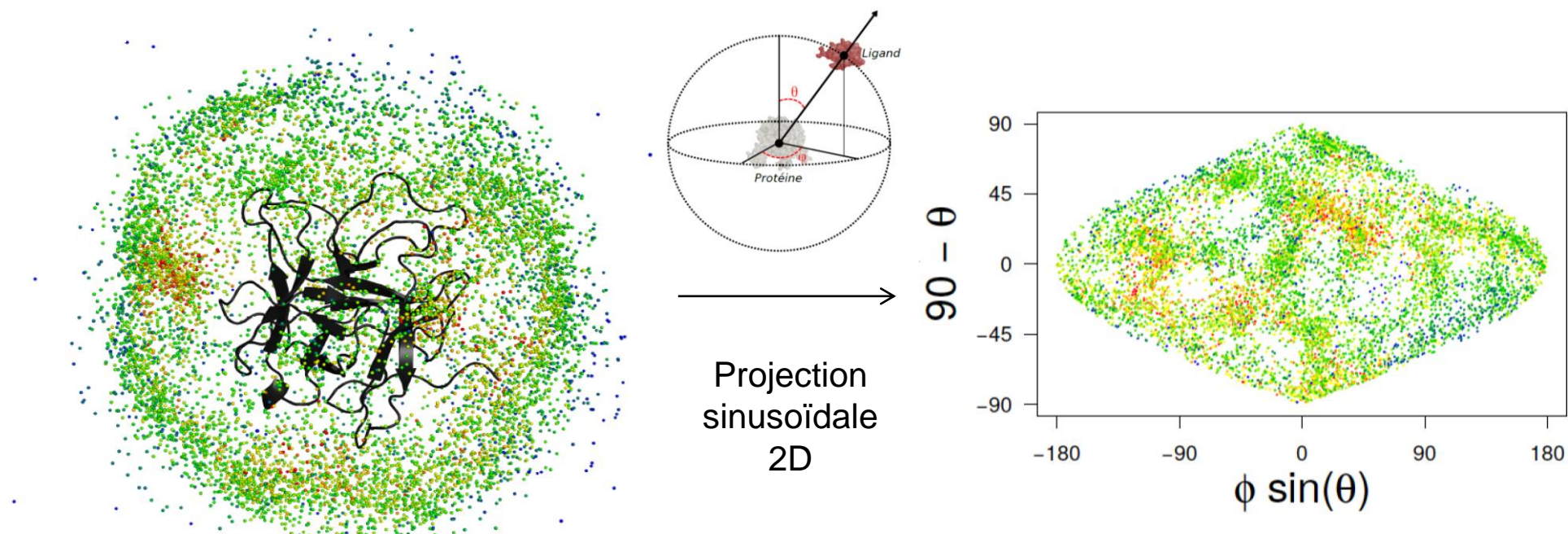
Visualisation des résultats

PAYSAGE ÉNERGÉTIQUE D'INTERACTION



Visualisation des résultats

PAYSAGE ÉNERGÉTIQUE D'INTERACTION



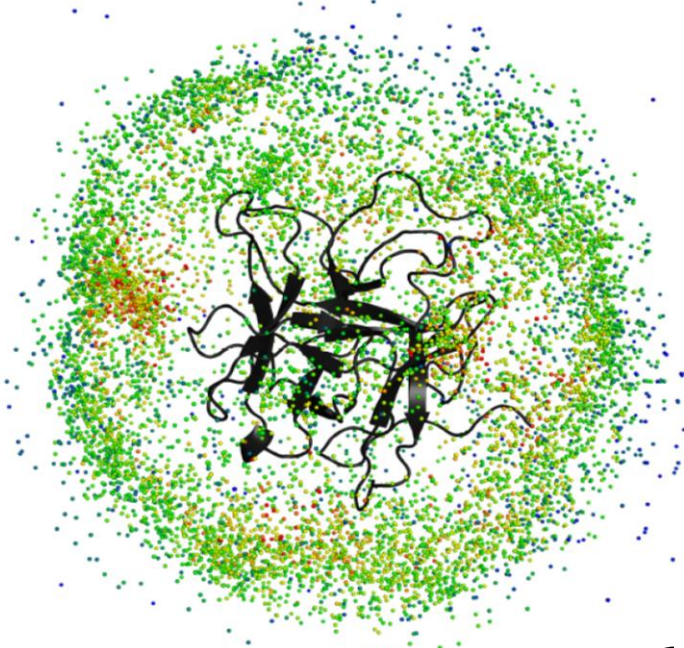
SURFMAP: A Software for Mapping in Two Dimensions Protein Surface Features

Hugo Schweke, Marie-Hélène Mucchielli, Nicolas Chevrollier, Simon Gosset, and Anne Lopes

Journal of Chemical Information and Modeling **Article ASAP**

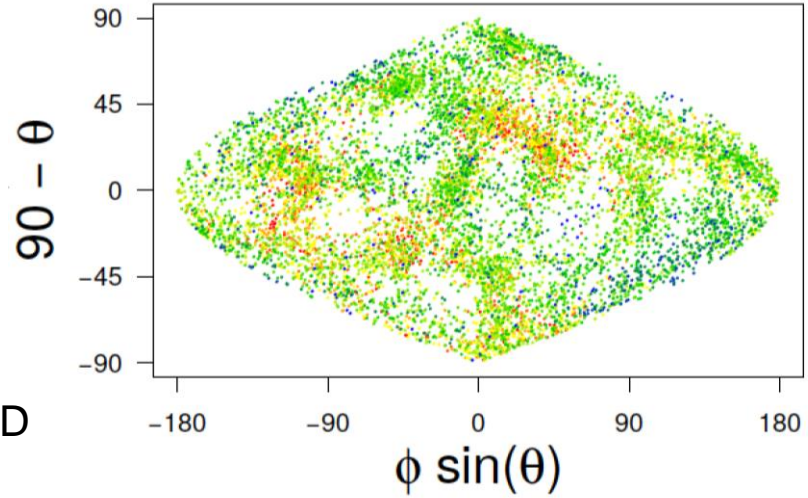
DOI: [10.1021/acs.jcim.1c0126](https://doi.org/10.1021/acs.jcim.1c0126)

Lissage des cartes d'énergies



2D energy map

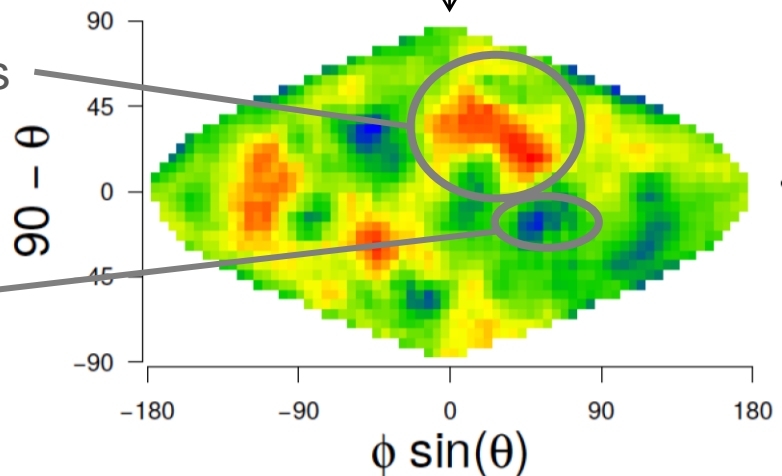
Projection
sinusoïdale en 2D



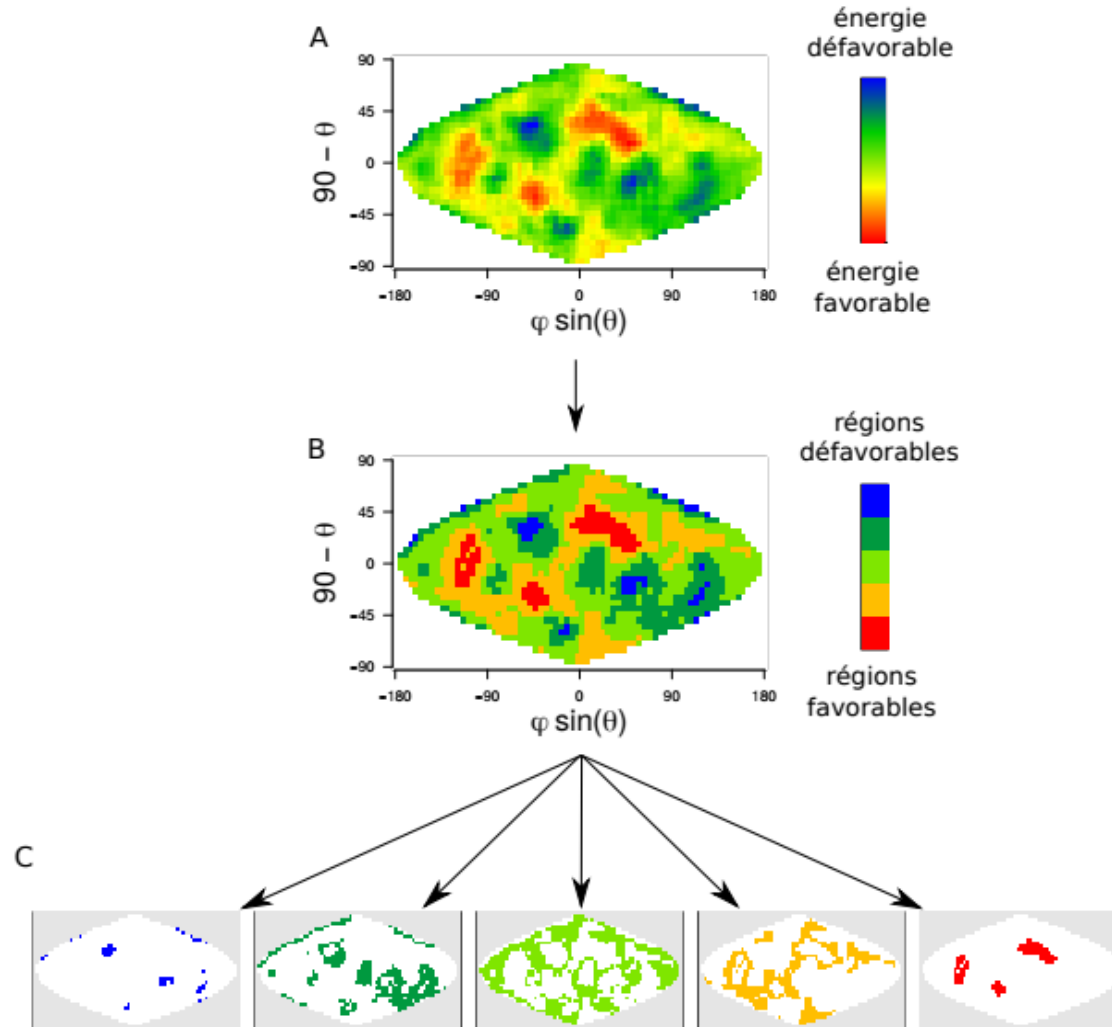
lissage

régions chaudes

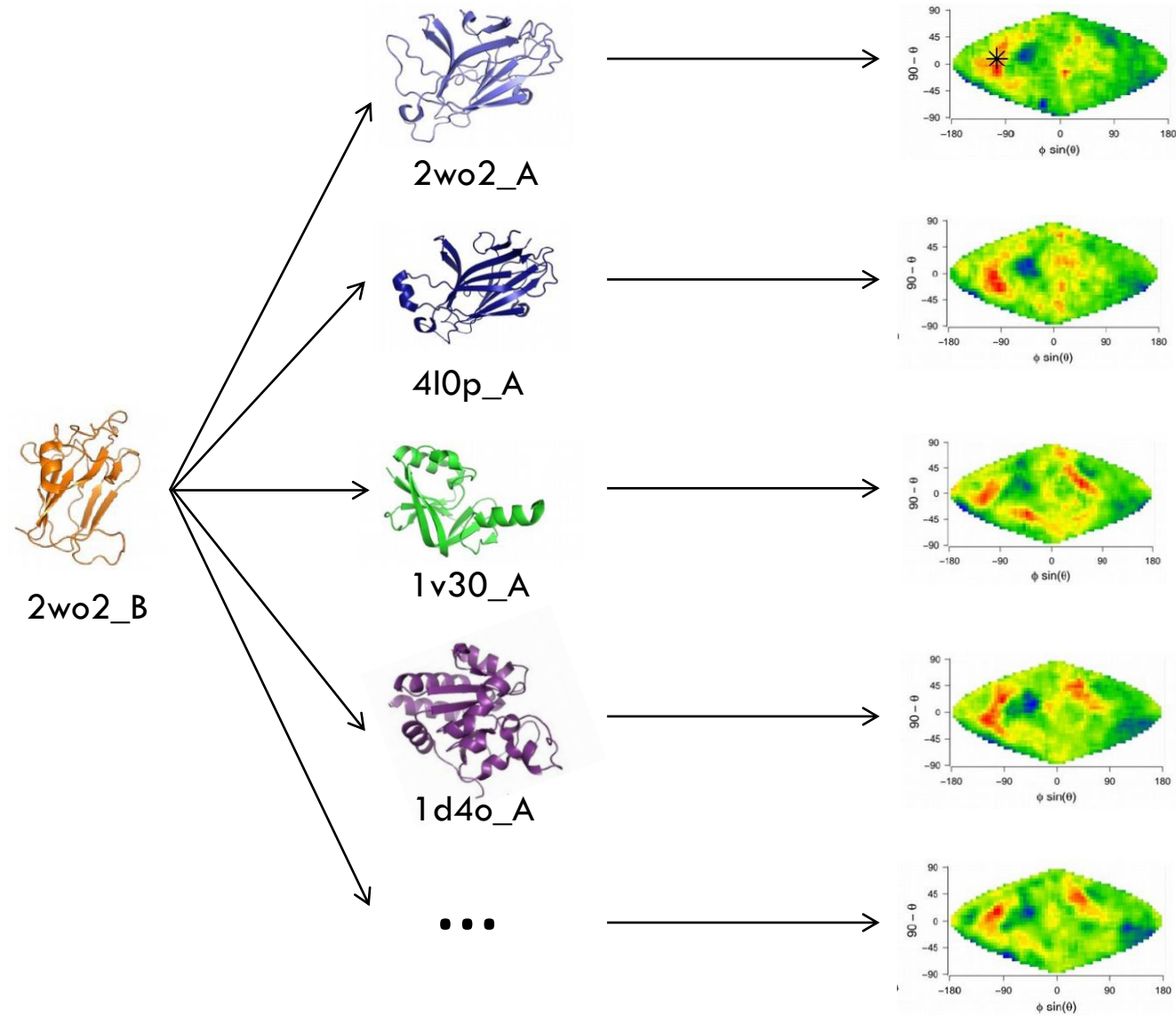
régions froides



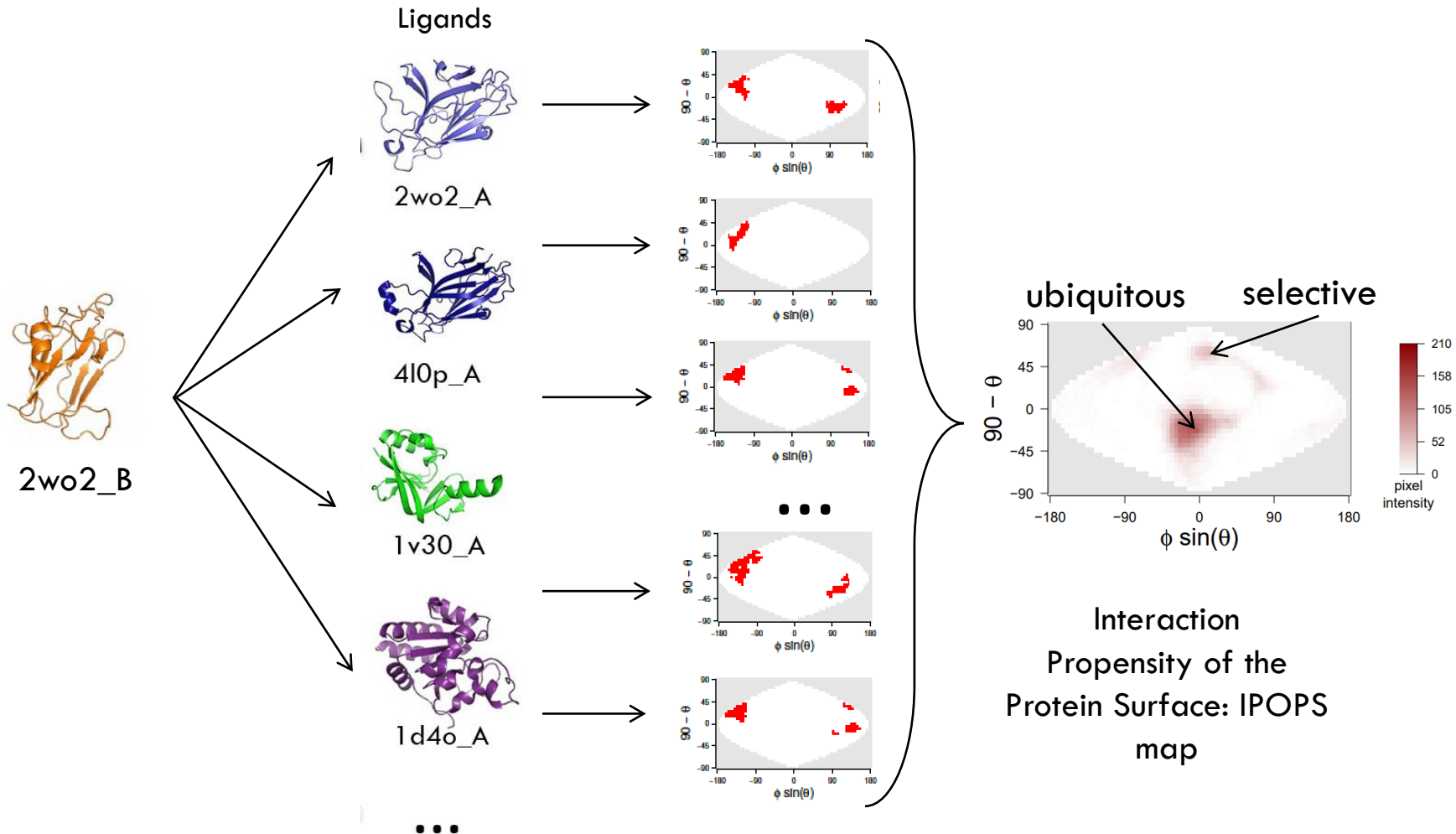
Discrétisation et séparation des classes



Obtention des cartes IPOPS



Obtention des cartes IPOPS



























Schweke, Mucchielli, Sacquin-Mora, Bei & Lopes, JMB, 2020

Obtention des cartes IPOPS



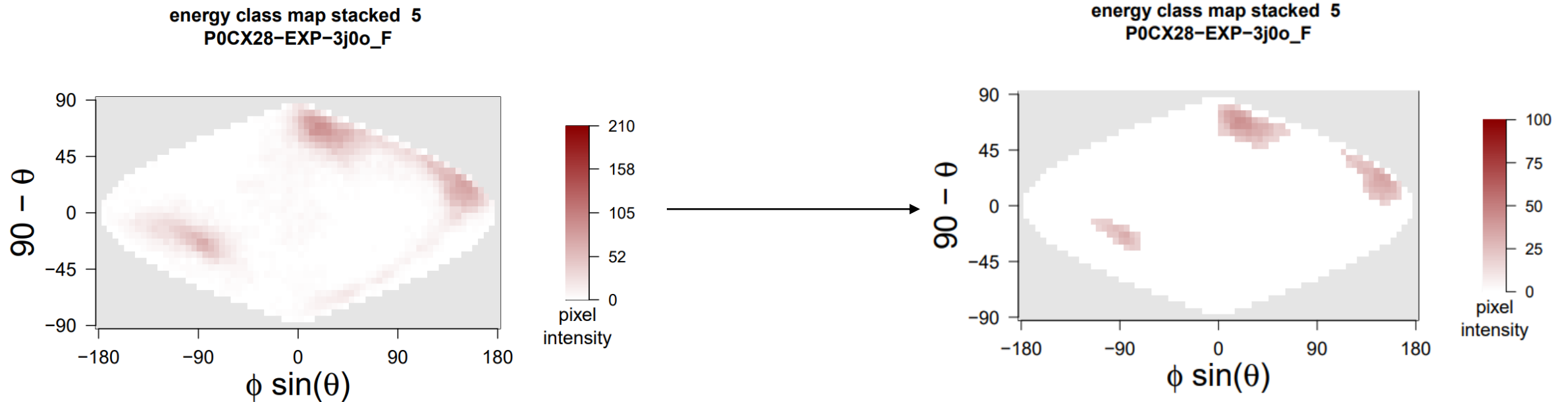
Ligands

Récepteurs

				...					
				...					
				...					
				...					
...
				...					
				...					
				...					
				...					
				...					

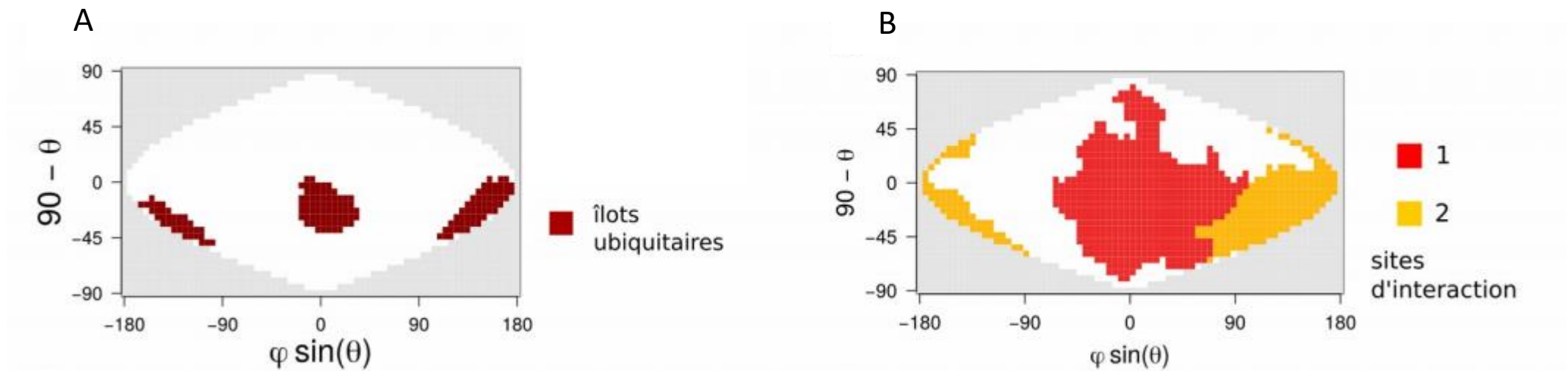
Traitement des cartes – extraction des îlots

- Les cartes sont normalisées puis nettoyées



Carte IPOPS

- Correspondance entre zone favorable à l'interaction et site d'interaction



Carte IPOPS des zones favorables à l'interaction (A) et cartes des site d'interactions (B) pour 3rpf_B

Carte IPOPS

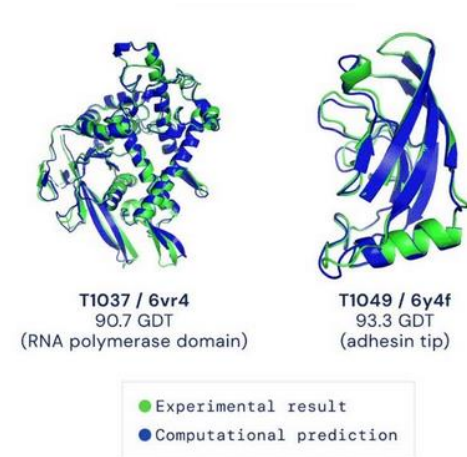
- Correspondance entre zone favorable à l'interaction et site d'interaction
- Utile pour la prédiction des sites mais qu'en est-il des partenaires ?

Pipeline IPOPS

- Perspective à court terme :
 - Le pipeline sera intégré à la plateforme RPBS, collaboration avec Sjoerd De vries
 - Observation sur des cartes 2d de différentes propriétés des protéines

Réflexion et changement de stratégie

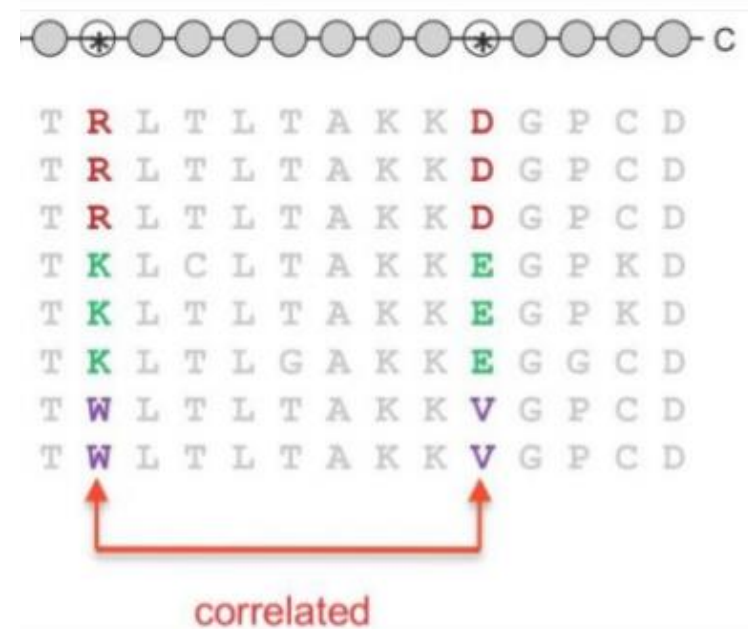
- Utilisation d'Alphafold pour la prédiction de PPI
 - ➔ Nécessité de revoir notre stratégie de prédiction



Highly accurate protein structure prediction with AlphaFold
<https://www.nature.com/articles/s41586-021-03819-2>

Réflexion et changement de stratégie

- Utilisation d'Alphafold pour la prédiction de PPI
 - Nécessité de revoir notre stratégie de prédiction
- Prédiction notamment à partir de données de coévolution



Prédiction de partenaires avec Rosetta et Alphafold

- **Computed structures of core eukaryotic protein complexes** - Humphrey et al; science novembre 2021

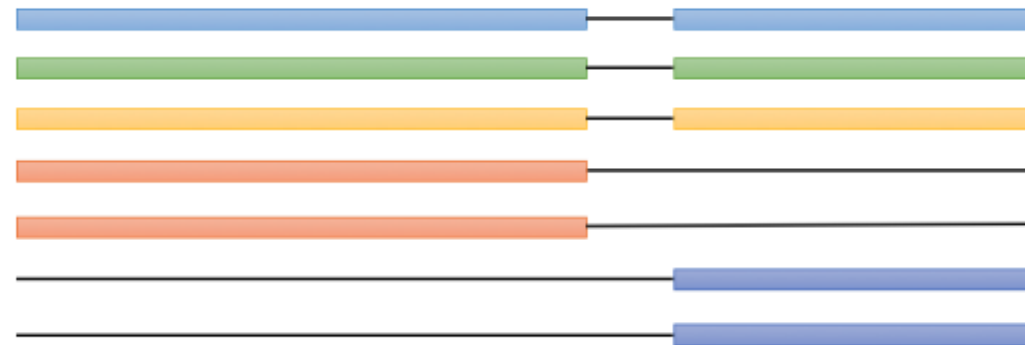
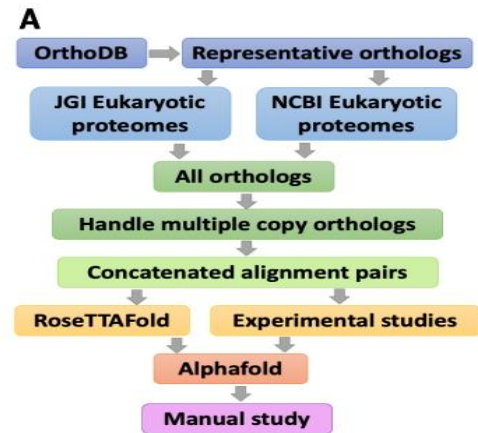
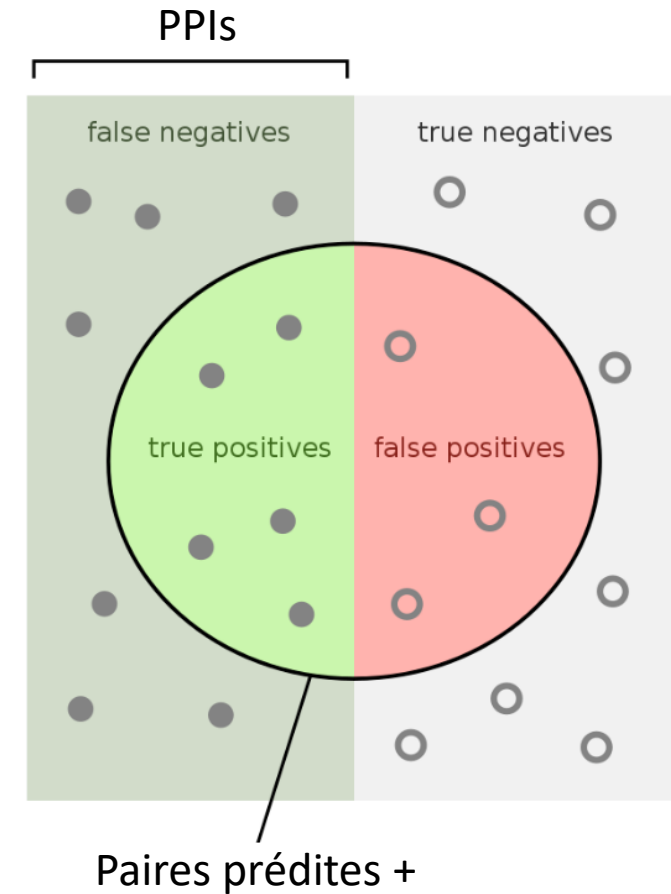


Figure S2. Diagram for paired multiple sequence alignments.

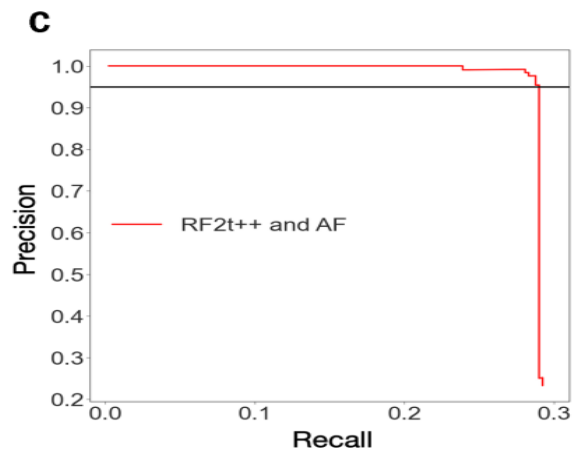
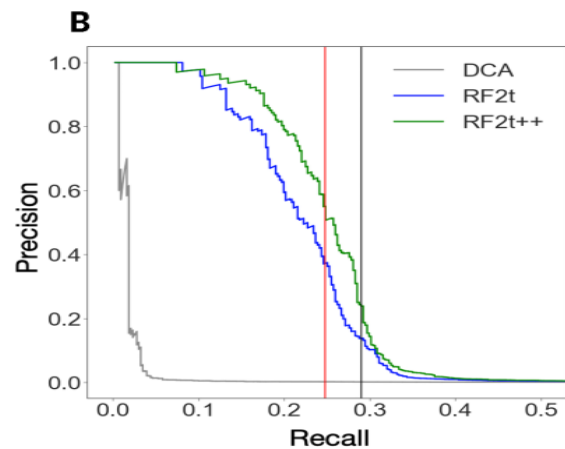
Prédiction de partenaires avec Rosetta et AlphaFold

- TP = PPIs prédites correctement
- FN = PPIs prédites comme non-PPIs
- TN = non-PPIs prédites correctement
- FP = non-PPIs prédites comme PPIs



Prédiction de partenaires avec Rosetta et Alphafold

- **Computed structures of core eukaryotic protein complexes** - Humphrey et al; science novembre 2021



Légendes

B. RoseTTA results sur un set de 768 PPI +768000 non interaction de *S. cerevisiae*

C. Résultats d'AlphaFold2 sur les PPIs prédites par Rosetta Fold

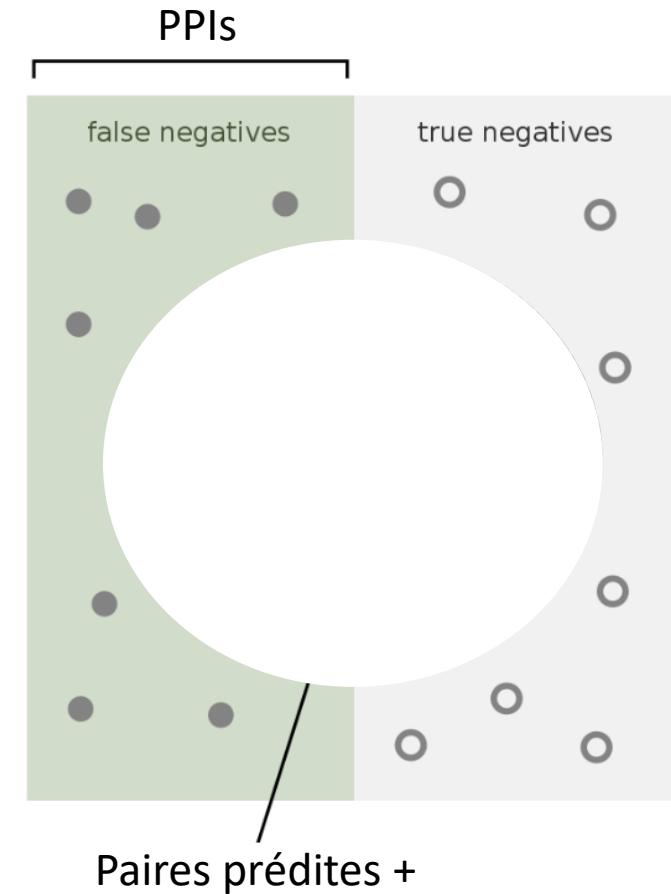
Precision=TP / (TP+FP)=0.95

Recall=TP / (TP+FN)=0.29

Les PPIs sont prédites de manière très sûr mais un grand nombre de PPIs sont prédites incorrectement comme non-PPI

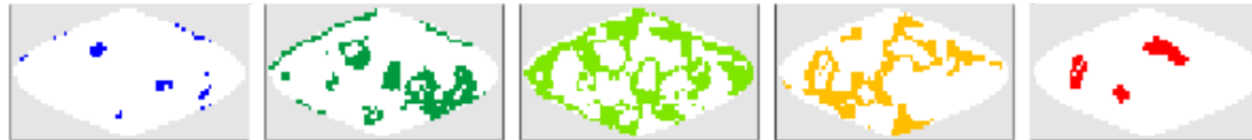
Réflexion et changement de stratégie

- La précision étant très bonne
→ Se concentrer sur ce qui a été prédit comme négatif



Nouvelle stratégie - Chez *S. cerevisiae*

- Chez *S. cerevisiae*, étudier la propension à l'interaction des surfaces d'interactions pour les FN et les VP
- Toutes les données disponibles
- Dans les négatifs, chercher des marqueurs ou des métriques pour séparer les PPIs des non PPIs



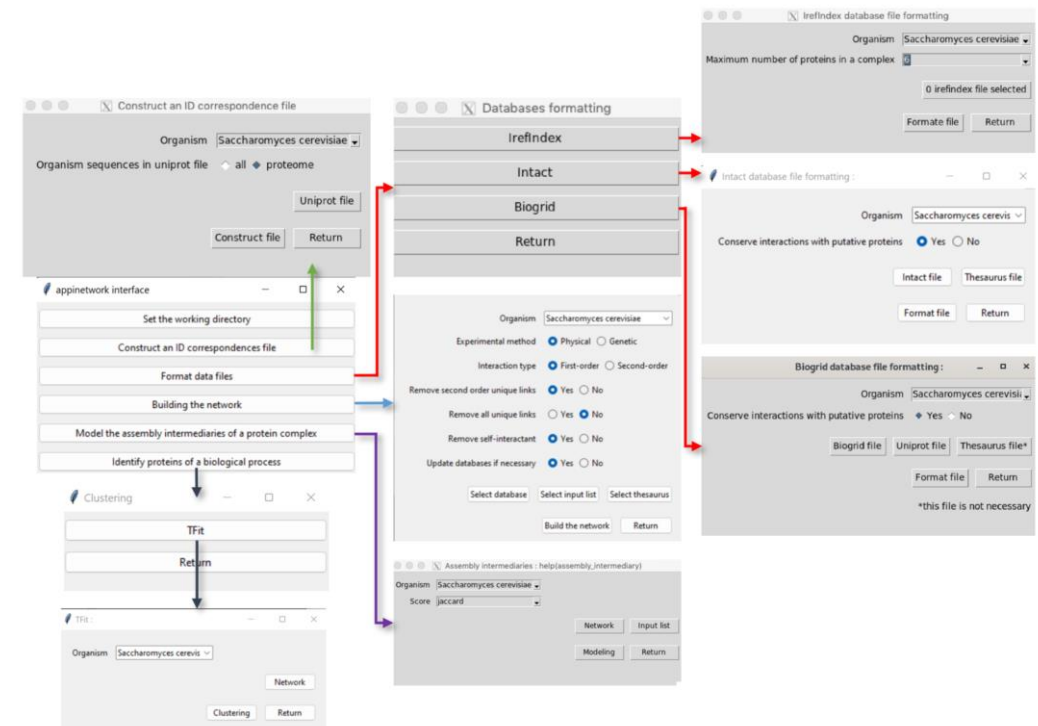
Nouvelle stratégie - Chez *A. thaliana*

- Application sur les protéines du chloroplaste :
 - Set de protéines réduit
 - Données PPIs et expertise dans le labo
- Set de 1500 protéines chloroplastiques
 - Récupérer leurs structure sur la pdb/base de donnée Alphafold
 - Faire le tri dans ces structures en sélectionnant les structures de qualités
 - Préparer les structures sélectionnées au docking contre 100 protéines de backgrounds

Package APPINetwork

- Permet de construire un réseau de PPI à partir d'une liste en cherchant des PPIs dans des bases de données publiques ou dans ses données personnelles
- Inclut des outils d'analyse du réseau
- Interface et guide utilisateur avec des exemples pour une utilisation facilitée

PeerJ



APPINetwork: an R package for building and computational analysis of protein-protein interaction networks

Simon Gosset^{1,2*}, Annie Glatigny^{3*}, Méлина Gallopin³, Zhou Yi³, Marion Salé³, Marie-Hélène Mucchielli^{1,2}

Collaboration - Remerciements

- Equipe GNET
- Frederic Desprez
- Dario Monachello
- Franck Samson
- Anne Lopes
- Hugo Schweke
- Sjoerd de Vries

