

ADT VSM-G3 (2021-2023)

CAPSID

Marie-Dominique Devignes

Malika Smail-Tabbone

Bernard Maigret

SED

Olivier Demengeon

(50%)

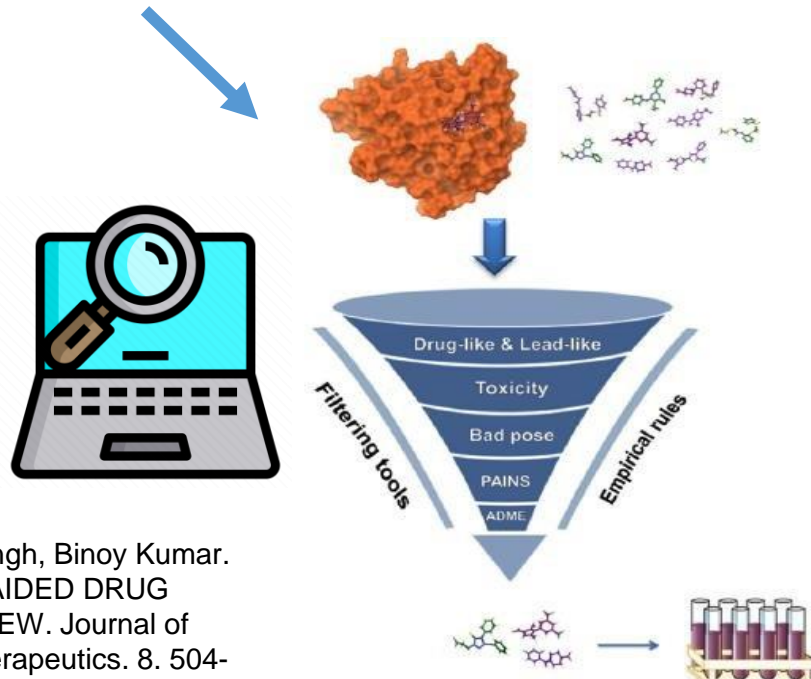
Contexte du projet VSM-G3

- Capsid: Computational Algorithms for Protein Structure and Interactions
- Expertise en docking de petites molécules: Bernard Maigret (DR CNRS émérite)
 - Nombreuses publications et collaborations, nombreux succès dans la découverte de petites molécules d'intérêt thérapeutique
 - Intérêt applicatif essentiellement, réponse à des demandes locales de nos collaborateurs
 - Importance de capitaliser et valoriser cette expertise en criblage virtuel: automatiser et rendre accessible ce qui peut l'être

VSM-G3

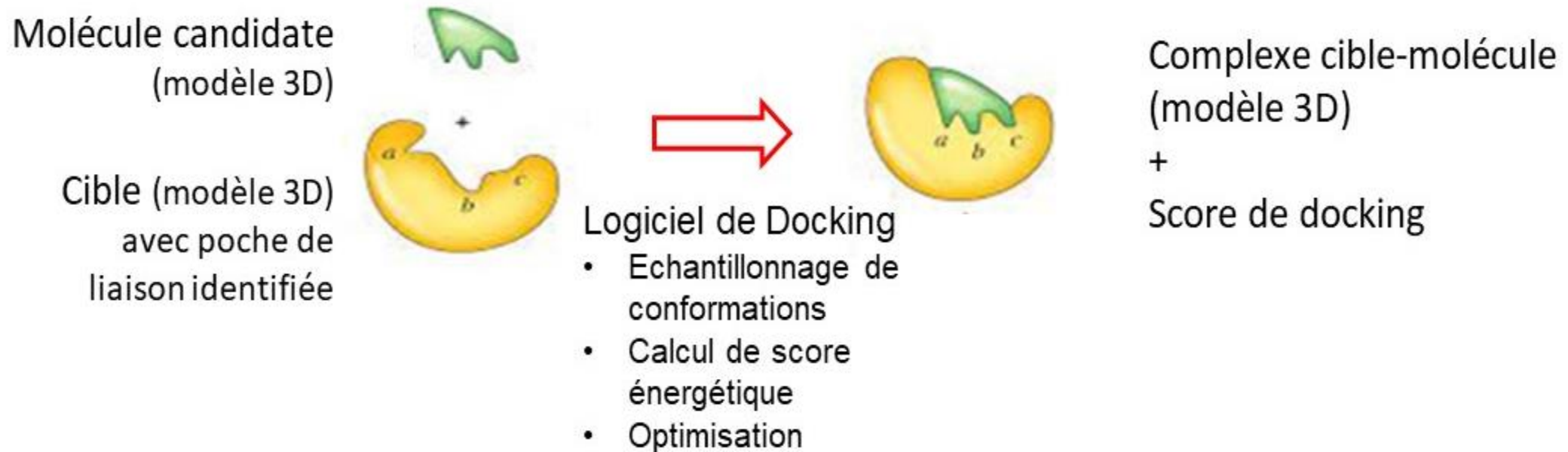
Virtual Screening Manager- Generation 3

- Criblage : passer de centaines de milliers de molécules à quelques candidats pour un test particulier
- ‘Virtuel’: par opposition à ‘Expérimental’



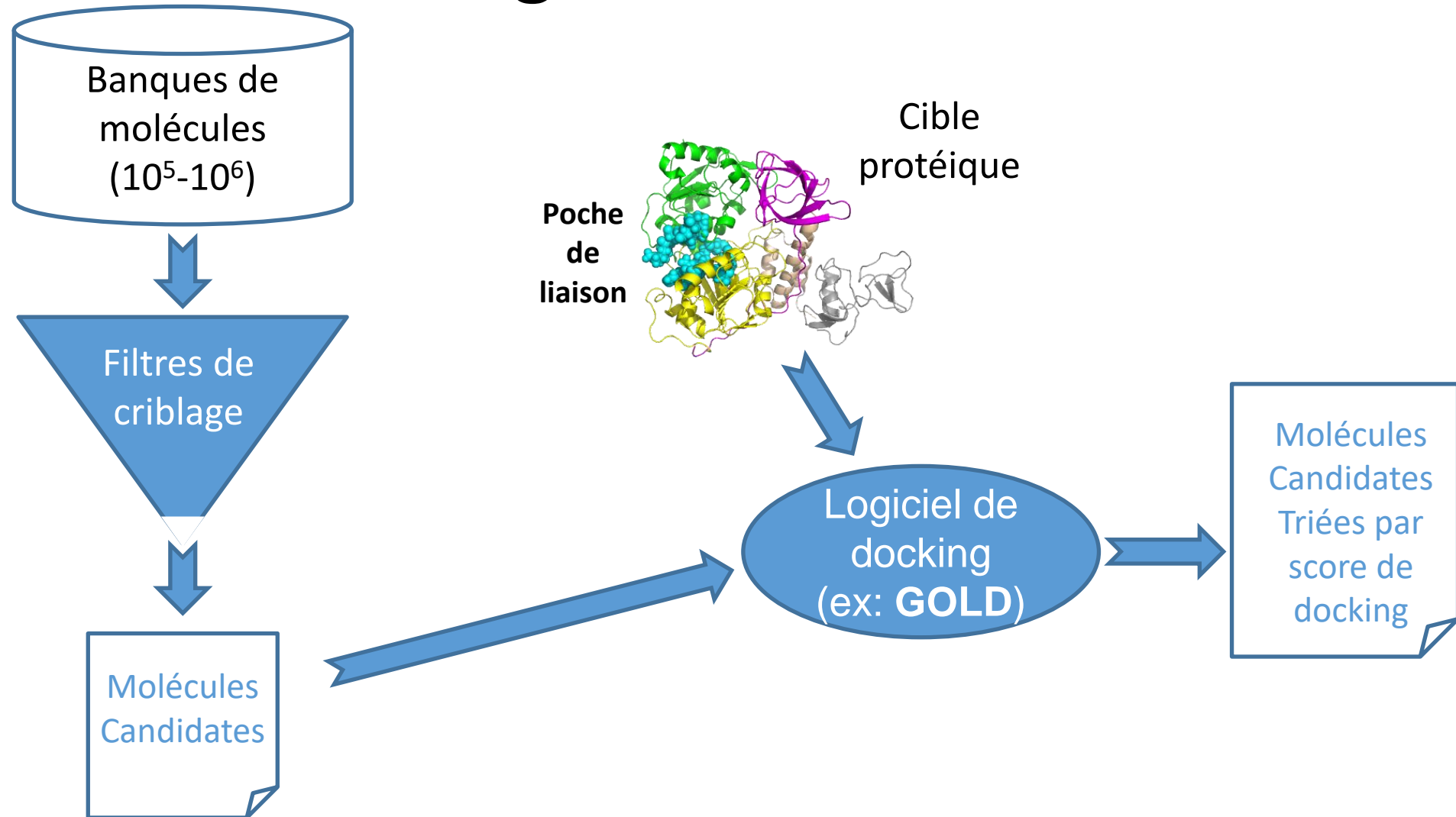
Une étape-clé du criblage virtuel : le docking moléculaire

Principe du docking de petites molécules sur une protéine cible



**Etape essentielle mais coûteuse en calcul -> distribution sur la grille de calcul
Impossible à réaliser sur des banques de 10^5 molécules**

Scénario de criblage virtuel



VSM-G3

- M comme Manager ... de **scripts et de programmes** (sources externes et développements “maison”)
 - Gérer la variété, et l’hétérogénéité des programmes!
 - Rendre possible la création de pipelines
 - Solution open-source et FAIR ‘[Galaxy Platform](#)’ (consortium international)
- G3 : 3^{ème} génération (après VSM, puis VSM-G: thèse A. Beautrait en 2008)
 - G comme **Grid**
 - G comme **GOLD**
 - G comme **Galaxy**

Objectifs de l'ADT

- Regroupement et standardisation des programmes de VSM-G dans l'environnement de workflow Galaxy
 - Etat initial : scripts dispersés, non maintenus, sans garantie d'exécutabilité.
 - Etat final : scripts accessibles sous une forme exécutable dans l'interface
 - Création de pipeline facilitée, possibilité d'interfacer des modules d'analyse supplémentaires
 - Possibilité de prise en main par des utilisateurs non bioinformaticiens
- Augmentation de l'impact de l'équipe dans le domaine de la bioinformatique structurale et du criblage virtuel
 - Formation de collaborateurs, d'étudiants (master classes)
 - Relations avec le GT Galaxy de l'Institut Français de Bioinformatique
- En pratique : ADT non financée (décision CDT 12 juillet 2021) mais ingénieur SED mis à disposition à 50% depuis juillet 2021

Etat d'avancement (automne 2023)

- Instance Galaxy opérationnelle (V1)
 - update en cours (V2), impératifs de sécurité
- GOLD via Galaxy
 - préparation d'une expérimentation sur un cluster (cluster MBI)
 - lancement de l'exécution sur le cluster
 - récupération des résultats
- Ajout d'autres outils développés en local ou partagés par d'autres
 - Recherche de poches (F-pocket)
 - Calculs de barycentre à partir d'une sélection d'acides aminés
 - Génération de fichiers « poches » au format pdb à partir d'une sélection d'acides aminés
 - A venir : logiciel de chimie combinatoire
- Diffusion : utilisation en master class au Brésil (23 et 29 novembre 2023)
 - Franc succès auprès d'étudiants non informaticiens

Les 3G (Galaxy, Gold, Grille) sont opérationnels !
(plus de détails par **Olivier**)

Détails techniques par Olivier Demengeon

**Les 3G ce n'était
pas gagné
d'avance !**



Objectifs

- Mise en place d'une plate-forme logicielle pour faciliter les expérimentations (GALAXY)
- Reprendre un outil permettant d'explorer une dynamique moléculaire afin de pouvoir généraliser son utilisation (Surface)

GALAXY

- <https://galaxyproject.org/>
- Une interface web pour réaliser des expérimentations
- L'utilisateur n'a pas à se soucier de
 - la réalisation de l'expérimentation
 - l'archivage des résultats et des données utilisées

GALAXY (SED)

- Installation de GALAXY dans un docker
 - Virtualisation légère via des conteneurs logicielle
- Sécurisation / Configuration de GALAXY
- Intégration pour utiliser le LDAP de l'INRIA pour l'authentification
- Mettre à disposition GALAXY = déployer une image docker

Tools



search tools



Upload Data

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

CAPSID Tools

Built-in Converters

WORKFLOWS

All workflows

Prepare with list of residues and run GOLD Running GOLD with list of residues on cluster (Galaxy Version 0.9.9)



protein



19: RelST3_AF2_chainA.pdb



ligands



20: List6candidates.mol2



Nb Conformation

50

List of residues

ASP72 ASP177 TYR252 LYS254 GLU257 GLU280

Number of ligand per batch

8

Execute

Prepare dataset containing all files for gold docking.

Parameters :

- **Protein** : Target protein in pdb format.
- **ligands** : Ligands in 3D and corrected by Corina or obabel
- **Coordinates** : **Coordinates of the pocket center of the binding site. Spaces separated, x y z like.**
Example : 5.64 9.85 7.52
- **Center radius** :
- **Target cluster** : Cluster on which to launch docking
- **Number of thread**

History



Rechercher des données



Unnamed history



3.49 MB

19

17



36 : Prepare and run GOLD on data 16 and data 17



35 : Prepare with list of residues and run GOLD on data 20 and data 19



265.3 KB

format **tgz**, génome de référence ?

Args: olivier.demengeon /galaxy /database/objects/4/b/f



binary data

34 : Prepare and run GOLD on data 16 and data 17



33 : Prepare with list of residues and run GOLD on data 20 and data 19



GALAXY (SED)

- Ajout d'outils
 - Docking moléculaire (via le logiciel GOLD) sur cluster
 - Intégration
 - d'outils de CAPSID
 - d'outils « third party »

GALAXY (SED)

- Problème faire fonctionner GOLD sur un cluster
 - GRID 5000 : refus de support
 - CAPSID ne possède que des licences par postes
- Adaptation de scripts fournis par l'éditeur de GOLD pour partager la charge de travail d'un docking sur un cluster (SLURM)

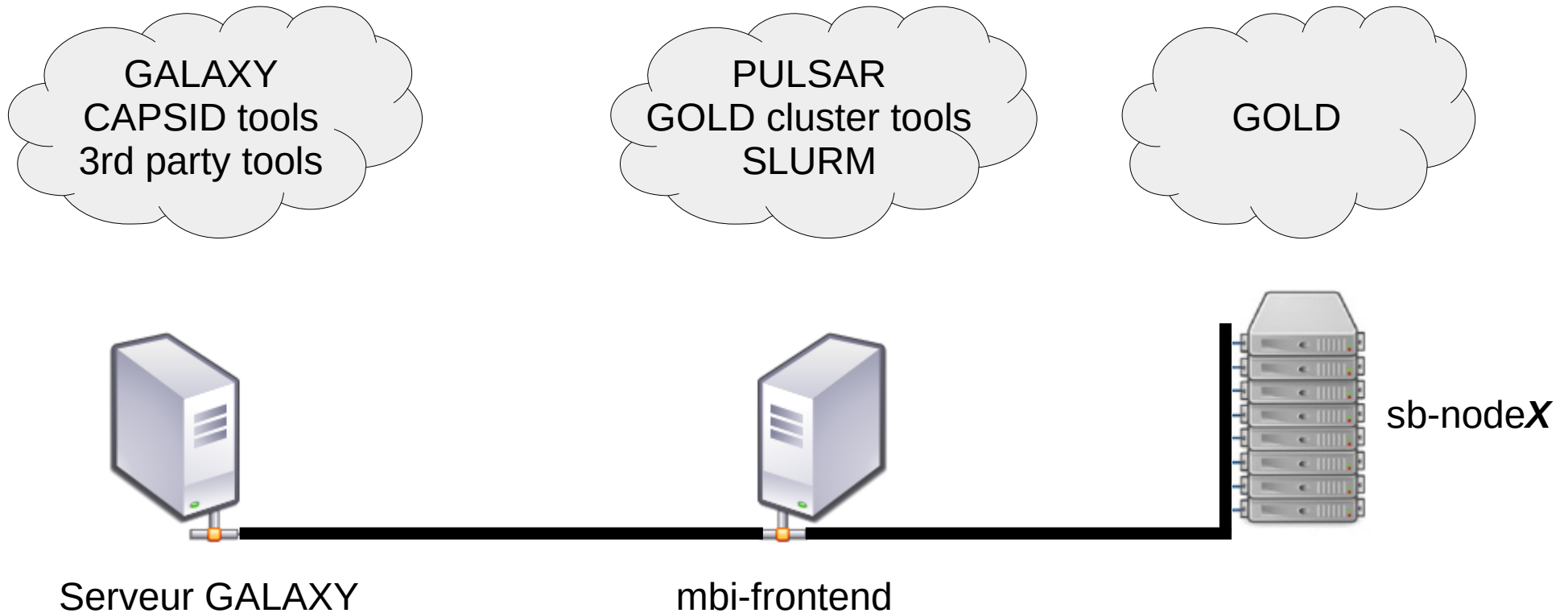
GALAXY (SED)

- Installation / Sécurisation / Configuration de PULSAR sur le cluster de CAPSID pour recevoir des travaux du serveur GALAXY et renvoyer les résultats
- Écriture des « plugins » GALAXY pour proposer des expérimentations avec GOLD

GALAXY (SED)

- Écriture des « plugins » GALAXY pour proposer des expérimentations avec les outils de CAPSID / « third party »

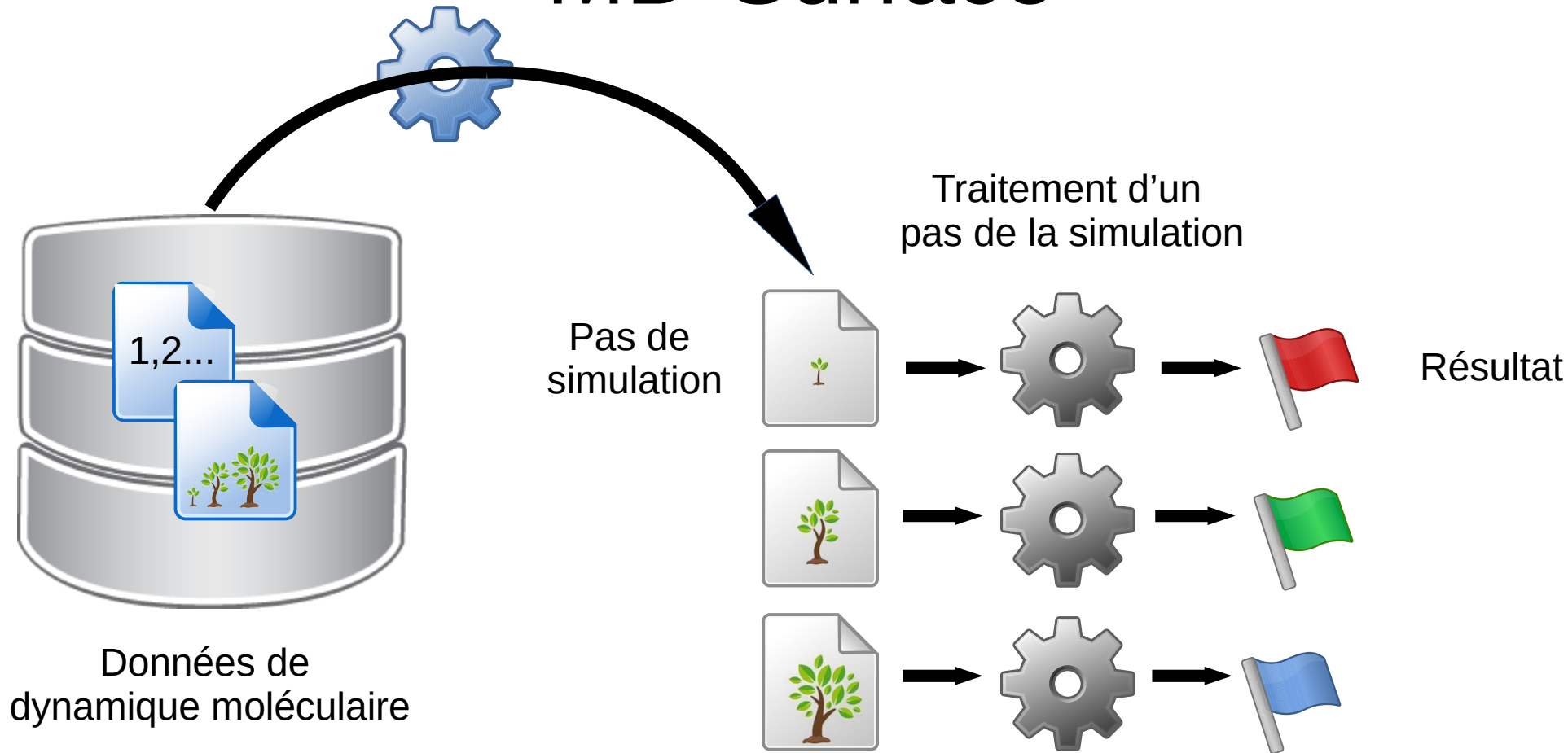
GALAXY (SED)



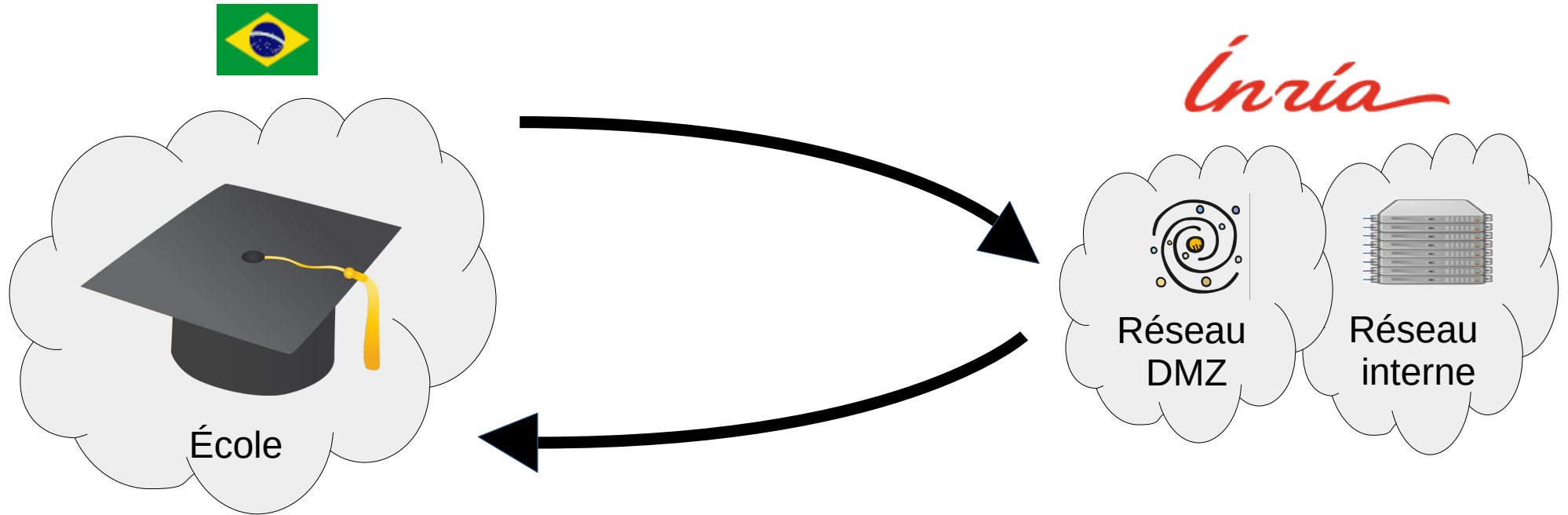
MD-Surface (SED)

- Reprise d'un logiciel pour extraire chaque pas d'une simulation de dynamique moléculaire
 - Pas de documentation conception / développement
- Tester avec un logiciel (fingerprint) calculant surface / volume d'une molécule
 - Développement d'un logiciel pour visualiser les résultats
- Réflexion sur le moyen de l'intégrer à GALAXY de manière générique

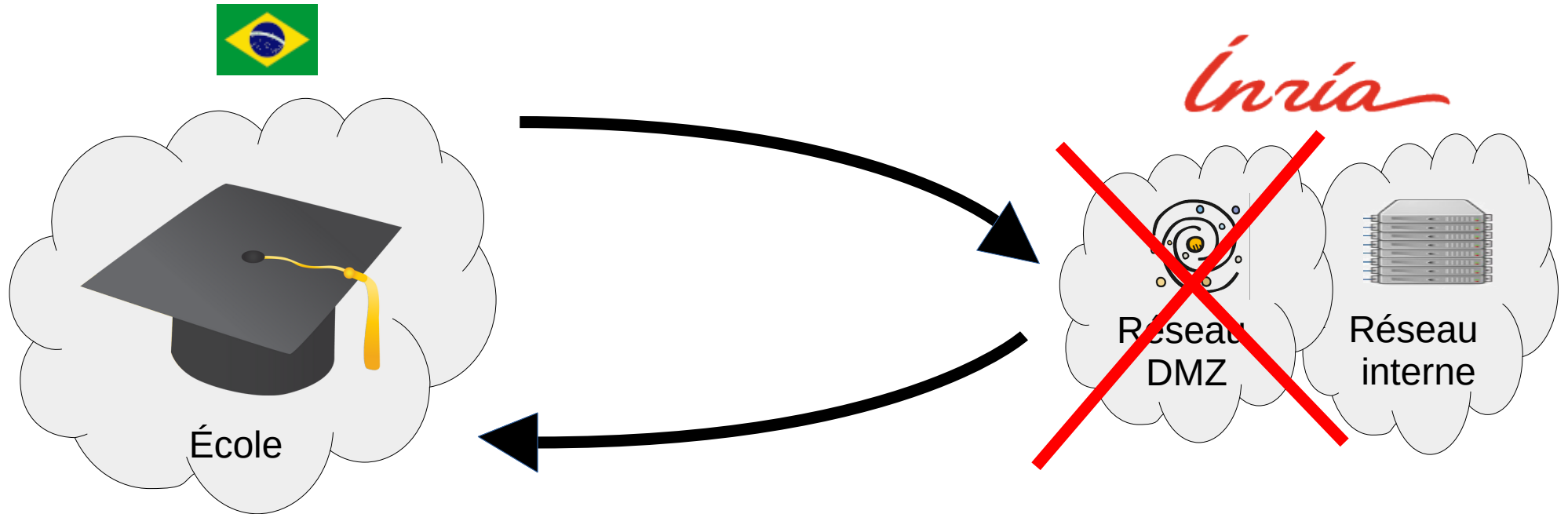
MD-Surface



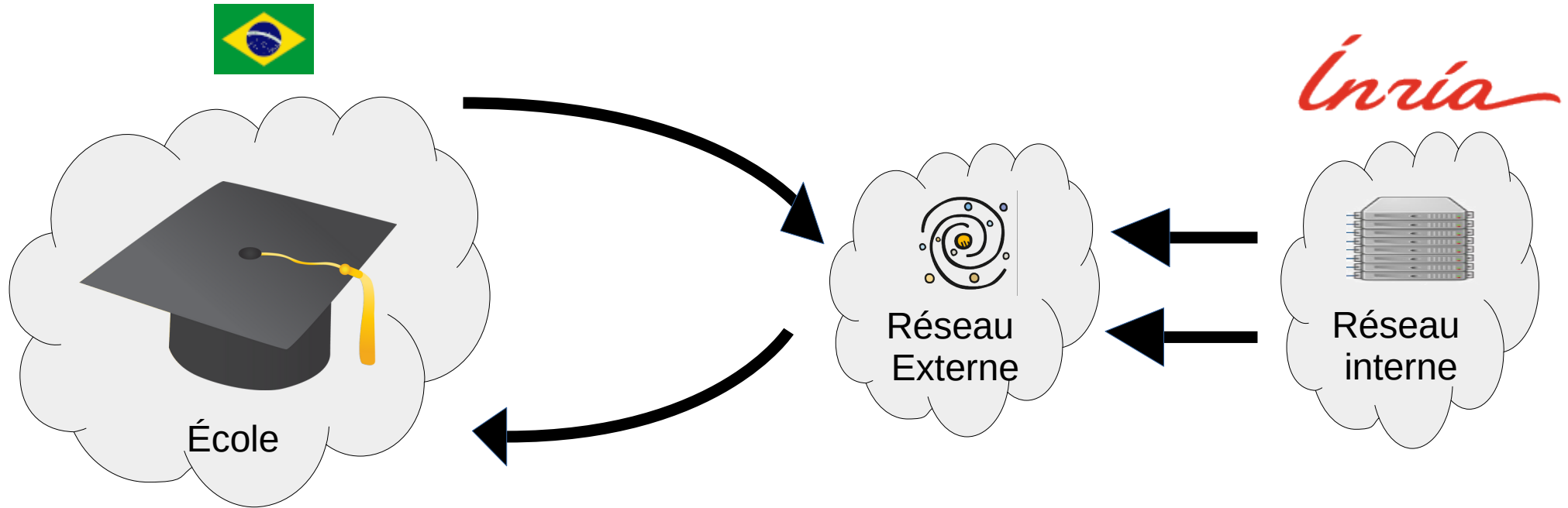
Utilisation de GALAXY



Utilisation de GALAXY

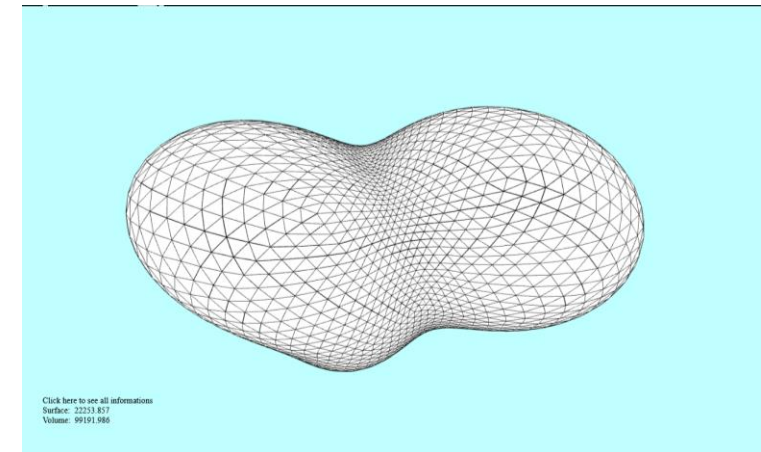


Utilisation de GALAXY



ADT VSM-G3 En cours actuellement

- Reprise par Olivier d'un outil d'analyse des dynamiques moléculaires « **MD-surface** »
 - MD (Molecular Dynamics) simulation numérique de la dynamique des structures moléculaires et de leurs interactions
 - Analyse : pouvoir exécuter des calculs sur tous les pas d'une simulation (10^3)
- Prototype : calcul de volumes et de surfaces sur chaque frame
 - Résultat: visualisation de l'évolution des formes au cours du temps



[Aperçu du résultat](#)

Perspectives

- Phase de test de VSM-G3 par les utilisateurs non informaticiens
 - Retours positifs pendant les master class au Brésil
 - A poursuivre avec certains collègues d'IMOPA ou de DynAMic
 - projet Nsp16 avec Harmonic Pharma
 - Projet Relaxase
- Extension de MD-Surf à de nouveaux outils d'analyses de DM
 - Calculs préparatoires pour les approches Deep Learning (Travaux de Yasaman Karami)
 - Interfaçage possible avec Galaxy pour faciliter la gestion des expériences

Quelques références de résultats obtenus par VSM-G

- Beautrait A, Leroux V, Chavent M, Ghemtio L, Devignes MD, Smaïl-Tabbone M, Cai W, Shao X, Moreau G, Bladon P, Yao J and **Maigret B** (2008) *Multiple-step virtual screening using VSM-G. Overview and validation of fast geometrical matching enrichment* . J. Mol Modeling, 14:135-48.
- Capoci IRG, Faria DR, ..., **Maigret B**. *Repurposing Approach Identifies New Treatment Options for Invasive Fungal Disease*. Bioorg Chem. 2019.
- Kioshima ES, ..., **Maigret B**. *Selection of Potential Anti-Adhesion Drugs by in Silico Approaches Targeted to ALS3 From Candida Albicans*. Biotechnol Letter 2019. doi: 10.1007/s10529-019-02747-6.
- Bresso E, ..., **Maigret B**, Martins NF. *GPCRs from fusarium graminearum detection, modeling and virtual screening - the search for new routes to control head blight disease*. BMC Bioinformatics. 2016.
- Bresso E, Furlan A, Noel P, Leroux V, Maina F, Dono R, **Maigret B**. *Large-Scale Virtual Screening Against the MET Kinase Domain Identifies a New Putative Inhibitor Type*. Molecules. 2020 Feb 19;25(4):938. doi: 10.3390/molecules25040938.
- Pierre Couvineau, Hugo De Almeida, Vincent Leroux, Bernard Roques, **Bernard Maigret**, Catherine Llorens-Cortes, Xavier Iturrioz; *Structural insight into the catalytic mechanism and inhibitor binding of aminopeptidase A*. Biochem J 13 November 2020; 477 (21): 4133–4148. doi: <https://doi.org/10.1042/BCJ20200307>
- Atanasova V, Bresso E, **Maigret B**, Martins NF, Richard-Forget F. *Computational Strategy for Minimizing Mycotoxins in Cereal Crops: Assessment of the Biological Activity of Compounds Resulting from Virtual Screening*. Molecules. 2022 Apr 16;27(8):2582. doi: 10.3390/molecules27082582.
- Mortari MR, Cunha AOS, Dos Anjos LC, Amaral HO, Quintanilha MVT, Gelfuso EA, Homem-de-Mello M, de Almeida H, Rego S, **Maigret B**, Lopes NP, Dos Santos WF. *A new class of peptides from wasp venom: a pathway to antiepileptic/neuroprotective drugs*. Brain Commun. 2023 Feb 17;5(1):fcad016. doi: 10.1093/braincomms/fcad016.