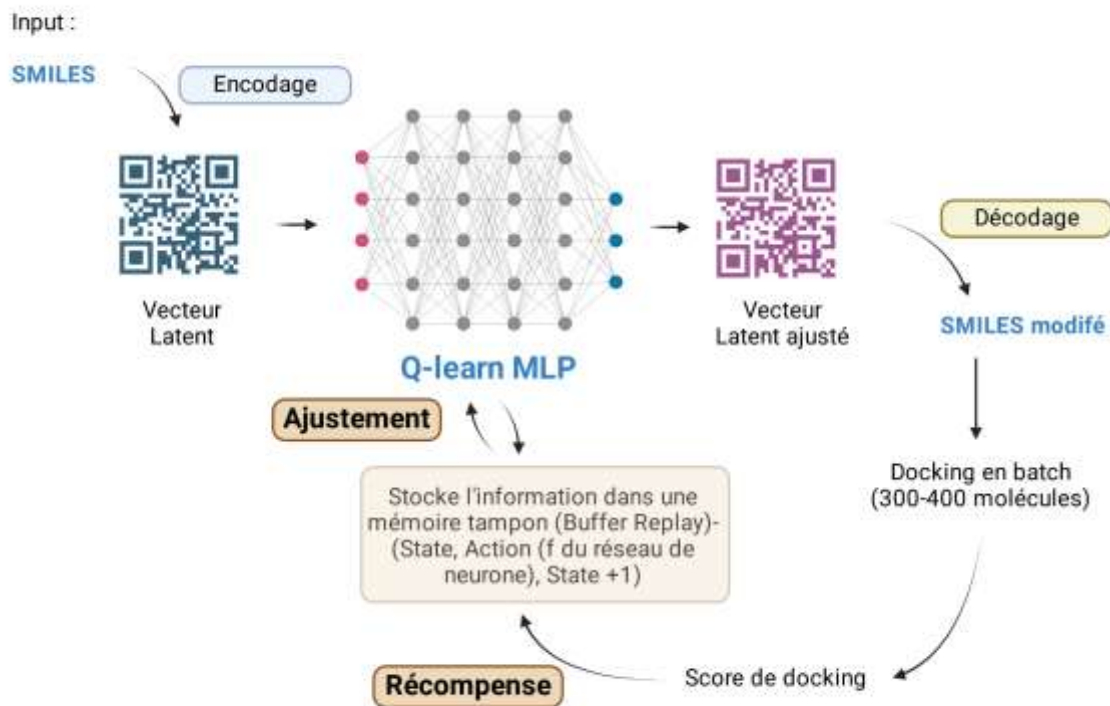


Réunion start up 02/04/2025

- Agent d'optimisation autoencodeur

Voilà le schéma global pour l'ajout d'un agent d'optimisation de type **Batch Deep-Q learning** sur lequel on doit travailler pour optimiser la génération de ligands de forte affinité :



- Outils de recherche de Off target :

○ PROZE et IA

Objectifs : améliorer l'outils existant PROZE pour la recherche systématique de ligands :

Idée : Optimiser PROZE (peut-être avec des agents IA) pour effectuer la recherche de cible off target. On part de la structure de notre protéine cible et plus particulièrement du domaine de liaison avec le ligand chimique d'intérêt et on utilise PROZE pour rechercher d'autres protéines qui ressemblent structurellement à la cible d'intérêt.

Ensuite utiliser un docker pour valider / invalider virtuellement l'interaction du ligand chimique avec notre liste off target

livrable : liste de off target potentielles

○ IA à partir de la structure

Développer un modèle de Machine Learning entraîner sur une importante base de données structurales d'interaction protéine-protéine au format PDB.

Réflexion sur le meilleur moyen de transformer l'information d'une interaction protéine-protéine ou protéin ligand en un objet (Coordonnées relatives des atomes (issues du fichier PDB) pour le site d'interaction des protéines, matrice d'adjacence atomique, fingerprint atomique des chaînes latérales de la poche d'interaction,...)

ML qui apprend à identifier un fingerprint particulier et faire une recherche systématique sur la base PDB ou sur alphafold et si on retrouve ce fingerprint à la surface d'une autre protéine qui sera un OFF TARGET POTENTIEL

- Module SAR :

Objectifs : Identifier les motifs structuraux communs aux molécules de haute affinité et inversement, les motifs défavorables.

Idée :

1. Ligands générés par autoencodeur (bruit, sd : 0.05 pour augmenter la diversité chimique) + score de docking
2. Récupérer chaque best pose (output) format PDB → orientation dans le site
3. Clustering Moléculaire :
 - a. Convertir chaque pose du ligand en vecteur ou représentation graphique ou fingerprint moléculaire
 - b. Faire du clustering par similarité
4. Analyse Statistique + SAR :
 - a. Associer à chaque cluster les scores de docking
 - b. Repérer au sein de chaque cluster des motifs récurrents ou des propriétés chimiques récurrentes
 - c. Associer ces propriétés communes à partir des BEST POSE pour identifier quels groupes chimiques interagissent avec quels parties de la protéine
 - d. Analyse des interactions (HH, π π , hydrophobe) pour les meilleurs ligands

livrable : Fournir un rapport des propriétés chimiques favorables et défavorables à modifier sur la structure de départ (le client vient avec un HIT) pour faire de l'optimisation chimique. En gros notre logiciel, donne des pistes d'amélioration virtuelle pour faire de l'optimisation chimique sur un HIT découvert par une société pharma.

