

ANR-05-CIGC-008 : Projet DOCK

Compte-rendu de la réunion du 05 Novembre 2007

OBJET : Mise en place d'une stratégie de validation	Lieu : Paris, La Villette	Version 2 du 16/11/2007
Rapporteur : Even Gaël Corrections et compléments : Loraine Brillet, Sylvaine Roy		

Participants : El-ghazali Talbi, Even Gaël (LIFL) Dragos Horvath (UGSF), Sylvaine Roy, Loraine Brillet (CEA)

L'ordre du jour était le suivant :

- ⇒ Présentation du workflow de l'interface docking@GRID avec une démonstration des premiers résultats d'intégration obtenue
- ⇒ Mise au point de la stratégie de validation avec le CEA

1.REFONTE DE L'APPLICATION POUR UNE INTÉGRATION DE LIGANDS À GRANDE ÉCHELLE – PRISE EN COMPTE DES FORMATS DE DONNEES

Gaël Even a présenté le workflow de l'application [Docking@GRID](#), la visualisation des résultats des calculs d'un Ligand, d'un Site actif et du Docking entre ces deux molécules. Cette démonstration est accessible via le web à l'adresse http://docking.futurs.inria.fr/docking_dev (login : demo@lifl.fr, passwd : demo).

Au niveau de la prise en compte des format de données, Gaël à intégré deux outils de conversion de format (Vega et Babel) qui permettent de prendre en compte une multitude de formats molécules, une liste des formats acceptés par l'application est en cours de réalisation.

Planification des tâches :

L'application actuelle permet d'ajouter les ligands (pour le calcul de conformation) uniquement un par un. Or dans le cadre de la validation du projet (et de son utilisation future) l'application doit permettre d'intégrer des sets de données de 300000 à 4 millions de Ligands.

La logique d'intégration et de visualisation de ligand doit être revue en prenant également en compte les problématiques de stockage. La base de données ZINC est une base de données de ligands qui peut être utile pour les tests.

Prochain point et présentation de la refonte : **3 décembre**

2. INTÉGRATION D'OUTILS DE VISUALISATION ET D'ANALYSE DE RÉSULTATS DANS L'APPLICATION

Dans le cadre de la validation avec le CEA, une stratégie de visualisation et d'analyse des résultats au niveau du docking doit être mise en place.

Sylvaine Roy et Lorraine Brillet ont avancé l'idée de mettre en place des outils de visualisation et d'analyse pour les résultats de Docking.

Plusieurs problématiques :

En fonction de quels critères doit-on présenter les résultats (scores...) ?

Comment les visualiser de façon optimale ?

Peut-on intégrer des outils de validations directement dans l'application (donc outils open source ou libre) ?

Tout ceci doit être fait dans le cadre de la prochaine partie :

3. MISE EN PLACE D'UN PROTOCOLE D'ÉVALUATION AUTOMATIQUE POUR LA PARTIE VALIDATION DES COMPLEXES PROTEINE-LIGANDS.

Lorraine Brillet a présenté les résultats obtenus dans le cadre du projet WISDOM concernant le docking avec FlexX sur la tubuline de mammifère et sur différentes conformations de la tubuline de *Plasmodium* générées avec Docking@Grid (résultats initialement prévus pour servir, entre autres, de benchmarks pour la validation de Docking@Grid).

→ voir présentation Power Point « Validation Docking@Grid »

Résumé :

- Les résultats de trois des quatre sites actifs étudiés sur la tubuline de mammifère ne sont pas fiables (erreurs dans la préparation des fichiers pour le docking, programme de docking non adapté au site actif)
- comparaison conformation fermée / ouverte (p_7) pour le site de l'oryzaline de la tubuline de *Plasmodium* : peu de différence dans le positionnement des actifs sur la distribution de score, interactions meilleures et plus nombreuses pour la conformation p_7, notamment pour les composés les mieux scorés
- mauvaise exploitation de la poche du site actif de l'oryzaline lors du docking en raison d'une définition mal adaptée du site actif pour le docking (en comparant avec la littérature)

Suite :

- tests biologiques sur une sélection de composés (sélection sur le score, les interactions, la disponibilité, la diversité...)
- La dynamique moléculaire de la tubuline de *Plasmodium* va être poursuivie pour comparer les conformations obtenues avec celles générées par Docking@Grid (pas d'ouverture du site de l'oryzaline lors des 5ns de simulation effectuée jusqu'ici)

Le benchmark pour la validation de Docking@Grid pourra être complété avec le programme Dock dont le CEA dispose gratuitement et un des programmes payants FlexX, Gold ou Glide dont le CEA envisage d'acquies une licence.

Pour la validation de Docking@Grid, Loraine Brillet a proposé d'utiliser les 134 complexes PDB de validation de GOLD, appelé 'original gold set validation', qui, comme son nom l'indique est un set de données utilisé pour valider le logiciel de docking GOLD, validation qui a fait l'objet d'une publication (G. Jones, P. Willett, R. C. Glen, A. R. Leach & R. Taylor, *J. Mol. Biol* **267** (1997) 727-748).

Planification des tâches :

Jean-Charles et Alexandru ont utilisé des complexes moléculaires issue du set de données ASTEX. Une étude est à réaliser pour choisir un set commun de validation entre nos deux équipes.

Le protocole doit être mis en place avec le CEA (Lorraine) et doit permettre deux choses principales :

- ⇒ Valider les complexes Protéine-Ligand de façon complète
- ⇒ Permettre une comparaison aisée avec d'autres logiciels de docking (DockGOLD ; AutoDock...)

Un travail bibliographique est nécessaire, pour connaître les processus de validation utilisés par les autres logiciels (notamment une analyse des protocoles d'évaluation de Oliver Korb). Puis une solution de validation sera proposée au CEA qui fournira son expertise.

Ensuite il faudra se poser la question du formatage des résultats de validation afin qu'ils soient exploitables pour la comparaison avec les logiciels existants.

Date d'envoi de la proposition au CEA : **30 Novembre**

4. VALIDATION BIOLOGIQUE AVEC LA CK2

S. Roy a présenté une stratégie de validation biologique possible avec le site ATP de la CK2.

La validation devait se faire avec initialement avec l'oryzaline, mais il y a deux avantages à utiliser la CK2 en parallèle:

- ⇒ Plus simple à valider (littérature abondante, données cristallographiques, connaissance de vrais négatifs, plus de facilité pour les expériences biologiques)
- ⇒ Plus connues donc résultats plus comparables.

Il faut cependant noter que le site ATP est moins flexible que celui de l'oryzaline, et permettra donc peut-être moins de mettre en évidence l'originalité de Docking@Grid..

Planification des tâches :

Possibilité de validation de la CK2 courant **printemps 2008**.

5. UTILISATION DE NAMD POUR LA DYNAMIQUE MOLÉCULAIRE (DRAGOS)

Utilisation de logiciel de dynamique moléculaire pour une éventuelle intégration dans l'algorithme de docking afin d'affiner les solutions.

Dans le cadre d'une parallélisation de NAMD, il faut se poser la question de la licence du logiciel, est-il open source ou pas ?

NAMD possède une licence spécifique éditée par l'université de L'Illinois (US).

C'est une licence non-exclusive et non-commerciale. Le code source est accessible, est libre de droit, du moment que l'utilisateur précise qu'il a intégré NAMD (ou une version modifiée) à son application.

La licence est accessible à cette adresse :

<http://www.ks.uiuc.edu/Research/namd/license.html>

6. ORGANISATION DU PROCHAIN WORKSHOP DOCKING@GRID A META 2008

EI-Ghazali Talbi a proposé d'organiser le prochain workshop de docking@GRID lors de META 2008 (Tunisie) - <http://www2.lifl.fr/META08/>

Le workshop sera ouvert et des invités acteurs du Docking Français et Européen qui se verront proposés des invitations.

E.-G. Talbi, D. Horvath et S. Roy contactent les personnes susceptibles d'être intéressées dans leur domaine respectif.