

Compte-rendu du workshop du 11 au 14 mars 2007

OBJET : Bilan des développements distincts et planification des tâches restantes.

Lieu : Dinant, Belgique.

Version 1 du 14 mars 2007

Rapporteur : Cyrielle BOUTROUE.

Correcteur : Jean-Charles BOISSON.

Participants : Dragos HORVATH, Benjamin PARENT, El-Ghazali TALBI, Nouredine Melab, Alexandru TANTAR, Jean-Charles BOISSON et Cyrielle BOUTROUE.

Programme :

- Permettre à chaque équipe de présenter les travaux effectués jusqu'à présent
- Faire un point sur la manière de regrouper les réalisations effectuées
- Planifier les tâches restantes.

1) Présentation des travaux

a) Modélisation des molécules (par B. Parent)

Cette introduction a donné le contexte et les notions utiles à la manipulation des molécules. Benjamin a présenté les différents types de docking possibles puis s'est attardé sur les contributions énergétiques liées au repliement des molécules avec les densités de probabilités et les champs de force spécifiques à l'algorithme.

Les différents problèmes mis en avant sont au nombre de trois : l'exploration, l'entropie et l'expertise expérimentale. L'idée sous-jacente retenue est d'effectuer une comparaison en terme de spectre.

b) Les modèles pour le problème de docking (par J-C. Boisson)

Jean-Charles a exposé les différentes possibilités de modélisation pour le docking. Même s'il existe deux grandes familles de représentation (liées au problème de prédiction de structure de protéines et au repliement de protéines), seul la représentation atomique permet de prendre en compte les champs de force liés aux différents modèles énergétiques existants. Au vu de différentes références, Jean-Charles va utiliser des critères d'énergie (modèle existant...), d'enfouissement (surface ...) et de robustesse (entropie ???) afin de mettre en place un modèle tri-objectif.

c) Les algorithmes évolutionnaires pour le problème de la prédiction de structures de la protéine et du docking moléculaire (par A. Tantar)

Après avoir présenté un premier descriptif sur la modélisation et l'analyse de la complexité (avec la fonction d'énergie, la modélisation des conformations et un panorama des méthodes et des algorithmes existants), Alexandru nous a expliqué la mise en oeuvre d'une méthode parallèle hybridée (algorithme générique Lamarckien) déployée grâce à ParadisEO-G (Globus). En conclusion, l'utilisation de l'hybridation avec le recuit simulé et les algorithmes génétiques seront les méthodes employées.

d) ParadisEO sur Globus (par N. Melab)

Nouredine a commencé par une explication des outils Globus pour la gestion de problèmes techniques pour le développement d'outils, de services et d'application pour les grilles. Afin de communiquer avec la grille en utilisant les services présentés précédemment, la bibliothèque de communication MPICH est nécessaire. Nouredine a donc dévoilé l'architecture de cette bibliothèque et a poursuivi sur la construction de l'organisation virtuelle Globus et a fini sur l'utilisation de ParadisEO sur ce système (ParadisEO-G).

e) Interface web de l'application (par C. Boutroue)

Cyrielle a commencé par résumer le service demandé puis a expliqué la modularité de l'interface avec l'implémentation des trois programmes pour le docking.

Ayant mis en place un prototype élaboré, chaque partie de l'interface a pu être présentée graphiquement d'un point de vue utilisateur et administrateur. En dernier point, la validation par l'équipe du CEA a été abordée.

f) Algorithme pour le folding (par D. Horvath)

Dragos a proposé un aperçu de son algorithme en « direct » en l'expliquant étapes par étapes (algorithme génétique, mise en place sur la grille de calcul, obtention des résultats, sélection des bonnes solutions, affinage de celles-ci si nécessaire et visualisation des résultats). Le problème majeur de cette implémentation a été illustré par le cas du zipper : difficulté de trouver la solution d'énergie correcte sans intensification immodérée du problème (problème du coût de calcul).

2) Regroupement des réalisations

La seconde partie du workshop a été principalement tournée sur la manière de « repenser » le code. Un gros problème d'organisation des classes a été levé. A l'heure actuelle, il est difficile de l'utiliser pour développer un modèle multi-objectif. A la fin de la réunion, un modèle simplifié réunissant efficacité et accessibilité a conforté les deux équipes.

3) **Planification des tâches**

a) **L'interface graphique**

L'interface graphique doit être réalisée en 3 mois d'exercice (ce qui nous amène à fin juin). L'intégration des différents algorithmes se faisant par threads, il est tout à fait possible de mettre à jour cette interface avec de nouveaux algorithmes par la suite. Afin de développer correctement cette version web on-line, des programmes non fonctionnels mais possédant des entrées/sorties adéquates doivent être fournis à Cyrielle (cf. Alexandru).

b) **La remodellisation du code**

Les quatre intéressés par ce thème sont Alexandru, Jean-Charles, Benjamin et Dragos. La première phase de reprogrammation (avec ajout de commentaires) de ce code va se faire très rapidement (environ une journée) auquel une période de test supplémentaire va être nécessaire. Cela permettra à Dragos et Benjamin de continuer de travailler sur leur approche sans devoir connaître les modifications de conception ou d'implémentation qui pourront être pratiquées dans le cœur du code. Ensuite, un travail important de conception et de programmation devra être réalisé par Alexandru et Jean-Charles pour notamment permettre la réutilisation du code telle une « boîte noire » qui peut cependant être étendue (plus ou moins) facilement.

c) **Parallélisation avec ParadisEO**

Alexandru est chargé de mettre en place l'hybridation algorithme génétique / recuit simulé et de le déployer sur la grille de calcul.

d) **Modèle multi-objectif.**

Jean-Charles va utiliser les critères d'enfouissement et de robustesse en plus d'un critère énergétique pour essayer de proposer un nouveau modèle tri-objectif. Une fois le code plus facilement portable sur ParadisEO, il va pouvoir implémenter ce modèle.

e) **Modélisation**

Un travail d'amélioration du calcul des contributions énergétiques au repliement des molécules sera continué par Benjamin.

Quant à Dragos, il est chargé d'affiner son algorithme (ajustement des contraintes) afin d'obtenir des résultats convaincants pour des molécules telles que le zipper.