

Christine CARAPITO¹, Jérôme PANSANEL², Patrick GUTERL¹, Alexandre BUREL¹, Fabrice BERTILE¹, Stéphane GENAUD³, Alain VAN DORSSELAER¹, Christelle ROY²

¹ Laboratoire de Spectrométrie de Masse BioOrganique, DSA, IPHC, UMR7178, CNRS, Université de Strasbourg, 25 rue Becquerel 67087 Strasbourg, France

² Département Recherches Subatomiques, DRS, IPHC, UMR7178, CNRS, Université de Strasbourg, 23 rue du Loess 67037 Strasbourg, France

³ Laboratoire de Sciences de l'Image, de l'Informatique et de la Télédétection, équipe ICPS (LSIIT/ICPS), Université de Strasbourg, CNRS UMR7005, Strasbourg, France

Introduction

Ces 20 dernières années, la spectrométrie de masse (MS) a joué un rôle majeur pour l'analyse protéomique et est aujourd'hui considérée comme l'une des méthodologies les plus prometteuses permettant des progrès dans la compréhension de nombreux processus biologiques. Les importants progrès techniques en MS ont conduit au développement d'instruments générant des données MS/MS de plus en plus volumineuses (du fait d'une grande rapidité d'acquisition des spectres de fragmentation). Par ailleurs, la soumission des résultats d'identification de protéines à partir de ces données MS/MS est de plus en plus réglementée par les journaux du domaine qui recommandent l'utilisation d'algorithmes transparents (open-source) et multiples si possible.

Dans ce contexte, afin de répondre au besoin croissant de puissance de calcul nécessaire pour l'analyse protéomique, une interface permettant d'effectuer l'interprétation des données MS/MS avec un logiciel libre (OMSSA : Open Mass Spectrometry Search Algorithm¹) sur la grille de calcul a été développée.

Méthode

Les différentes étapes à suivre pour le lancement d'une identification de données MS/MS sur la grille sont les suivantes :

1. Préparation de la banque de séquences protéiques
2. Préparation de fichiers de sous-ensembles de spectres MS/MS à partir des fichiers de données MS/MS d'origine
3. Soumission des calculs sur la grille avec l'outil JJS² (via la VO Biomed ou la VO régionale)
4. Récupération des résultats et post-traitement

Lors de l'exécution des calculs sur la grille, les éventuelles erreurs sont gérées par JJS permettant ainsi d'assurer l'intégrité des données (tous les fichiers sont récupérés et correctement traités).

Résultats

Afin d'illustrer l'amélioration des performances obtenue en calculant sur la grille de calcul, quatre sets de données caractéristiques d'expériences protéomiques courantes ont été utilisés. Les gains de temps optimisés sont décrits dans la figure 1. Les paramètres clés ayant permis d'accélérer les recherches d'un facteur 80 sont les suivants :

- Nombre de soumissions par minute
- Nombre de spectres MS/MS par fichier
- Temps de calcul minimal autorisé sur un site
- Nombre minimal de CPU sur un site

Un second bénéfice de la grille réside dans le fait que l'ensemble des recherches peut être lancé en simultanément. Comme illustré ci-dessous, les multiples utilisateurs peuvent donc bénéficier des améliorations de performances optimisées sans passer par une file d'attente (figures 2 et 3).

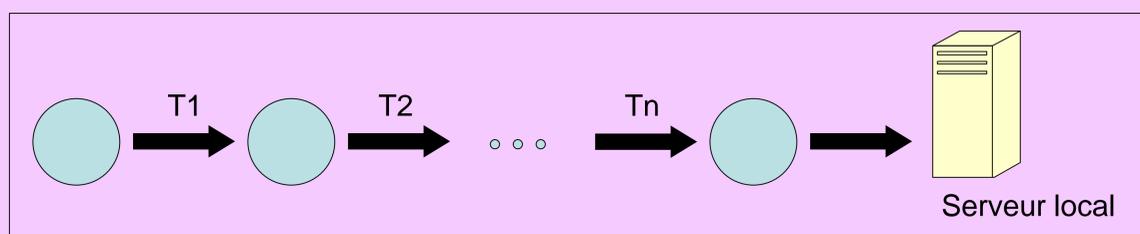


Figure 2 : File d'attente des utilisateurs sans la grille

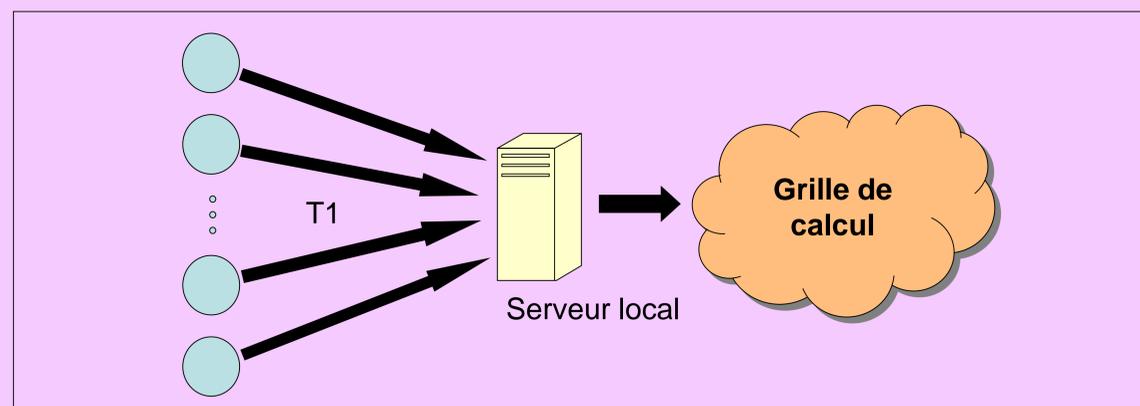


Figure 3 : Avec la grille, les utilisateurs accèdent directement au serveur

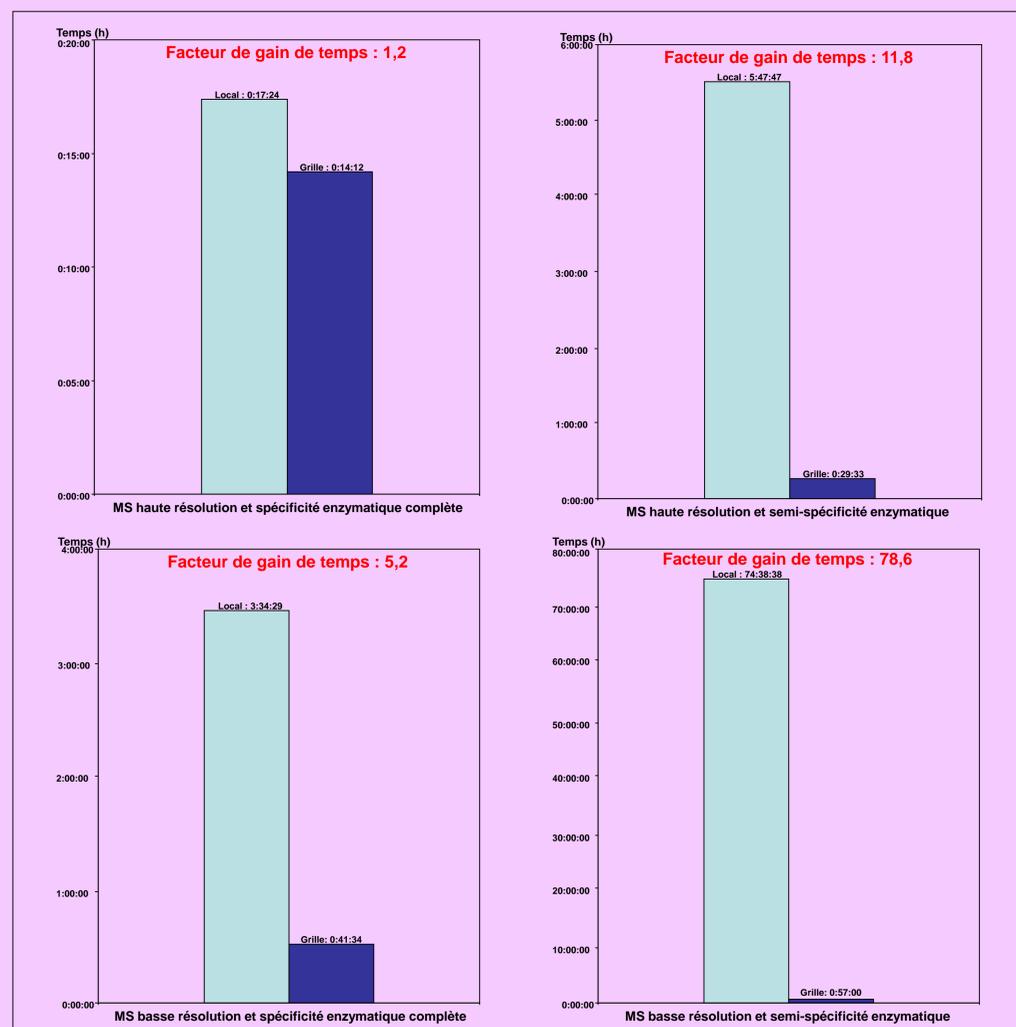


Figure 1 : Facteurs de gains de temps optimisés pour chaque type d'expérience

Conclusion

Le développement de l'interface pour OMSSA sur la grille de calcul permet de répondre aux importants besoins de puissance de calcul non accessibles à ce jour dans les laboratoires de protéomique. Non seulement les performances sur une expérience sont nettement améliorées mais la puissance disponible nous a permis d'ouvrir le site à une large communauté de protéomistes tout en s'affranchissant de la notion de file d'attente.