

GénoProtéome

(Projet de Service 2004)



La protéomique : un outil d'annotation des génomes

<http://www-genoproteo.inrialpes.fr/>

CONTEXTE

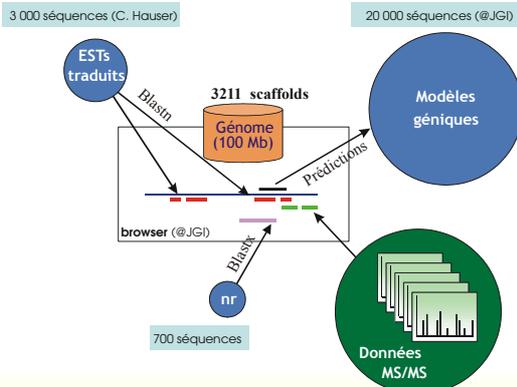
L'analyse "massive" des protéines présentes dans une cellule constitue un moyen d'identifier les gènes codant pour ces protéines, à condition de disposer des outils nécessaires à cette approche. En confrontant ces informations issues de données expérimentales avec les prédictions de régions codantes obtenues par des méthodes d'analyse « in silico », il devient possible d'infirmar ou de confirmer les prédictions de ces régions. La protéomique devient alors un outil pour l'annotation des génomes qui apporte aux méthodes de prédictions une dimension expérimentale supplémentaire.

OBJECTIFS

Notre objectif est de concevoir un nouveau module de la plate-forme Genostar dédié à la représentation des données expérimentales de protéomique et d'y intégrer les composants Taggor et PepMap qui vont permettre de : (1) manipuler et d'analyser de manière intégrée des données expérimentales de MS/MS, et (2) rattacher les données expérimentales à la séquence génomique de façon à prolonger le processus d'annotation des génomes. La communication et le partage des données entre modules qu'offre la plate-forme GenoStar permettra de confronter ces résultats aux données d'annotation issues de GenoAnnot.

CONTEXTE de VALIDATION

L'annotation du génome de l'algue verte *Chlamydomonas reinhardtii* avec les données de protéomique de type MS/MS constitue le cadre de validation de l'outil développé.



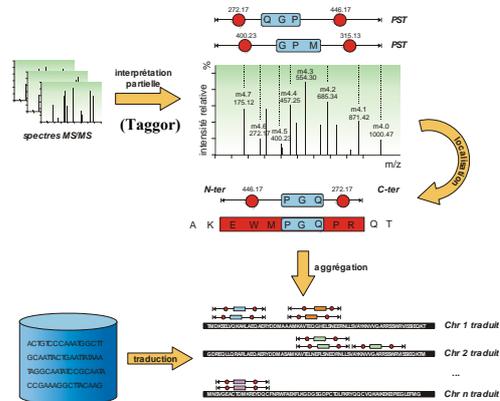
LABORATOIRES IMPLIQUES

Laboratoire de Chimie des Protéines - ERM 0201 CEA/INSERM
Laboratoire de Biologie, Informatique et Mathématique - CEA
Projet Helix - INRIA Rhône-Alpes

MOTS CLES

Protéomique, Spectrométrie de masse, Annotation des génomes.

PepLine

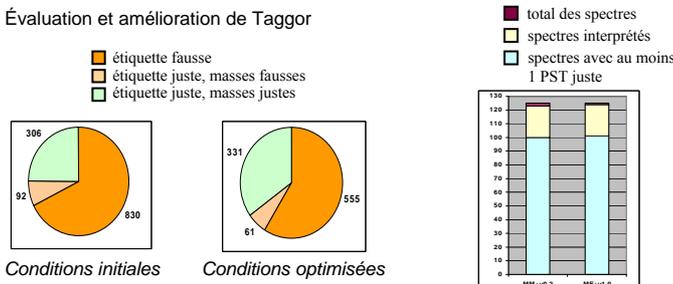


PepLine, extrait des étiquettes peptidiques (PST) à partir des spectres MS/MS et les localise sur des séquences d'ADN génomique traduites. Un PST est constitué d'une courte séquence de trois acides aminés, et de deux masses latérales correspondant aux masses N et C terminales du fragment tryptique. La partie syntaxique correspond à la région du spectre aisément interprétable et les deux masses flanquantes représentent les zones du spectre non interprétées. Les PST sont ensuite localisés sur les traductions de la séquence génomique.

GénoProtéome

RESULTATS

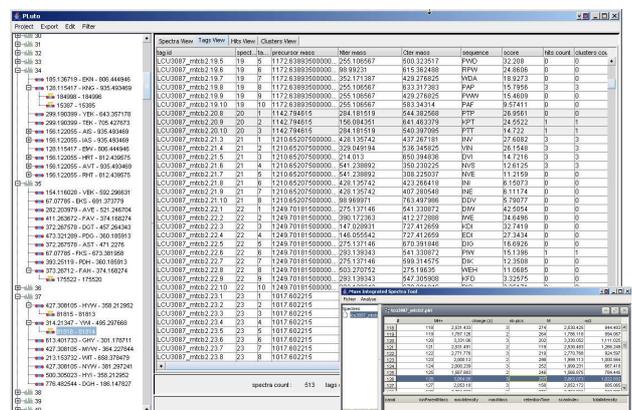
Évaluation et amélioration de Taggor



Les étiquettes peptidiques (PST) générées sont comparées aux séquences des protéines dont les spectres sont issus pour évaluer leur exactitude. Cette évaluation est répétée en faisant varier les paramètres de génération des spectres et des PST. L'analyse des causes d'erreur sur les PST incorrects a permis de corriger et d'améliorer cette première étape du processus d'annotation (Taggor).

CALENDRIER

- 2004 Octobre – Évaluation et amélioration des étiquettes peptidiques.
- 2005 Janvier – Développement d'éditeurs spécialisés.
- 2005 Septembre – Recrutement d'un développeur Génostar



Développement d'éditeurs spécialisés pour les données issues des résultats de PepLine : spectres, PST, localisation des PST sur la séquence et agrégation de PST.

