

Prédiction des N-terminaux mature des protéines mitochondriales et chloroplastiques (CDD 12 mois pour un IR en bio-informatique avec expérience)

Contexte scientifique

Les organites liés à la conversion énergétique (ECO) tels que les mitochondries et les chloroplastes ont perdu, au fil de l'évolution, une grande partie de leur matériel génétique qui est maintenant intégré au génome nucléaire. La majorité des protéines de ces organelles provient donc du cytosol ou a lieu la synthèse du génome nucléaire ; certaines protéines sont redirigées spécifiquement aux organelles grâce à l'information située au niveau du peptide N-terminal dit de transit (TP). Le TP est généralement excisé durant l'importation et parfois acétylé au niveau du néo-N-terminal [1, 2]. Le site de clivage du TP n'a été caractérisé expérimentalement que pour un petit nombre de protéines. Ainsi, ces sites de clivages sont généralement déterminés par des outils de prédiction tel que TargetP [3] ou ChloroP [4]. Malheureusement, il est apparu que ces prédictions sont peu fiables et ne tiennent aucun compte de la modification du néo-N-terminal [1].

Le principal biais des outils existant vient de l'échantillon de données restreint à quelques protéines seulement ne permettant pas d'avoir une vue globale du phénomène. Nous avons donc développé une technique permettant de cibler préférentiellement les N-terminaux des protéines matures ce qui nous a permis de caractériser expérimentalement des milliers de N-terminaux de protéines relocalisées dans les ECO. Avec ces informations, il est maintenant envisageable d'exploiter ces résultats expérimentaux pour développer de nouveaux outils de prédictions basés sur ce large échantillonnage. Ces données expérimentales seront utilisées pour le développement de plusieurs nouveaux outils de prédiction ciblant en particulier la localisation subcellulaire des protéines (chloroplaste, mitochondrie, ...), le site de clivage du TP, les modifications associées au neo-N-terminal... Finalement il sera nécessaire de déterminer la puissance prédictive de ces nouveaux outils et de comparer leur valeur prédictive par rapport à ceux existant.

Nous recherchons un ingénieur de recherche (IR) en bio-informatique (avec expérience dans le domaine) dans le cadre d'un projet collaboratif financé par une ANR-Blanc (ANR-13-BSV6-0004). Son projet consistera principalement, dans un premier temps, à collecter et d'unifier les données des différents partenaires puis de les valoriser au travers le développement de nouveaux outils de prédictions. Cela nécessitera la mise en place d'une banque de données centralisée (mise à disposition de la communauté) et l'utilisation de ces données pour le développement d'outils de prédiction en utilisant la méthode la plus performante (NN, SVM,...).

Compétences requises

- Master ou Ph D. en bio-informatique avec expérience confirmée en protéomique;
- Connaissance d'un langage de programmation et en particulier le Python;
- Expertise concernant le développement d'outils de prédiction pouvant utiliser des systèmes à réseaux de neurones ou SVM...
- Bonne connaissance/interaction avec la biologie et les données de protéomique
- Bonne communication en français ou anglais
- Connaissance en statistique, un plus.

Contact :

Willy Bienvenu (willy.bienvenu@i2bc.paris-saclay.fr)

site web : <http://www.i2bc.paris-saclay.fr/spip.php?article130>

Équipe d'accueil : Maturation et destinée cellulaire des protéines

Durée du contrat : 12 mois

Date de début : 01/09/2016

Localisation : **Protein Maturation, Cell fate and Therapeutics** group, Institut de Biologie Intégrative de la Cellule, Département « Biologie des génomes », Bât 21, CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette

Candidature :

Les candidats intéressés par cette offre devront faire parvenir une lettre de motivation et un CV à W. Bienvenu ainsi que les coordonnées de trois références.

Prediction tools development for mature proteins N-terminal characterisation in mitochondria and chloroplast (12-month position for an experienced bio-informatician in proteomics).

Energy conversion organelles (ECO) such as mitochondria or chloroplasts lost much of their genetic material over the evolution. The majority of the genetic material of these organelles is now integrated into the nuclear genome. Thus, Most of the organelles proteins are coming from the cytosol where the nuclear genome is translated then, they are relocated specifically to the relevant organelles based on the transit peptide coded information (TP) that is usually excised during the import step and then, sometimes acetylated at the neo-N-terminus [1, 2]. TP cleavage site has been characterized experimentally for a small number of proteins but these positions are generally obtained using prediction tools such as TargetP [3] or ChloroP [4]. Unfortunately, it appeared that they are unreliable and they do not consider the neo-N-terminus modifications [1].

Poor predictions of these existing tools are primarily due to the limited experimental dataset used during their development. Thus, we developed a technique to preferentially target and characterize the mature proteins N-termini which allowed us to experimentally characterize thousands of protein N-terminus relocated to the ECO. With such information, it is now possible to valorise these experimental results through the development of new prediction tools. These experimental data will be used for the development of several new prediction tools targeting especially the subcellular localization of proteins (chloroplasts, mitochondria...), cleavage site of the TP, post translational N-terminal modifications... Finally, it will be necessary to determine the predictive power of these new tools compared to the existing ones.

We are looking for a computer programming engineer (with experience in proteomics field) as part of a collaborative project (LSMBO, S. Cianferani, Strasbourg; CEA, T. Rabilloud, Grenoble; I2BC, C. Giglione, Gif/Yvette) funded by ANR-Blanc (ANR -13- BSV6-0004). This work will require first to collect and unify data from different partners (through a centralized database, available to the community) and to valorize them through the development of new prediction tool using the most efficient method (NN, SVM, ...).

Required Skills

- Master or PhD in bioinformatics with confirmed experienced in proteomics;
- Solid experience in programming language (especially Python);
- Expertise on the development of predictive tools using, for example, neural networks or SVM...
- Good knowledge in biology and proteomics;
- Good communication skills in French or English;
- Knowledge in statistics, a plus.

Contact: Willy Bienvenu (willy.bienvenu@i2bc.paris-saclay.fr)

Host team: Protein Maturation, Cell fate and Therapeutics;

site web : <http://www.i2bc.paris-saclay.fr/spip.php?article130>

Duration: 12 mois

Starting date: 01/09/2016

Address: Institut de Biologie Intégrative de la Cellule, Département « Biologie des génomes », Bât 21, CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette

Applicants:

Interested candidates should send a cover letter and *curriculum vitae* to W. Bienvenu and contact information of three referees.

- [1] W.V. Bienvenut, D. Sumpton, A. Martinez, S. Lilla, C. Espagne, T. Meinel, C. Giglione, Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features, *Mol Cell Proteomics*, 11 (2012) M111 015131.
- [2] B. Zybaylov, H. Rutschow, G. Friso, A. Rudella, O. Emanuelsson, Q. Sun, K.J. van Wijk, Sorting signals, N-terminal modifications and abundance of the chloroplast proteome, *PLoS One*, 3 (2008) e1994.
- [3] O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP and related tools, *Nat Protoc*, 2 (2007) 953-971.
- [4] O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites, *Protein Sci*, 8 (1999) 978-984.