

# Inférence de réseaux de filiation de peptides à partir de données peptidomiques

Ousmane Suwareh<sup>1</sup>, Françoise Nau<sup>1</sup>, David Causeur<sup>2</sup>

<sup>1</sup>STLO, INRAE, Institut Agro, Rennes, France

<sup>2</sup>IRMAR, UMR6625, CNRS, Institut Agro, Rennes, France

E-mail de correspondance: [ousmane.suwareh@agrocampus-ouest.fr](mailto:ousmane.suwareh@agrocampus-ouest.fr)

**Résumé:** La digestion est un processus dynamique complexe impliquant plusieurs protéases dont l'action consiste à décomposer les protéines en fragments de plus en plus petits appelés peptides. Cependant, le mode d'action de toutes les protéases, notamment de la pepsine, n'est pas totalement compris. L'inférence sur le mécanisme de génération des peptides est rendue possible par des modèles de digestion *in vitro*, combinés à l'identification par spectrométrie de masse des peptides libérés à différents temps. Un modèle de régression logistique non paramétrique pour la probabilité qu'une protéase clive un peptide dont le profil de variables explicatives physico-chimiques est connu pourrait expliquer la spécificité de l'action protéasique. Cependant, la traçabilité du peptide parent dont le clivage a produit un peptide enfant donné est souvent incertaine et soulève une ambiguïté dans l'étiquetage de la variable réponse binaire de clivage. Nous proposons de remédier à ce problème d'étiquetage en utilisant un algorithme d'estimation de type espérance-maximisation *ad hoc*, tenant compte d'une part de l'observation incomplète du réseau de parenté des peptides identifiés à deux points de temps successifs, et d'autre part de leurs profils de variables explicatives physicochimiques. Le gain en termes d'ajustement de l'estimation est démontré via un dispositif de digestion *in vitro* avec pour objectif d'étudier l'action dynamique de la pepsine. Ces observations sont confirmées par des simulations basées sur des données réelles, en incluant la simulation d'un réseau de parenté du même type.

**Mots clés:** Algorithme EM; Données peptidomiques; Régression logistique non paramétrique; Pepsine; Protéolyse.

Suwareh, O., Causeur, D., Jardin, J., Briard-Bion, V., Le Feunteun, S., Pezenec, S., Nau, F. (2021). Statistical modeling of *in vitro* pepsin specificity. *Food Chemistry* 362, 130098.  
Suwareh, O., Nau, F., Causeur, D. (2022). An Expectation-Maximization algorithm for modeling peptide filiation using peptidomic data. submitted in *Biometrics Practice*, ID :BIOM2022515P.