

# THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641  
*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*  
Spécialité : *Bioinformatique*

Par

**Emile Benoist**

**Aspects algorithmiques de l'identification de peptides et de l'inférence de protéines par spectrométrie de masse**

Thèse présentée et soutenue à Nantes, le 05 décembre 2024  
Unité de recherche : UMR 6004 - LS2N

## Rapporteurs avant soutenance :

Christine CARAPITO Directrice de recherche l'Université de Strasbourg  
Hélène TOUZET Directrice de recherche l'Université de Lille

## Composition du Jury :

Examinat.eur.rice.s :	Annie CHATEAU	Professeure des universités	Université de Montpellier
	Thomas SCHIEX	Directeur de recherche	Centre Inrae Occitanie-Toulouse
	Christine CARAPITO	Directrice de recherche	l'Université de Strasbourg
	Hélène TOUZET	Directrice de recherche	l'Université de Lille
Dir. de thèse :	Guillaume FERTIN	Professeur des universités	Nantes Université
Enc. de thèse :	Géraldine JEAN	Maître de conférences	Nantes Université
	Dominique TESSIER	Ingénieure de recherche	Inrae de Nantes

## Invité(s) :

Hélène ROGNIAUX Ingénieure de recherche Inrae de Nantes

---

**Titre :** Aspects algorithmiques de l'identification de peptides et de l'inférence de protéines par spectrométrie de masse

**Mots-clés :** Protéomique, Peptidomique, Inférence de protéines, Identification des peptides, Spectrométrie de masse, Complexité algorithmique

**Résumé :** Durant ce doctorat, nous nous sommes intéressés à deux problèmes que l'on rencontre en protéomique lorsque l'on souhaite identifier les protéines d'un échantillon biologique par spectrométrie de masse : l'identification des peptides (morceaux de protéines) et l'inférence de protéines. Nous avons dans un premier temps proposé plusieurs modélisations de ces deux problèmes, et cela dans le but de les étudier sous l'angle de la complexité algorithmique. Ensuite, nous avons réunis l'application de ces deux pro-

blèmes dans un seul et même modèle global, permettant de limiter la propagation d'erreurs. Nous avons cette fois-ci étudié ce modèle global d'un point de vue pratique, en le confrontant à des jeux de données représentatifs de situations réelles. Enfin, nous avons implémenté un nouveau logiciel appelé SpecPeptidOMS, capable d'identifier des peptides modifiés dans le cadre de la peptidomique, cadre dans lequel l'objectif est d'identifier les peptides d'un échantillon biologique.

---

**Title:** Algorithmic aspects of peptides identification and proteins inference by mass spectrometry

**Keywords:** Proteomics, Peptidomics, Proteins inference, Peptides identification, Mass spectrometry, Algorithmic complexity

**Abstract:** During this PhD, we focused on two problems encountered in proteomics when identifying proteins in a biological sample by mass spectrometry: peptides identification (protein pieces) and proteins inference. We first proposed several models for these two problems, with the aim of studying them from the point of view of algorithmic complexity. Then, we combined the application of these two problems in a single global model, en-

abling us to limit the propagation of errors. This time, we studied this global model from a practical point of view, confronting it with datasets representative of real-life situations. Finally, we implemented a new software package called SpecPeptidOMS, capable of identifying modified peptides in the context of peptidomics, where the aim is to identify peptides in a biological sample.