

## Introduction

La lignine, un coproduit de l'industrie papetière, est le deuxième biopolymère le plus abondant sur Terre après la cellulose. Actuellement, cette lignine résiduelle est majoritairement brûlée pour produire de la chaleur. Étant donnée la complexité de ces molécules, on peut en espérer une meilleure valorisation, mais cette amélioration passe par la connaissance de sa structure moléculaire.

## Problématique

La structure moléculaire de la lignine est difficile à connaître pour plusieurs raisons. D'abord, elle est produite dans les plantes par l'assemblage combinatoire d'unités fondamentales (les monomères H, G et S), créant des polymères à structure variable. On connaît cependant les règles de base de l'assemblage de ces monomères.

Ensuite, la structure de la lignine est généralement altérée par les procédés d'extraction, rendant compliquée l'analyse de ses propriétés dans son état naturel.

Enfin, les méthodes d'analyse ne donnent qu'un portrait partiel de la lignine à analyser. Plusieurs méthodes d'analyses corrélées entre elles donneraient une meilleure information sur sa structure complexe.

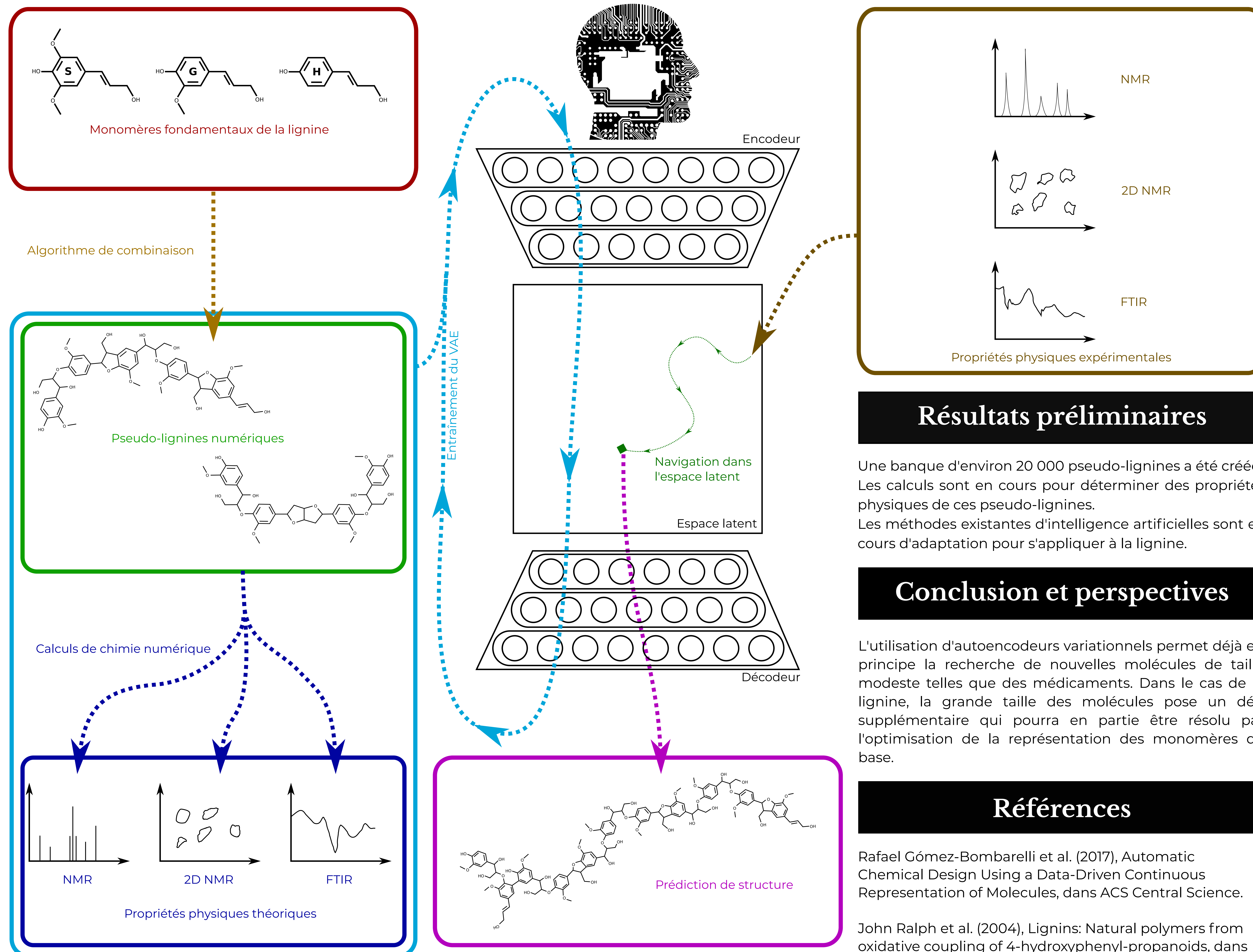
## Méthodologie

Pour déterminer la structure moléculaire d'une lignine à partir d'un ou de plusieurs résultats expérimentaux, il est proposé de créer numériquement une multitude de pseudo-lignines et de calculer théoriquement les propriétés physiques de chacune. Cette banque de pseudo-lignines et de propriétés théoriques servira à l'entraînement d'une intelligence artificielle (IA).

On choisit une IA de type autoencodeur variationnel (VAE). Un VAE peut encoder dans un «espace latent» multidimensionnel les structures moléculaires qu'on lui fournit pendant son entraînement, en les classant en fonction des propriétés théoriques.

Après l'entraînement, le VAE peut prendre en entrée une série de propriétés expérimentales et naviguer dans l'espace latent pour se diriger vers une structure moléculaire qui reproduit ces résultats. La nature probabiliste du VAE fait en sorte qu'il peut générer de nouvelles structures moléculaires en interpolant entre les structures des pseudo-lignines utilisées pendant l'entraînement, pour finalement suggérer une structure moléculaire inédite à la lignine étudiée.

## Illustration



## Résultats préliminaires

Une banque d'environ 20 000 pseudo-lignines a été créée. Les calculs sont en cours pour déterminer des propriétés physiques de ces pseudo-lignines. Les méthodes existantes d'intelligence artificielle sont en cours d'adaptation pour s'appliquer à la lignine.

## Conclusion et perspectives

L'utilisation d'autoencodeurs variationnels permet déjà en principe la recherche de nouvelles molécules de taille modeste telles que des médicaments. Dans le cas de la lignine, la grande taille des molécules pose un défi supplémentaire qui pourra en partie être résolu par l'optimisation de la représentation des monomères de base.

## Références

- Rafael Gómez-Bombarelli et al. (2017), Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, dans ACS Central Science.
- John Ralph et al. (2004), Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids, dans Phytochemistry Reviews.