

Projeto de Iniciação Científica - PIBIC/CNPEM

Título: **Aprendizado de máquina para determinação de entalpias de reação em redes de reações químicas de lignina**

Pesquisador Responsável: **Carlos Driemeier**

Unidade do CNPEM: **LNBR**

Introdução

A biomassa lignocelulósica é a mais abundante fonte de carbono renovável disponível na Terra. Esse tipo de biomassa constitui a parte estrutural e não comestível das plantas, sendo formada majoritariamente por carboidratos (celulose e hemiceluloses, tipicamente 60-70% em massa) e lignina (cerca de 20-30%). A transformação industrial dos carboidratos tem sido historicamente bem-sucedida, como demonstra a indústria de papel e celulose e, mais recentemente, a indústria de etanol celulósico. No entanto, o mesmo não ocorre com a lignina, que vem sendo utilizada quase exclusivamente como combustível de caldeira. Em busca de maior valorização da lignina, estudos recentes têm buscado transformar a lignina em biocombustíveis líquidos e bioquímicos aromáticos, aproveitando a baixa oxigenação e caráter aromático da macromolécula [1]. Esse potencial é particularmente interessante se associado à lignina residual do etanol celulósico, que é um novo tipo de lignina disponível em escala industrial [2]. Esse tipo de processo para aproveitamento de lignina é apresentado na Figura 1.

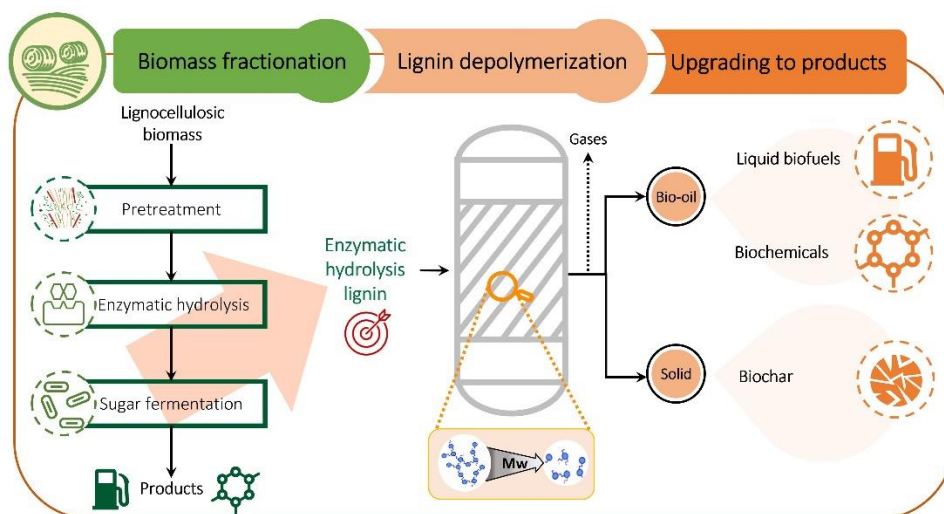


Figura 1: Representação esquemática de processo no qual aos carboidratos da biomassa lignocelulósica são direcionados para produtos de fermentação (e.g., etanol celulósico) e a lignina residual é submetida a processos de despolimerização e beneficiamento (*upgrading*) para obtenção de biocombustíveis líquidos, bioquímicos e biocarvões. Adaptado de [2].

Um desafio dos processos de despolimerização da lignina e beneficiamento do seu bio-óleo advém da complexidade do meio reacional, com grande diversidade de compostos químicos ($\sim 10^2$ - 10^3 compostos) e inúmeras reações químicas ocorrendo de forma paralela e sequencial. Esse tipo de problema tem sido abordado através de redes de reações químicas com técnicas de aprendizado de máquina [3,4]. No entanto, ainda não existe aplicação dessa abordagem para redes de reações envolvendo lignina e este projeto busca avançar o conhecimento nessa direção.

Estado da arte

Está havendo um rápido avanço no uso de redes de reações químicas associado a aprendizado de máquina para descrição de fenômenos envolvendo grande número de compostos e reações [3,4]. Esses desenvolvimentos empregam bases de dados criados com cálculos de química quântica contendo dezenas a centenas de milhares de compostos e reações [5,6,7]. No entanto, essas bases de dados são limitadas a compostos pequenos (*e.g.*, de até 9 átomos pesados) e não contém o conjunto de compostos necessários para modelar reações de despolimerização de lignina e beneficiamento do seu bio-óleo.

Objetivos

O objetivo geral do projeto é avançar no uso de técnicas de aprendizado de máquina para descrição de redes de reações químicas envolvendo lignina. Para alcançar esse objetivo geral, o projeto buscará os seguintes objetivos específicos:

- i)* Prospecção de algoritmos que permitam aprender a partir de grandes bases de compostos e reações [5,6,7] para posterior transferência do aprendizado para bases de fragmentos de lignina [8].
- ii)* Implementação em linguagem Python dos algoritmos selecionados para predição de entalpias de reação envolvendo fragmentos de lignina.
- iii)* Testes de desempenho dos algoritmos de predição de entalpias de reação de fragmentos de lignina, buscando atingir acurácia química (~ 1 kcal/mol).

Metodologia

Os desenvolvimentos serão realizados em linguagem Python, usando as bibliotecas RDKit [9] para manipulação de informação química, e scikit learn [10] para aprendizado de máquina, análise e mineração de dados. Serão consultadas bases de compostos e reações químicas [5,6,7,8]. O projeto seguirá as seguintes etapas: (i)

preparação e armazenamento de dados; (ii) análise de dados e inferência estatística; (iii) modelagem preditiva; e (iv) validação de modelo e avaliação do seu desempenho.

Referências

- [1] Schutyser W, Renders T, Van Den Bosch S, et al (2018) Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chem Soc Rev* 47:852–908.
- [2] Menezes FF, Nascimento VM, Gomes GR, Strauss M, Junqueira TL, Driemeier C (2023) Depolymerization of enzymatic hydrolysis lignin: Review of technologies and opportunities for research. *Fuel* 342:127796.
- [3] Wen M, Spotte-Smith EWC, Blau SM, et al (2023) Chemical reaction networks and opportunities for machine learning. *Nat Comput Sci* 3:12–24.
- [4] Stocker S, Csányi G, Reuter K, Margraf JT (2020) Machine learning in chemical reaction space. *Nat Commun* 11:1–11.
- [5] Ramakrishnan R, Dral PO, Rupp M, Von Lilienfeld OA (2014) Quantum chemistry structures and properties of 134 kilo molecules. *Sci Data* 1:1–7.
- [6] St. John PC, Guan Y, Kim Y, et al (2020) Quantum chemical calculations for over 200,000 organic radical species and 40,000 associated closed-shell molecules. *Sci Data* 7:1–6.
- [7] Grambow CA, Pattanaik L, Green WH (2020) Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Sci Data* 7:1–8.
- [8] Colombari FM, Nascimento VM, Liu YL, Rocha GJM, Driemeier C (2022) Density Functional Theory with Implicit Solvents for Accurate Estimation of Aqueous and Organic Solvation Free Energies of Lignin Fragments. *ACS Sustain Chem Eng* 10:10870–10878.
- [9] <https://www.rdkit.org/docs/GettingStartedInPython.html>
- [10] <https://scikit-learn.org/stable/>