

# Aprendizado de Máquina e Bioinformática: A indução do Ciclo de Krebs em anaerobiose para o aumento na produção de biohidrogênio pela bactéria *Escherichia coli*

Palavras-Chave: biohidrogênio, bioinformática, machine learning

Autores:

Cristian Leonardo Ferraz Ferreira, IB – UNICAMP  
Prof. Dr. Renato Vicentini dos Santos, IB- UNICAMP

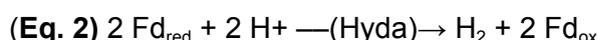
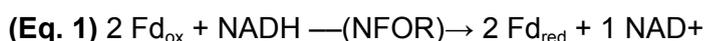
---

## INTRODUÇÃO:

Frente às adversidades ambientais presenciadas cotidianamente pela sociedade, se mostra crescente a ideia de que, caso a humanidade vise mitigar os efeitos das mudanças climáticas, a maior parte das reservas remanescentes de combustíveis fósseis devem permanecer no subsolo (Krane, 2017). Diante disso, torna-se essencial a reformulação dos processos vigentes de produção energética, demandando aos pesquisadores a busca por meios de produção sustentáveis e aplicáveis em larga escala.

Nesse sentido, no que diz respeito ao biohidrogênio, este se mostra como uma opção atraente para complementação das fontes de energia vigentes, dada sua alta densidade energética e possível produção via fermentação escura de biomassa residual (Zhang *et al.*, 2020). A produção de H<sub>2</sub> através da fermentação escura ocorre por meio da ação do metabolismo bacteriano em substratos orgânicos, este que, ao ser quebrado em compostos mais simples, leva à liberação do biocombustível (Meena; Patane, 2025).

De acordo com a literatura, são conhecidas três principais vias metabólicas que implicam na produção de H<sub>2</sub> através da fermentação escura, dentre as quais ressalta-se a via do NADH, composta pelas reações catalisadas pelas enzimas NAD(P)H ferredoxina oxidoreductase (NFOR) e hidrogenase dependente de ferredoxina (Hyda) (**Eqs. 1 e 2**) (OH *et al.*, 2011).



A partir do aumento da razão NADH/NAD<sup>+</sup> disponível intracelularmente, constata-se que a via da NFOR é capaz de produzir hidrogênio sob altos rendimentos, ultrapassando os limites estipulados

experimentalmente (Oh *et al.*, 2008). Nesse sentido, sabendo da capacidade de produção do NADH pelo Ciclo de Krebs, 3 mols por volta do ciclo (Blanco; Blanco, 2022), este pode servir como fonte da molécula para a geração de H<sub>2</sub> pela via do NADH (Cao *et al.*, 2022). Porém, dada a alta sensibilidade da Hyda para o O<sub>2</sub> (Vincent *et al.*, 2005), a realização do Ciclo de Krebs deve ocorrer sob anaerobiose, evidenciando a necessidade de modificações no metabolismo escolhido.

Dessa maneira, tomando como base o modelo metabólico em escala genômica iJO1366, da bactéria anaeróbica facultativa *Escherichia coli*, este estudo tem o objetivo de induzir a realização do Ciclo de Krebs na ausência de O<sub>2</sub> pela bactéria. Além disso, a partir da introdução da via do NADH no modelo, esperamos maximizar a produção de H<sub>2</sub> pela via, restringindo a utilização do metabólito para a síntese do biocombustível.

## METODOLOGIA:

Para a realização das análises, foram implementadas abordagens inteiramente computacionais, estas voltadas para estudos nas áreas de engenharia metabólica e aprendizado de máquina. Deste modo, a condução dos experimentos baseou-se na utilização de diferentes pacotes desenvolvidos para a linguagem de programação Python, os quais designados para a manipulação em modelos metabólicos em escala genômica (cobrapy, dFBA e escher) e criação de algoritmos de machine learning (scikit-learn).

### Criação do modelo preditivo

Primeiramente, para a avaliação das principais enzimas capazes de suportar a indução do Ciclo de Krebs em anaerobiose, realizamos o treinamento de um modelo de machine learning com 1.000.000 de combinações distintas entre os fluxos das enzimas envolvidas no ciclo. As enzimas selecionadas foram tanto aquelas diretamente atuantes no Ciclo de Krebs em aerobiose, quanto as necessárias para a entrada no ciclo (Tabela 1). Para cada enzima selecionada, foram gerados 10 valores de fluxo no intervalo correspondente aos fluxos mínimo e máximo possíveis para a enzima em anaerobiose. Dessa maneira, os valores presentes nas combinações geradas consistiram em fluxos aleatórios para as enzimas dentro dos seus respectivos limites estabelecidos.

CS	ACONTa	ACONTb	ICDHyr	AKGDH	SUCOAS	FUM	MDH	PDH	GLCptspp	PPC
----	--------	--------	--------	-------	--------	-----	-----	-----	----------	-----

Tabela 1: Enzimas selecionadas para a geração do modelo preditivo: CS (Citrato sintase); ACONTa (Aconitase - meia reação A); ACONTb (Aconitase - meia reação B); ICDHyr (Isocitrato desidrogenase); AKGDH (2-Oxoglutarato desidrogenase); SUCOAS (Succinil-CoA sintetase); FUM (Fumarase); MDH (Malato desidrogenase); PDH (Piruvato desidrogenase); GLCptspp (Transporte de glicose via PEP:Pyr (periplasma)); PPC (Fosfoenolpiruvato carboxilase)

A partir do pacote cobrapy (versão 0.19), foi realizada uma Análise de Balanço de Fluxo (FBA) para cada combinação de fluxos enzimáticos gerada, a fim de averiguar a produção de NADH e biomassa pelo modelo, quando modificado para a expressão dos fluxos impostos. Desta forma, devido à diferença nas produções de NADH e biomassa entre as combinações geradas, o modelo de aprendizado de máquina foi capaz de determinar a importância de cada enzima para a expressão do

fenótipo desejado. O algoritmo escolhido para a geração do modelo de machine learning foi o Random Forest Classifier, disponível pelo pacote scikit-learn. Dada a necessidade imposta pelo algoritmo selecionado, foi necessária a divisão dos dados gerados em duas classes distintas, as quais fornecem a base para o treinamento do modelo

A divisão das classes foi realizada de modo a classificar como favorável uma combinação de fluxos que resultou em uma taxa de produção de NADH maior do que 30 mmol/gDW/h, concomitantemente à uma taxa de produção de biomassa maior que 0.1 mmol/gDW/h. Dessa maneira, as combinações classificadas como favoráveis foram rotuladas com o número 1, sendo as combinações remanescentes rotuladas com o número 0. Após a rotulação, os dados gerados foram separados em: variáveis preditoras, compostas pelas combinações de fluxos enzimáticos; variáveis resposta, compostas pelos rótulos estabelecidos (0 para combinações desfavoráveis e 1 para favoráveis).

O algoritmo de aprendizado de máquina foi então treinado com 90% dos dados (900.000 combinações), sendo os 10% restantes utilizados para a validação do modelo. Ao fim do treinamento, o modelo criado apresentou precisão de 0.99, assim como uma curva ROC AUC equivalente a 1.00, indicando alta precisão preditiva do modelo. Finalizando as etapas equivalentes à criação do modelo preditivo, foram encontradas as principais enzimas responsáveis pela determinação das classes, sendo estipulada a importância relativa, em porcentagem, da enzima para a expressão do fenótipo. Cabe ressaltar que a métrica da importância relativa indica apenas o quão importante a característica avaliada é para a predição do modelo, sem que seja esclarecida se essa importância é positiva ou negativa.

## **Indução do Ciclo de Krebs em anaerobiose e expressão da via do NADH no modelo iJO1366**

As 3 principais enzimas determinadas pelo algoritmo preditivo (PDH, MDH e AKGDH) foram induzidas a funcionar sob seus fluxos máximos permitidos, de maneira a ser possível a expressão mínima do objetivo estabelecido pelo modelo iJO1366, a produção de biomassa. Para as induções, foi desenvolvido um algoritmo interativo capaz de avaliar a síntese de NADH e biomassa pelo modelo quando modificado o fluxo de uma ou mais das três enzimas escolhidas. Além disso, o código criado permitiu ainda a visualização dos fluxos das demais enzimas que constituem o Ciclo de Krebs, quando modificações nas três enzimas são realizadas.

Posto isso, foram induzidos dois conjuntos de fluxos enzimáticos para a PDH, AKGDH e MDH que resultaram na maximização da síntese de NADH, tornando o Ciclo de Krebs parcialmente operável sob anaerobiose, sem o comprometimento do crescimento bacteriano. Por fim, para a avaliação da capacidade de produção de H<sub>2</sub> pela via do NADH, esta foi introduzida ao modelo iJO1366 a partir das equações descritas por Oh *et al.* (2011) (**Eqs. 1 e 2**). Após a introdução da via do NADH ao modelo, os

conjuntos de fluxos enzimáticos para a PDH, AKGDH e MDH que impulsionaram a produção de NADH pela bactéria foram fixados, sendo avaliada a concentração de H<sub>2</sub> gerado pela bactéria recombinante.

## RESULTADOS E DISCUSSÃO:

De acordo com os resultados obtidos pelo algoritmo de machine learning, as três principais enzimas determinantes para a produção de NADH e biomassa foram, respectivamente, em porcentagem de importância relativa, a PDH ( $\approx 22\%$ ), a MDH ( $\approx 17\%$ ) e a AKGDH ( $\approx 15\%$ ). A PDH, ao converter o piruvato em acetil-CoA, além de preparar a inicialização do Ciclo de Krebs, é altamente capaz de estimular a produção de NADH intracelular, mesmo que fora do ciclo. Além disso, de maneira semelhante, a MDH, presente dentro do Ciclo de Krebs, ao converter o oxaloacetato em malato, também pode impulsionar a produção do metabólito.

Dessa maneira, quando analisadas com o algoritmo iterativo, a PDH e a MDH, uma vez induzidas a funcionarem sobre fluxos equivalentes a 25.33 e 9 mmol/gDW/h, respectivamente, são capazes de manter uma taxa de produção de NADH de 54.358866 mmol/gDW/h, 1.52 vezes maior que aquela observada no Ciclo de Krebs em aerobiose (35.61788 mmol/gDW/h). Além disso, para estes fluxos estabelecidos, observamos também uma manutenção da taxa de produção de biomassa equivalente a 0.103219 mmol/gDW/h, indicando um crescimento bacteriano aceitável.

Por outro lado, quando incluída nas análises iterativas, a AKGDH foi capaz de manter uma taxa ainda maior de produção de NADH (60.027871 mmol/gDW/h), sendo os fluxos enzimáticos das enzimas PDH, MDH e AKGDH equivalentes a, respectivamente, 31.96, 8.02 e -19.48 mmol/gDW/h. Neste caso, a taxa de produção de biomassa também se mostrou aceitável, sendo equivalente a 0.123101 mmol/gDW/h, maior que aquela observada anteriormente. Dessa maneira, é possível inferir que a AKGDH, quando induzida a funcionar de maneira reversa, é capaz de manter altas taxas de fluxo na PDH, principal enzima responsável pela produção de NADH no organismo.

Quando introduzida ao modelo, a via do NADH demonstrou alta capacidade de conversão do excesso de NADH gerado em H<sub>2</sub>. No caso das modificações dos fluxos da PDH e MDH para 25.33 e 9 mmol/gDW/h, respectivamente, a via do NADH foi capaz de aproveitar 96.1126% do metabólito produzido pela bactéria, levando à produção de 4.578 mol H<sub>2</sub> / mol glicose consumida, resultado 2.8 vezes maior que aquele observado para a bactéria selvagem (1.66 mol H<sub>2</sub> / mol glicose consumida). Quanto à produção de biomassa, esta também teve um aumento quando comparada à produção pela bactéria selvagem em anaerobiose (2.460921 g/L), sendo observada uma concentração final de 3.544658 g/L.

Quanto às análises referentes à introdução da via do NADH e manipulação das enzimas PDH, MDH e AKGDH, estas indicaram uma produção menor de H<sub>2</sub>, o que se deve ao próprio consumo de NADH pela AKGDH quando induzida a funcionar de maneira reversa. Neste cenário, dos 60.027871

mmol/gDW/h de NADH gerados, a via do NADH foi capaz de aproveitar apenas 64.3134%, sendo o restante consumido pela AKGDH. Deste modo, a produção de H<sub>2</sub> pela bactéria foi limitada a 3.61476 mol H<sub>2</sub>/ mol glicose consumida, sendo a concentração final de biomassa equivalente a 2.37 g/L.

## CONCLUSÕES:

Os resultados obtidos demonstram a possibilidade de indução parcial do Ciclo de Krebs em anaerobiose, seja pela modificação apenas nas enzimas PDH e MDH, assim como na inclusão da reação inversa da AKGDH na bactéria. Se mostra também evidente a importância da PDH para o aumento na síntese de NADH pela bactéria, o qual pode ser utilizado por diversas outras rotas metabólicas economicamente importantes além daquela de produção do H<sub>2</sub>. Deste modo, torna-se clara a necessidade de modificações na enzima que visem sua superexpressão em ensaios experimentais *in vitro* destinados ao aumento da quantidade de NADH produzido.

Destaca-se também a capacidade de aproveitamento do NADH produzido pela via recém introduzida, esta que não somente é capaz de impulsionar a produção de H<sub>2</sub> a rendimentos altos, mas também possibilita um destino melhor para o NADH, uma vez que, quando presente no organismo, a via é capaz de estimular um maior crescimento bacteriano. Em suma, o presente projeto serve como base para futuras experimentações *in vitro* destinadas ao aumento da produção de biohidrogênio através da via do NADH, provendo conhecimento base para métodos de produção do biocombustível em larga escala.

## BIBLIOGRAFIA

- BLANCO, Antonio; BLANCO, Gustavo. **Chapter 14 - Carbohydrate metabolism**. In: BLANCO, Antonio; BLANCO, Gustavo (Orgs.). *Medical Biochemistry (Second Edition)*. [S.I.]: Academic Press, 2022. p. 315–358.
- CAO, Y. et al. **Debottlenecking the biological hydrogen production pathway of dark fermentation: insight into the impact of strain improvement**. *Microbial Cell Factories*, v. 21, n. 1, p. 166, 2022.
- KRANE, Jim. **Climate change and fossil fuel: An examination of risks for the energy industry and producer states**. *MRS Energy & Sustainability*, v. 4, p. E2, 2017.
- MEENA, Pradeep Kumar; PATANE, Prashant Mahadev. **Biohydrogen: Advancing a sustainable transition from fossil fuels to renewable energy**. *International Journal of Hydrogen Energy*, v. 100, p. 955–970, 2025.
- OH, Y. K. et al. **Metabolic-flux analysis of hydrogen production pathway in *Citrobacter amalonaticus* Y19**. *International Journal of Hydrogen Energy*, v. 33, n. 5, p. 1471–1482, 2008.
- OH, Y. K. et al. **Current status of the metabolic engineering of microorganisms for biohydrogen production**. *Bioresource Technology*, v. 102, n. 18, p. 8357–8367, 1 set. 2011.
- VINCENT, Kylie A. et al. **Electrochemical Definitions of O<sub>2</sub> Sensitivity and Oxidative Inactivation in Hydrogenases**. *Journal of the American Chemical Society*, v. 127, n. 51, p. 18179–18189, 1 dez. 2005.
- ZHANG, Zhiping et al. **Investigation of the interaction between lighting and mixing applied during the photo-fermentation biohydrogen production process from agricultural waste**. *Bioresource Technology*, v. 312, p. 123570, 2020.