

“Implementation of FBA problem”

O perfil metabólico do modelo da *Escherichia coli* (estirpe K12 e sub-estirpe MG1655), disponibilizado pelo BiGG através do identificador “iML1515”, apresenta 1877 metabolitos, 2712 reações e 1516 genes (**Anexo 1**). O foco principal deste trabalho passa por tentar encontrar uma estratégia teórica em termos de engenharia biológica para maximizar a produção de lactato desta sub-estirpe. Para isso, utilizaram-se as funcionalidades do MewPy.

Em primeiro lugar, fez-se a simulação do modelo (com o método “simulate”) e especificou-se o algoritmo **pFBA** (Parsimonous Flux Balance Analysis) para tentar otimizar a função objetivo com o mínimo de alterações ao organismo *wt* possível (de forma que possa ser aplicável em laboratório). Foi necessário, em seguida, encontrar o nome dado ao lactato para este modelo (para descobrir qual a produção *wild-type* deste metabolito), à fonte de carbono (glucose), ao oxigénio consumido, e à biomassa. Para isso, procurou-se pelo nome “lac”, “glc”, “o2” e “BIO”, respetivamente, com o método “find” (**Anexo 2**). O resultado obtido mostra-nos que existem 2 tipos de produção de lactato (“EX_lac__L_e” e “EX_lac__D_e”), e que em condições normais nenhuma destas reações ocorrem (fluxo = 0). Para além disso, mostra-nos o identificador da glucose consumida (“EX_glc__D_e”), do oxigénio consumido (“EX_o2_e”) e da biomassa (“BIOMASS_Ec_iML1515_core_75p37M”). Desta forma podemos depois averiguar se as alterações efetuadas para maximizar a produção de lactato afeta a produção de biomassa. Assim, também já é possível repetir a simulação aplicando as restrições ambientais impostas pelo enunciado, isto é, sabemos que a cultura se encontra em condições anaeróbias (fluxo de $O_2 = 0_{\text{mmol/gDW/h}}$) e com um fluxo de glucose igual a $15_{\text{mmol/gDW/h}}$.

Em relação ao lactato, verificou-se também o fluxo máximo dos 2 metabolitos com o método “FVA”, obtendo-se $1.11_{\text{mmol/gDW/h}}$ para o L-lactato e $12.26_{\text{mmol/gDW/h}}$ para o D-lactato (**Anexo 3**).

Estamos em condições de simular modificações de forma a melhorar o fluxo de lactato. Primeiro, verificou-se se é possível melhorar o fluxo com o *knockout* de apenas um gene. Criou-se um ciclo no Python para percorrer todos os genes existentes do modelo, um de cada vez, simulando os respetivos *knockouts* e, caso melhorasse o fluxo de lactato, adicionava-se o nome do respetivo gene e o novo fluxo (**Anexo 4**). Os resultados indicam que, para o L-lactato, não houve melhorias de fluxo, mas para o D-lactato, houve uma deleção de um gene (“s0001” - um gene requerido na realização de diversas reações) que permiti aumentar o seu fluxo para $6.86_{\text{mmol/gDW/h}}$. Precisamos agora de ver se o *knockout* do gene “s0001” afetou a produção da biomassa (**Anexo 5**). Vemos que o fluxo da biomassa passou a ser 0, pelo que podemos concluir que, apesar do *knockout* melhorar o fluxo de D-lactato (em teoria), não iria haver produção de lactato à mesma visto que não haveria produção de biomassa.

Para ver quais as modificações em termos genéticos (através de *knockouts* e sub/sob-expressões), correu-se o algoritmo evolutivo de otimização (“Evolutionary Algorithms”). Definiram-se as restrições ambientais, o identificador do produto (D-lactato / L-lactato), o da

biomassa e o da fonte de carbono. De seguida, correu-se o algoritmo para primeiro efetuar a otimização através de sub/sob-expressão de genes, e depois através de *knockouts*, em ambos os casos restringindo o número máximo de alterações para 5 (tendo já em consideração a viabilidade em contexto laboratorial) e o número de gerações para 10. Isto gerou no total 4 simulações: produção de L-lactato com sub/sob-expressões, L-lactato com *knockouts*, D-lactato com sub/sob-expressões, e D-lactato com *knockouts*.

Para o L-lactato, nem a simulação de *knockouts* (**Anexo 6**), nem a das sub/sob-expressões (**Anexo 7**) melhorou a produção do composto sem comprometer a produção de biomassa. Já para o D-lactato, foi possível aumentar a produção do composto e manter a biomassa em crescimento. Desta forma, decidiu-se repetir as simulações, mas desta vez para 30 gerações, obtendo-se os resultados finais das simulações *knockout* (**Anexo 8**) e sub/sob-expressão (**Anexo 9**).

De entre as soluções propostas, optou-se por escolher uma solução de entre os da simulação de *knockouts*, visto que se obtiveram valores muito superiores relativamente ao WYIELD em comparação com a simulação de sub/sob-expressões. A solução selecionada foi a do *knockout* dos genes que produzem a enzima Acetato Cinase (**ACKr**) e os que produzem a enzima Acetaldeído Desidrogenase (**ACALD**), uma vez que correspondem apenas a 2 alterações enzimáticas para a obtenção de um WYIELD de 26.54 mmol/gDW/h (ou seja, induz um elevado fluxo de D-lactato) e de um BPCY de 0.37 mmol/gDW/h (indicando que continua a haver produção de biomassa) (**Anexo 10**). Aplicaram-se as novas alterações ao modelo e verificaram-se as novas produções de lactato e de biomassa (**Anexo 11**). Comprova-se assim que o fluxo de D-lactato aumentou de forma muito significativa, e o fluxo de biomassa continua com um valor muito próximo do *wt*.

Após uma verificação no BiGG relativamente aos genes responsáveis pela produção de ACKr e ACALD, inferiu-se que existem 3 genes (*tdtD*, *purT* e *ackA*) que produzem ACKr, tendo uma regra GPR do tipo (**G1 or G2 or G3**) (**Anexo 12**), e 2 genes (*adhE* e *mhpF*) que produzem ACALD tendo uma regra GPR do tipo (**G1 or G2**) (**Anexo 13**). Desta forma, para impedir a produção destas enzimas e, consequentemente, impedir as reações que despoletam, é necessário realizar o *knockout* dos 5 genes. Em contexto laboratorial, isto parece-nos viável, pois a quantidade de alterações genéticas não é insensata e permite obter uma quantidade considerável de D-lactato, tendo em conta que pela estirpe *wt* não era produzido.

Assim, podemos concluir que o objetivo desta ficha é alcançado, na medida em que, utilizando as ferramentas informáticas disponibilizadas, foi possível encontrar uma estratégia que aumentasse a produção do composto desejado, nas condições propostas e mantendo o fluxo de biomassa positivo, tendo em conta que o a produção *wt* de lactato era nula. Inferimos, deste modo, que a melhor estratégia para alcançar o objetivo de maximizar a produção de lactato, passa pelo *knockout* de alguns genes, o que, na nossa perspetiva, é um procedimento possível de transpor da área bioinformática para a prática laboratorial.

Trabalho realizado por:

André Silva (PG); António Duarte (PG); Roberto Bullitta (PG45474);
Rute Castro (PG); Vânia Miguel (PG45971).