# **Relatório da Task 2**:

# Desenvolvimento de Modelos Metabólicos à Escala Genómica para os organismos *Synechocystis* e *Chlorella* *vulgaris* e respetiva Identificação de Vias Metabólicas Relacionadas ao CH4 para ambos os microrganismos.

Os principais objetivos desta tarefa passam pelo desenvolvimento de modelos metabólicos à escala genómica dos organismos *Synechocystis* e *Chlorella* *vulgaris*. A partir destes, e juntamente com outros tipos de análises pretende-se identificar possíveis vias metabólicas relacionadas com o metano.

O desenvolvimento de modelos metabólicos à escala genómica pode ser subdividido nas seguintes sub-tarefas: Revisão da literatura, Anotação Funcional e Estrutural dos Genomas, Curação Manual, Conversão para Modelo estequiométrico e Validação do modelo.

1. **Revisão da literatura**:

Esta tarefa envolveu a revisão e avaliação da literatura existente relacionada ao metabolismo da *Synechocystis* e *Chlorella vulgaris*. Para além disso, serviu para avaliar os modelos metabólicos previamente reconstruídos para estes microrganismos. Para tal, foi examinada a sua cobertura metabólica, precisão e utilidade para os objetivos específicos do nosso projeto. Esta análise passou pela compreensão de como os modelos foram construídos, quais dados foram incorporados e como foram validados. No final, houve uma procura por ferramentas bioinformáticas e recursos web que poderiam ser úteis para a reconstrução dos modelos.

No caso da *Synechocystis*, foram encontrados 2 modelos na base de dados ‘BiGG Models’, nomeadamente iJN678 e SynCJ816, estes foram analisados e alguns dos dados destes modelos foram utilizados na reconstrução do modelo GSM deste organismo.

Até ao momento, a literatura apenas menciona a existência de um modelo metabólico estabelecido em 2016 para a *Chlorella* *vulgaris*. Embora este modelo tenha sido validado experimentalmente em relação à taxa de crescimento e à absorção de vários metabolitos, é importante considerar que, ao longo dos anos, a informação pode ter sido atualizada, possibilitando uma abordagem mais precisa e detalhada do genoma e do metabolismo desta espécie. Além disso, é relevante destacar que este modelo não incorpora o metabolismo do metano, que é o foco principal do deste projeto.

1. Anotação funcional e estrutural

De forma a dar início ao processo de reconstrução, foi primeiro necessário obter os genomas de ambas as espécies. Esta informação foi obtida a partir da base de dados NCBI. A ferramenta *merlin* foi a escolhida para a reconstrução dos modelos pela sua capacidade de simplificar todo o processo e pela sua versatilidade em lidar com organismos tanto procarióticos quanto eucarióticos.

O processo de anotação das enzimas foi efetuado através de buscas de homologia contra as bases de dados UniProtKB/Swiss-Prot e UniProtKB/TrEMBL, utilizando o DIAMOND e o BLAST para a *Synechocystis* e para a *Chlorella*, respetivamente. A ferramenta “Automatic Workflow” presente no *merlin*, permitiu ainda efetuar a anotação funcional do genoma. O processo de avaliação da ferramenta é baseado na distância taxonómica, atribuindo uma pontuação a cada organismo encontrado nos resultados do DIAMOND\BLAST. Assim, a ferramenta utiliza uma lista de organismos filogeneticamente relacionados para realizar a anotação. Esta lista pode ser gerada automaticamente levando em consideração o número de vezes que cada organismo ocorre nos resultados e a pontuação taxonómica ou manualmente selecionando organismos filogeneticamente próximos.

Em ambos os casos, a anotação dos transportadores foi conduzida e integrada no Merlin utilizando o Transport Systems Tracker (TranSyT). Adicionalmente, a compartimentalização foi realizada utilizando o PSORTb 3.0, um software preciso de localização subcelular de proteínas bacterianas para a *Synechocystis*, e o LocTree para a *Chlorella*, considerando sua capacidade de previsão em organismos eucarióticos.

Por fim, a estrutura e conectividade da rede metabólica à escala genómica foram definidas pela integração dos dados metabólicos carregados do Kyoto Encyclopedia of Genes and Genomes (KEGG) e da anotação funcional das enzimas.

Os resultados desta etapa para ambos os organismos são os seguintes:

*Synechocystis***:**

* Os três organismos, *Synechococcus elongatus, Thermosynechococcus elongatus* e *Nostoc sp.,* foram identificados como cianobactérias que são filogeneticamente próximas ao *Synechocystis sp;*
* 1.493 dos 3.565 genes encontrados no genoma foram anotados com base em entradas do Swiss-Prot, enquanto as 2.072 anotações restantes foram baseadas em entradas do UniProtKB;
* 424 reações de transporte automaticamente geradas pelo Transyt, mas apenas 210 foram adicionadas ao modelo;
* 2.185 reações automaticamente e 724 adicionadas manualmente;
* 5 compartimentos: citoplasma, membrana citoplasmática, membrana externa, periplasma e extracelular;

Chlorella:

* Os organismos selecionados manualmente foram espécies do genus *chlorella, Chlamydomonas reinhardtii, Arabidopsis thaliana* e *Oryza sativa subsp. Japónica*;
* Foi verificado que grande parte da anotação foi baseada na *Arabidopsis* *thaliana*. Isto poderá ser explicado pelo facto de que este é um organismo muito estudado e por isso contém bastante informação em bases de dados. Dos 11.141 genes da *Chlorella*, 6.086 foram anotados com base na Swiss-Prot e os restantes 5.055 anotados com base no TrEMBL;
* 1.428 reações automaticamente e 240 adicionadas manualmente;
* 11 compartimentos: extracelular, citoplasma, cloroplasto, retículo endoplasmático, complexo de golgi, peroxissoma, mitocôndria, lúmen, núcleo, vacúolo e pfm.

1. Curação Manual/Gap-filling

Após a anotação do genoma, a rede metabólica preliminar foi montada e refinada. Como o modelo reconstruído automaticamente pode conter erros ou dados em falta (gaps), a curadoria do modelo implicou em corrigir esses problemas e preencher quaisquer lacunas na rede metabólica. Isso incluiu verificar o balanço de massa de reações individuais, associações de reações e direcionalidade das reações. Para tal, foram utilizados bancos de dados como KEGG, BRENDA e MetaCyc. Este processo foi efetuado via a via para todo o modelo.

1. Conversão para Modelo Estequiométrico

A definição da equação da biomassa é um dos passos fundamentais na reconstrução de modelos metabólicos. Para ambos os casos, esta foi elaborada com base na literatura existente e nos modelos metabólicos já estabelecidos.

*Synechocystis***:**

A biomassa inclui proteínas, DNA, RNA, lípidos, glicogénio, LPS, peptidoglicano, pigmentos, cofatores e iões inorgânicos. É importante destacar que a maior parte da biomassa é composta por proteínas, lipídios, RNA e DNA. Top of FormA composição lipídica foi retirada da literatura, já que os lipídios nesses modelos não estão completamente definidos estruturalmente. Dada a importância metabólica e biotecnológica da biossíntese de ácidos gordos na *Synechocystis* sp. e a disponibilidade de informações sobre a biossíntese de lipídios, cinco classes de lipídios foram adicionadas: monogalactosildiacilglicerol (MGDG), digalactosildiacilglicerol (DGDG), sulfiquinovosildiacilglicerol (SQDG), o fosfolipídeo fosfatidilglicerol (PG) e triacilglicerol (TG). No final, a composição dos componentes individuais foi processada, e os dados resultantes foram exportados para o *merlin*.

*Chlorella*:

A composição das macromoléculas DNA e proteínas calculada automaticamente pelo *merlin*. A composição das restantes macromoléculas (cofatores, carbohidratos, pigmentos e RNA) foi retirada da literatura. Os lípidos foram retirados do modelo já existente.

Adicionalmente, nesta etapa foram também determinadas as necessidades energéticas dos organismos. As GAM (Growth-associated maintenance) e NGAM (Non-growth-associated maintenance) foram retiradas dos modelos já estabelecidos.

1. Validação do modelo

Para a validação inicial do funcionamento correto dos modelos ambos foram validados usando a biblioteca COBRApy. Diferentes aspetos do modelo foram verificados, tais como:

Um método para verificar se todas as reações nos modelos estão balanceadas em termos de massa, garantindo que a entrada seja igual à saída de massa para cada reação, um método que deteta ciclos de produção de energia no modelo, o que pode levar a previsões irreais. Este identifica metabolitos que estão envolvidos em ciclos de produção de energia, onde a energia é produzida sem a entrada de substrato.

Para além destes, há outro método que verifica se o modelo é capaz de produzir biomassa, um indicador essencial para garantir a viabilidade do crescimento celular simulado, e ainda um que identifica reações que estão bloqueadas no modelo, ou seja, que não têm fluxo através delas sob condições padrão, e outro que verifica se há reações com fluxo ilimitado, o que pode indicar problemas de modelação, como reações mal definidas ou limites de fluxo incorretos. Este calcula a variabilidade do fluxo para cada reação e relata aquelas com fluxo incorreto.

Um método que verifica se há reações que violam o equilíbrio do saldo de carga no modelo. O equilíbrio do saldo de carga é essencial para a precisão das previsões do modelo, pois garante que a conservação de carga elétrica seja mantida em todas as reações e, identifica e relata as reações que têm os mesmos metabolitos como reagentes e produtos.

Durante esta validação várias curações manuais foram feitas aos modelos nomeadamente, em termos de reversibilidade e balanceamento de reações. O *merlin* permite fazer isto automaticamente, utilizando modelos do banco de dados ModelSEED para esse fim, mesmo assim o banco de dados do ModelSEED também não tem todas as reações com a revesibilidadade correta e então no final ainda houve a correção de reversibilidade e balanceamento de reações apontadas através de base de dados como o MetaCyc.

Outro tipo de validações foi efetuado como a determinação de crescimento espontâneo (se cresce ao retirar o meio), taxa de crescimento sob diferentes condições e comparação com valores de absorção e secreção de metabolitos com valores experimentais.

A validação dos modelos, como o modelo de *Synechocystis* e *Chlorella vulgaris* é fundamental para garantir a confiabilidade e a precisão dos resultados obtidos a partir das simulações. Essa validação permite identificar erros, inconsistências e lacunas nos modelos, garantindo que estes capturem adequadamente o comportamento dos microrganismos em estudo. Além disso, uma vez validado o modelo serve como uma ferramenta confiável para investigar e compreender os processos biológicos subjacentes.

Com base nos resultados obtidos, pode-se observar que o modelo reconstruído iJG706 se mostra geralmente consistente com os resultados experimentais e com os outros modelos (iJN678) previamente feitos, nomeadamente em termos de taxa de crescimento da *Synechocystis* em diferentes condições metabólicas, incluindo heterotrofia, mixotrofia e autotrofia. Embora haja algumas pequenas discrepâncias entre os valores previstos pelo modelo e os valores experimentais, as diferenças são geralmente pequenas e dentro de uma faixa aceitável de variação. Portanto, podemos concluir que o modelo iJG706 é válido e representa adequadamente o comportamento metabólico da *Synechocystis* em diferentes condições de crescimento

Relativamente ao modelo iGA1305, este também é capaz de simular as 3 diferentes condições, ainda que, relativamente aos valores de taxa de crescimento experimentais sob fotoautotrofia e mixotrofia sejam longe do esperado. Ainda assim, esta diferença é também relatada na taxa de crescimento do modelo iRC1080 (pertencente à chlamydomonas) comparativamente aos seus valores experimentais.

**Overview dos modelos**

*Synechocystis*

O meio usado foi o seguinte: H2O, Fe2+, O2, Nitrate, CO2, Sulfato, H+, ortofosfato, *hv*, Mg2+, Na+, este é um meio modelo para crescimento de microrganismos.

Quanto ao modelo iJG706, este contém 2172 reações, 1619 metabolitos e 706 genes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Modelo | Genes | Reações | Metabolitos | Compartimentos |
| iJG706 | **706** | **2172** | **1619** | **[e] [p] [c] [u] [l]** |
| iJN678 | 678 | 863 | 795 | [e] [c] [p] [u] |
| iSyn811 | 811 | 956 | 911 | [e] [c] |
| iSyn669 | 669 | 882 | 790 | [e] [c] |

Compartimentos:[e] espaço extracelular; [p] periplasma; [c] citoplasma; [u] tilacoide; [l] carboxissoma.

O modelo contém um total de 1619 metabólitos diferentes. Esses metabólitos estão divididos pelos diversos compartimentos da seguinte forma: carboxissoma (8), metabólitos extracelulares (69), periplasma (77) e citoplasma (1465).

|  |  |  |
| --- | --- | --- |
|  | Modelos | Taxa de crescimento |
| Heterotrofia | iJN678 | 0.063 |
| Experimental | 0.076 |
| **iJG706** | **0.064** |
| Mixotrofia | iJN678 | 0.056 |
| Experimental | 0.059 |
| **iJG706** | **0.043** |
| Autótrofia | iJN678 | 0.088 |
| Experimental | 0.085 |
| **iJG706** | **0.085** |

**Nota:** Para a obtenção destes resultados foi usada a luz solar, devido ao facto do modelo iJN678 e os resultados experimentais também usaram este tipo de luz. Para além deste tipo de luz o modelo também foi testado com luz branca, uma vez que, a equipa laboratorial usa este tipo de luz nas suas experiências. Observou-se que quando se usava a luz branca em vez da luz solar as taxas de crescimento desciam ligeiramente.

*Chlorella:*

O meio constituinte para o modelo contém água, ortofosfato, oxigénio, sulfato, hidrogénio, nitrate, ferro e sódio. O modelo contém 2.659 reações, 2.207 metabolitos e 1.312 genes.

**Tabela:** Experimental and predicted growth rates reported for C. *vulgaris* and C. *reinhardtii*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Model*** | ***iCZ843*** | | ***iRC1080*** | |
| ***Predicted Growth Rate*** | ***Experimental*** | ***Predicted Growth Rate*** | ***Experimental*** |
|  | *h−1* | | | |
| **Photoautotrophy** | *0.0252* | *0.014–0.025* | *0.1538* | *0.035–0.09* |
| **Heterotrophy** | *0.0168* | *0.018–0.025* | *0.0299* | *0.059–0.084* |
| **Mixotrophy** | *0.0407* | *0.02–0.03* | *0.1817* | *0.066* |

**Tabela:** Predicted growth rates for the iGA1305 model (this work).

|  |  |
| --- | --- |
| *iGA1312* | *Predicted Growth Rate* |
| **Photoautotrophy** | *0.245* |
| **Heterotrophy** | *0.0236* |
| **Mixotrophy** | *0.27* |

**Tabela:** Experimental and predicted growth rates for medium containing Tryptophan, Methionine and Acetate.

|  |  |  |  |
| --- | --- | --- | --- |
| Heterotrophic conditions | *i*CZ843 | *i*GA1312 | Experimental |
| **Glucose + Tryptophan** | 0.040 | 0.0239 | 0.040 |
| **Glucose + Methionine** | 0.032 | 0.0239 | 0.032 |
| **Glucose + Acetate** | 0.017 | 0.0236 | 0.017 |

**Tabela: Overview do modelo, juntamente com os modelos da Chlorella e chlamydomonas**

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Genes | Reações | Metabolitos |
| *i*GA1305 | 1305 | 2659 | 2207 |
| *i*CZ843 | 843 | 2286 | 1770 |
| *i*RC1080 | 1086 | 2191 | 1706 |

**Principais limitações:**

As principais limitações passaram pela falta de informação presente na literatura relativamente à capacidade destes organismos (e semelhantes) utilizarem o metano. Outra limitação é o facto de não haverem evidências ao nível do genoma de enzimas relacionadas com o metabolismo do metano (quer produção, quer consumo). É crucial reconhecer que dados experimentais gerados pelo grupo têm o potencial de fornecer contribuições significativas para o refinamento dos modelos.