



Universidade do Minho

Legionella pneumophila

Biossíntese de Flavina

Cursos: Mestrado em Bioinformática

Mestrado em Engenharia Informática

Unidade Curricular: Laboratórios de Bioinformática

Grupo 6:

Ana Carolina Silva PG35351

Ana Daniela Fernandes PG33745

Daniel Soares PG35358

Carlos Martins PG36095

Eduardo Cunha A71940

Índice

Objetivos	2
Introdução	2
Análise e anotação dos genes e proteínas de interesse	3
Análise da Literatura	5
Análise de homologias por Blast	6
Alinhamento múltiplo e filogenia	9
Regulação da via metabólica	12
Fatores de virulência	13
Resistência a antibióticos	13
Domínios de proteínas	13
Análise dos <i>motifs</i> e estrutura das proteínas	15
Referências	17

Objetivos

Este trabalho tem como principal objetivo a utilização de algumas ferramentas bioinformáticas abordadas na unidade curricular de Laboratórios de Bioinformática. Mais concretamente será analisada a via metabólica – Biossíntese de Flavina - da *Legionella pneumophila subsp. pneumophila str. Philadelphia 1*, assim como proteínas e genes de interesse para esta via, não só ao nível metabólico como também ao nível da regulação. Para esta análise recorreu-se à pesquisa em bases de dados relevantes, assim como ao desenvolvimento de *scripts* de Biopython, que permitem automatizar a análise e anotação de genes e proteínas de interesse.

Introdução

A Doença do Legionário foi reconhecida pela primeira vez em 1976 após um surto de pneumonia numa convenção que decorria em Filadélfia. Após análise dos pacientes o microrganismo responsável pela infeção foi *identificado* como um bacilo gram-negativo a que se deu o nome de *Legionella pneumophila*. Atualmente sabe-se que esta bactéria é aeróbica e pertence ao filo *Proteobacteria*, classe *Gammaproteobacteria*, ordem *Legionellales* e à família das *Legionellaceae* (família esta que engloba mais de 50 espécies e 70 subespécies). A *Legionella pneumophila* em específico tem 16 estirpes que estão designadas de 1 a 16, sendo que as estirpes 1, 4 e 6 são as mais comumente implicadas na infeção em Humanos (a estirpe em que o nosso trabalho se focará é a *Legionella pneumophila subsp. pneumophila str. Philadelphia 1*, sendo esta a estirpe mais patogénica causando entre 70 e 90% dos casos de Legionelose em pacientes adultos)¹.

Este microrganismo desencadeia Legionelose em pacientes infetados, podendo esta condição referir-se a duas síndromes clínicas distintas: a **Doença do Legionário** e a **Febre de Pontiac**. Clinicamente a Doença do Legionário caracteriza-se como uma pneumonia aguda que se manifesta após um período de incubação que tem entre 2 e 10 dias acompanhada por febre geralmente elevada (40°C), expectoração podendo ou não conter sangue, dor no peito, náuseas, vômitos e diarreia. Por outro lado, a Febre de Pontiac é clinicamente menos grave sendo caracterizada como uma infeção do trato respiratório que acaba por desaparecer naturalmente, sendo muitas vezes confundida como uma gripe comum².

Os organismos desta espécie são bactérias gram-negativas e aeróbicas que crescem em habitats aquáticos naturais, como rios ou lagos, ou artificiais, como torres de água, sistemas de refrigeração, equipamento de terapia de respiração, chuveiros ou spas. O meio de crescimento tem que conter preferencialmente L-cisteína e ferro, quanto à temperatura ideal de crescimento esta situa-se entre os 28 e os 40°C, sendo que abaixo de 20°C a bactéria encontra-se dormente e acima de 60°C o microrganismo morre. A transmissão aos humanos ocorre através da inalação de aerossóis contaminados com a *Legionella pneumophila* ou através do contacto entre feridas e água contaminada.

A fisiopatologia da *Legionella* inicia-se quando os microrganismos são aspirados e chegam aos alvéolos pulmonares onde sofrem fagocitose pelos macrófagos aí presentes. Uma vez que estes macrófagos não têm a capacidade de matar a *Legionella* esta continua o seu crescimento e divisão dentro do macrófago até que este deixa de ter a capacidade de conter tantas bactérias e rompe

libertando as mesmas e infectando outros macrófagos, situação que ocorre ciclicamente. Para além desta capacidade a *Legionella* apresenta genes que codificam para fatores de virulência que potenciam a infeção dos macrófagos e inibem a ligação dos mesmos aos fagossomas. Esta imunidade celular desencadeada pelos macrófagos é acompanhada pela produção de citocinas que vão tentar inibir o crescimento da *Legionella*. Após este crescimento inicial a infeção evolui para uma pneumonia aguda com alveolite e bronquiolite, podendo em casos mais graves afetar os nódulos linfáticos, cérebro, rins, fígado, baço, medula óssea e o miocárdio³.

Análise e anotação dos genes e proteínas de interesse

De forma a proceder à anotação dos genes e proteínas de interesse realizou-se uma tabela com a seguinte informação:

- Identificação do gene: geneID; accession (NCBI); locus tag (NCBI); nome do gene (NCBI); strand (NCBI);
- Identificação da proteína: UniprotID; grau de revisão (Uniprot); accession number (Uniprot); protein name (Uniprot); número de aminoácidos (Uniprot); EC number (NCBI); descrição/função (NCBI); localização (Biocyc); gene ontology (Uniprot).

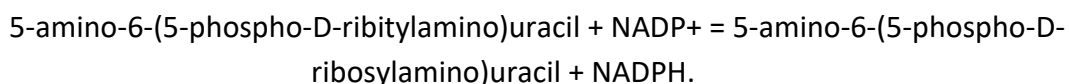
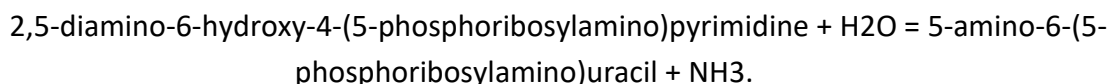
Esta informação foi extraída das bases de dados de forma manual ou com recurso a um *script* do Biopython e confirmada, manualmente, com a tabela disponibilizada no NCBI (https://www.ncbi.nlm.nih.gov/genome/proteins/416?genome_assembly_id=300116), tendo sido toda a informação validada.

Como auxílio para esta anotação, começou-se por verificar que proteínas estão relacionadas com a via metabólica, biossíntese de flavina. Desta forma, recorreu-se à plataforma Biocyc onde se verificou que existem cinco proteínas de interesse. Também foi criado um *script* que permitisse aceder ao NCBI e extrair o genoma completo do organismo *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1, guardando este num ficheiro do tipo genbank.

■ Gene **ribD**

Este gene codifica a proteína *riboflavin biosynthesis protein RibD*.

A *Riboflavin biosynthesis protein RibD* é uma enzima que entra na via metabólica da Flavina sendo responsável pela síntese da 5-amino-6-(D-ribitylamino)uracil a partir do GTP. Esta enzima tem como cofator o ião zinco e apresenta diversos locais de ligação, quer ao substrato quer ao NADP. Dentro da via é responsável pela catálise das seguintes equações químicas:

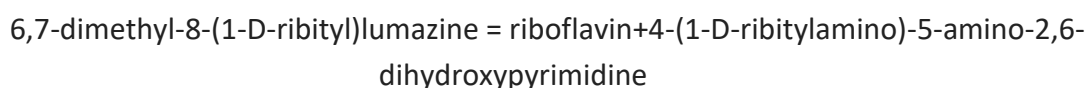


A localização subcelular da proteína foi inferida com a consulta da plataforma LOCTREE, tendo-se chegado à conclusão que a localização é o citoplasma, obtendo-se esta informação com um *score* de 100, sendo por isso uma informação confiável.

- **Gene ribE**

Este gene codifica a proteína *riboflavin synthase*.

A *Riboflavin synthase* é uma enzima que entra na via metabólica da Flavina e é responsável pela síntese de riboflavina e 5-amino-6-(D-ribitylamino) uracil a partir de duas moléculas de 6,7-dimethyl-8-ribityllumazine. Esta enzima é ativa por iões de sulfito e encontra-se no citosol da bactéria. Apresenta locais de ligação para a molécula de 6,7-dimethyl-8-ribityllumazine e a equação da catálise por si exercida é a seguinte:

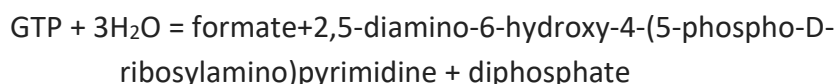


A localização subcelular da proteína foi inferida com a consulta da plataforma LOCTREE, tendo-se chegado à conclusão que a localização é o citoplasma, obtendo-se esta informação com um *score* de 92, sendo por isso uma informação confiável.

- **Gene ribA**

Este gene codifica a proteína GTP cyclohydrolase-2.

A *GTP cyclohydrolase-2* é uma enzima que entra na via metabólica da síntese de Flavina entrando no primeiro passo da via e sendo responsável pela síntese de 5-amino-6-(D-ribitylamino)uracil a partir do GTP. Apresenta como cofator o ião zinco e apresenta diversos locais de ligação para este ião e para o GTP. A equação química de degradação do GTP em que esta enzima está envolvida é a seguinte:

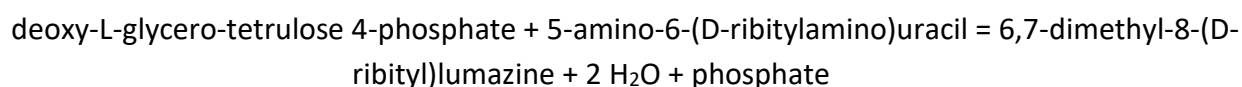


A localização subcelular da proteína foi inferida com a consulta da plataforma LOCTREE, tendo-se chegado à conclusão que a localização é o citoplasma, obtendo-se esta informação com um *score* de 100, sendo por isso uma informação confiável.

- **Gene ribH**

Este gene codifica a proteína *6,7-dimethyl-8-ribityllumazine synthase*.

A *6,7-dimethyl-8-ribityllumazine synthase* é uma enzima da via metabólica da biossíntese da Flavina que é responsável pela catálise de 6,7-dimethyl-8-ribityllumazine através da condensação 5-amino-6-(D-ribitylamino)uracil com 3,4-dihydroxy-2-butanone 4-phosphate. Estas moléculas ligam-se à enzima através de locais de ligação específicos da mesma. A equação química desta reação é:

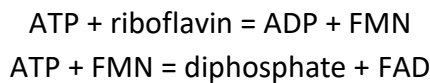


A localização subcelular da proteína foi inferida após consulta da plataforma LOCTREE, tendo-se chegado à conclusão que a localização é o citoplasma, contudo obteve-se esta informação com um *score* de apenas 10, sendo por isso uma informação pouco confiável.

▪ **Gene *ribF***

Este gene codifica a proteína *ribiflavin biosynthesis protein RibF*.

A *Riboflavin biosynthesis protein RibF* é uma enzima que entra na subvia da síntese do FAD a partir do FMN sendo a responsável por esta conversão. De igual forma está também relacionada com a biossíntese do FMN a partir da riboflavina. Esta enzima tem locais de ligação para o ATP, FAD, Riboflavina e FMN. As equações de conversão de riboflavina em FMN e FMN em FAD são, respetivamente, as seguintes:



A localização subcelular da proteína foi inferida com a consulta da plataforma LOCTREE, tendo-se chegado à conclusão que a localização é o citoplasma, obtendo-se esta informação com um *score* de 97, sendo por isso uma informação confiável.

Análise da Literatura

Primeiramente, começou-se por automatizar a seleção dos artigos relevantes que permitiam conhecer melhor o organismo, via metabólica e genes de interesse para o grupo. Esta automatização foi realizada com recurso ao Biopython que permitiu uma pesquisa rápida e eficiente na plataforma Pubmed. Assim, desenvolveram-se duas funções: uma que permite obter o número de artigos disponíveis relacionados com termo colocado como argumento e outra que permite obter outras informações (como autor, título, etc) importantes para identificação dos artigos relevantes.

A via da Biossíntese da Riboflavina, também conhecida como vitamina B2 (este composto é um nutriente essencial, sendo que os mamíferos não têm capacidade de o produzir), é uma via essencial à sobrevivência do organismo sendo os seus produtos, FMN e FAD, bastante utilizados como reagentes noutras vias metabólicas, essencialmente em reações redox.

De forma a produzir uma molécula de riboflavina a via (ilustrada na Figura 1) inicia-se com a reação entre uma molécula de GTP e duas moléculas de ribulose 5-fosfato, havendo a formação de 5-amino-6-ribitylaminuracilo devido a um conjunto de reações nomeadamente desaminações, reduções e desfosforilações. Por sua vez a condensação entre este composto e o 1-deoxi-L-glicero-tetrol-4-fosfato obtida da ribulose 5-fosfato leva à formação de 6,7-dimetil-8-(1-D-ribityl)lumazine. Por fim, devido à ação enzimática resulta a riboflavina^{4,5,6}.

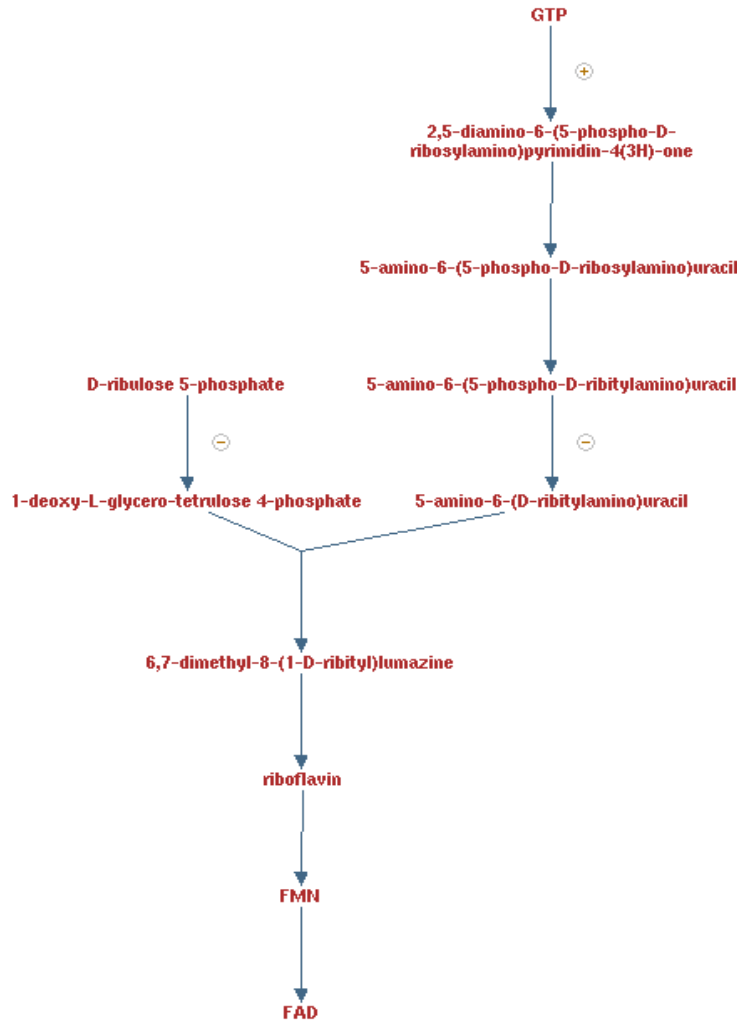


Figura 1. Via da Biossíntese da Flavina

Análise de homologias por Blast

Neste ponto, o objetivo era automatizar o processo de análise do Blast das proteínas, recorrendo à ferramenta Blast presente no package Biopython. Esta ferramenta permitiu obter os resultados deste processo em formatos XML e texto.

Uma homologia realmente sólida requer especial atenção ao comportamento de três fatores aquando a seleção de alinhamentos: o *e-value*, o *query cover* e o *ident*. O *e-value* deverá ser o mais próximo possível de zero, enquanto que ambos o *query cover* e o *ident* deverão tender para 100%.

Uma análise do ficheiro resultado do Blast, em XML, permite, entre outros, seleccionar os alinhamentos com um menor *e-value* de uma forma automatizada. Contudo, não é possível fazer o mesmo com o ficheiro de texto usando python, apesar de conter os valores *e-value* e *ident* de cada alinhamento.

Estas restrições não permitem automatizar uma análise dos resultados do Blast usando os três parâmetros necessários para uma homologia sólida. Assim, só foi considerado o ficheiro XML, que contém apenas o *e-value* dos alinhamentos, para fazer a análise com recurso ao Biopython. Os

alinhamentos com um *e-value* menor que 0,01 foram retirados de cada ficheiro resultado em XML, guardando a sequência alinhada, o *score* e o *e-value* num outro ficheiro.

Uma análise manual deste ficheiro com os alinhamentos retirou os cinco resultados com o menor *e-value* que correspondessem a espécies de *Legionella* diferentes da estudada. Contudo, como o grau de confiança nesta análise é baixo, pois só teve em consideração os valores de *e-value*, estes resultados foram comparados com os resultados do Blast usando o site do NCBI, sendo que só se detetou uma pequena diferença nos resultados da proteína codificada pelo gene RibE.

Por fim, através de pesquisas no site da Uniprot foram retiradas as funções das proteínas selecionadas anteriormente, de forma a comparar a sua função com a das proteínas estudadas. Uma vez que as sequências têm uma grande homologia, é expectável que as suas funções sejam iguais ou bastante semelhantes.

- **Gene ribF**

A proteína codificada por este gene é a riboflavin biosynthesis RibF, cujas funções são atividade catalítica, biossíntese de FAD e biossíntese de FMN. As proteínas e respetivos organismos com a maior homologia com a proteína em estudo foram ordenadas por ordem decrescente:

- bifunctional riboflavin kinase/FAD synthetase* da *Legionella norrlandica*, cujas funções são atividade catalítica, biossíntese de FAD e biossíntese de FMN;
- bifunctional riboflavin kinase/FAD synthetase* da *Legionella shakespearei*, cujas funções são atividade catalítica, biossíntese de FAD e biossíntese de FMN;
- bifunctional riboflavin kinase/FAD synthetase* da *Legionella anisa*, cujas funções são atividade catalítica, biossíntese de FAD e biossíntese de FMN;
- bifunctional riboflavin kinase/FAD synthetase* da *Legionella drancourtii*, cujas funções não se encontram descritas na Uniprot;
- bifunctional riboflavin kinase/FAD synthetase* da *Legionella fallonii*, cujas funções são atividade catalítica, biossíntese de FAD e biossíntese de FMN.

- **Gene ribD**

A proteína codificada por este gene é a riboflavin biosynthesis protein RibD, cujas funções são ser cofator e a biossíntese da riboflavina. As proteínas e respetivos organismos com a maior homologia com a proteína em estudo foram ordenadas por ordem decrescente:

- bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase* RibD da *Legionella norrlandica*, cujas funções são ser cofator e a biossíntese da riboflavina;
- bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase* RibD da *Legionella sainthelensi*, cujas funções não se encontram descritas na Uniprot;
- bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase* RibD da *Legionella gratiana*, cujas funções são ser cofator e a biossíntese da riboflavina;

- iv. bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase RibD da *Legionella santicrucis*, cujas funções são ser cofator e a biossíntese da riboflavina;
- v. bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase RibD da *Legionella longbeachae*, cujas funções são ser cofator e a biossíntese da riboflavina.

- **Gene ribE**

A proteína codificada por este gene é a riboflavin synthase subunit alfa, cuja função é a síntese de riboflavina. As proteínas e respetivos organismos com a maior homologia com a proteína em estudo foram ordenadas por ordem decrescente:

- i. riboflavin synthase da *Legionella norrlandica*, cujas funções não se encontram descritas na Uniprot;
- ii. riboflavin synthase da *Legionella moravica*, cuja função é a síntese de riboflavina;
- iii. riboflavin synthase da *Legionella waltersii*, cuja função é a síntese de riboflavina;
- iv. riboflavin synthase da *Legionella worsleiensis*, cuja função é a síntese de riboflavina;
- v. riboflavin synthase da *Legionella shakespearei*, cujas funções não se encontram descritas na Uniprot.

- **Gene ribA**

A proteína codificada por este gene é a riboflavin biosynthesis protein RibA, cujas funções são atividade catalítica, ser cofator e biossíntese de riboflavina. As proteínas e respetivos organismos com a maior homologia com a proteína em estudo foram ordenadas por ordem decrescente:

- i. bifunctional 3,4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II da *Legionella fallonii*, cujas funções são atividade catalítica, ser cofator e biossíntese de riboflavina;
- ii. bifunctional 3,4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II da *Legionella quateirensis*, cujas funções são atividade catalítica, ser cofator e biossíntese de riboflavina;
- iii. bifunctional 3,4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II da *Legionella steigerwaltii*, cujas funções são atividade catalítica, ser cofator e biossíntese de riboflavina;
- iv. bifunctional 3,4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II da *Legionella moravica*, cujas funções são atividade catalítica, ser cofator e biossíntese de riboflavina;
- v. bifunctional 3,4-dihydroxy-2-butanone-4-phosphate synthase/GTP cyclohydrolase II da *Legionella cherrii*, cujas funções não se encontram descritas na Uniprot.

- **Gene ribH**

A proteína codificada por este gene é a 6,7-dimethyl-8-ribityllumazine synthase, cujas funções são atividade catalítica e biossíntese de riboflavina. As proteínas e respetivos organismos com a maior homologia com a proteína em estudo foram ordenadas por ordem decrescente:

- 6,7-dimethyl-8-ribityllumazine synthase da *Legionella fallonii*, cujas funções são atividade catalítica e biossíntese de riboflavina;
- 6,7-dimethyl-8-ribityllumazine synthase da *Legionella quateirensis*, cujas funções são atividade catalítica e biossíntese de riboflavina;
- 6,7-dimethyl-8-ribityllumazine synthase da *Legionella moravica*, cujas funções são atividade catalítica e biossíntese de riboflavina;
- 6,7-dimethyl-8-ribityllumazine synthase da *Legionella waltersii*, cujas funções são atividade catalítica e biossíntese de riboflavina;
- 6,7-dimethyl-8-ribityllumazine synthase da *Legionella shakespearei*, cujas funções não se encontram descritas na Uniprot.

Alinhamento múltiplo e filogenia

O alinhamento múltiplo é importante pois permite analisar zonas de maior e menor conservação de forma a conseguir identificar-se domínios conservados. A filogenia é também de grande relevância uma vez que permite associar os diferentes organismos evolutivamente ao analisar, por exemplo, a transferência horizontal de genes. A transferência horizontal de genes é o movimento de informação genética entre organismos, que promove, por exemplo, a dispersão de resistência a antibióticos entre bactérias, contribuindo para a evolução do patogénio.

Começou-se por realizar um Blast (no NCBI) para cada uma das cinco sequências proteicas. Após os resultados, para cada um dos casos selecionaram-se cinco estirpes diferentes, sendo que foram escolhidas as com maior *ident* (o mais próximo de 100% possível) e menor *e-value* (o mais próximo de 0 possível), tendo em conta também o *query cover*. De seguida, acedeu-se à plataforma Clustal Omega, que permite realizar o alinhamento múltiplo das sequências. Desta forma, procedeu-se ao alinhamento múltiplo de cada conjunto de proteínas homologas, obtendo-se os resultados dos alinhamentos em formato *Clustal* e a informação para posterior criação da árvore em formato Newick. Através de BioPython, utilizou-se a função `Philo.read()` que permite ler o ficheiro em formato newick e, também, a `Philo.draw_ascii()` que permite desenhar a árvore.

- **RibD - lpg1177**

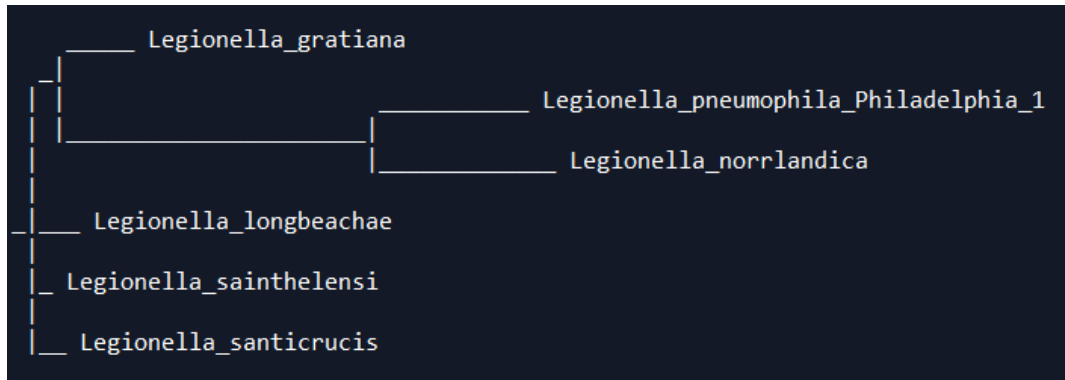


Figura 2. Árvore filogenética obtida através de Biopython

Após análise destes resultados, podemos verificar que as proteínas em questão que apresentam uma maior semelhança evolutiva são aquelas pertencentes às estirpes *Legionella sainthelensi*, *Legionella santicrucis* e *Legionella longbeachae*, pelo que se pode inferir que estas proteínas serão as mais semelhantes funcional e estruturalmente.

Quanto à proteína da *Legionella pneumophila Philadelphia 1*, esta tem mais semelhanças com a proteína da *Legionella norrlandica*.

- **RibE – lpg1178**

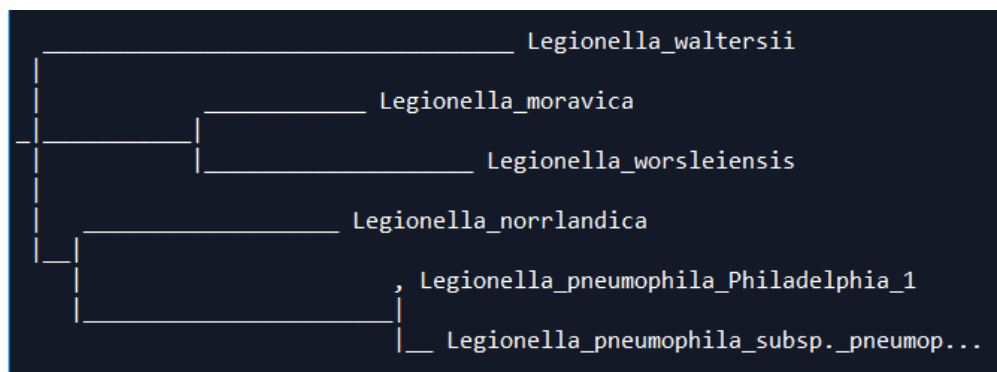


Figura 3. Árvore filogenética obtida através de Biopython

Na análise desta árvore filogenética podemos ver que existe uma grande proximidade entre a *Legionella pneumophila subsp. Pneumophila* e a *Legionella pneumophila Philadelphia 1* isto pode ser explicado uma vez que são estirpes da mesma subespécie, sendo possível inferir que apresentam estrutural e funcionalmente bastante analogia.

As restantes estão filogeneticamente mais afastadas, o que pode ser explicado por serem de espécies diferentes.

- **RibA – lpg1179**

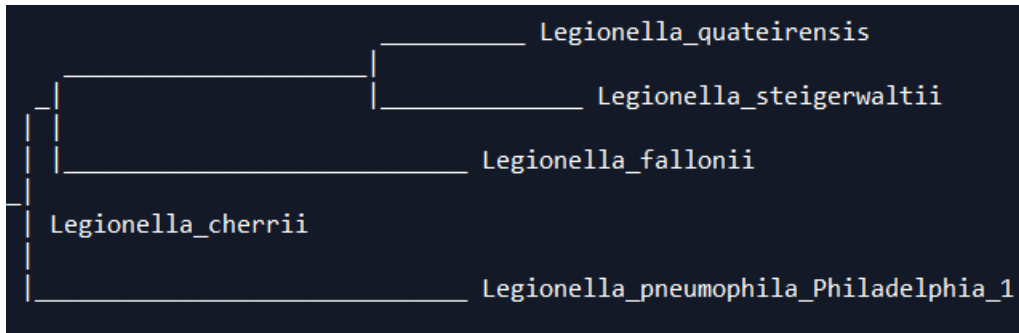


Figura 4. Árvore filogenética obtida através de Biopython

Após a análise desta árvore filogenética, podemos concluir que as proteínas das estirpes *Legionella quateirensis* e *Legionella steigerwaltii* são as mais semelhantes, pelo que serão as mais parecidas funcional e estruturalmente.

Quanto à proteína da *Legionella pneumophila Philadelphia 1*, esta tem mais semelhanças com a proteína de estirpe *Legionella cherrii*.

- **RibH – lpg1180**

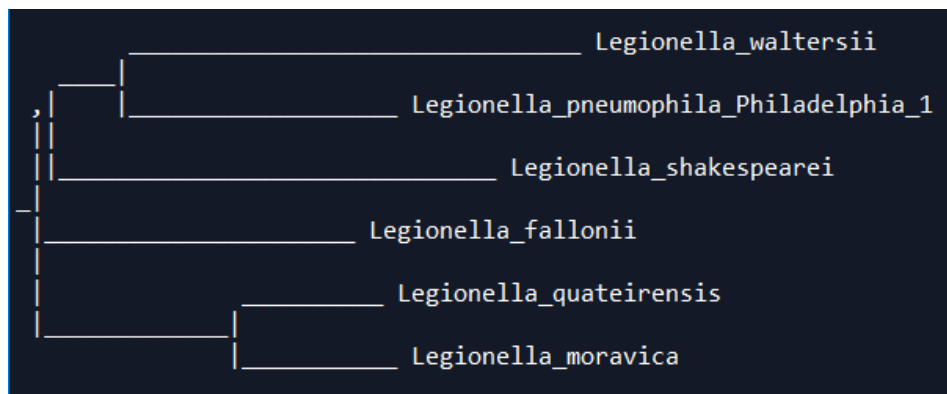


Figura 5. Árvore filogenética obtida através de Biopython

Após a análise desta árvore filogenética podemos ver que as espécies que apresentam maior semelhança entre si são a *Legionella quateirensis* e a *Legionella moravica*, quanto à *Legionella pneumophila Philadelphia 1* apresenta maior analogia com a *Legionella waltersii*, sendo possível concluir que apresenta maior semelhança com esta espécie, a nível estrutural e funcional, havendo por isso maior conservação dos domínios.

- **RibF – lpg0936**

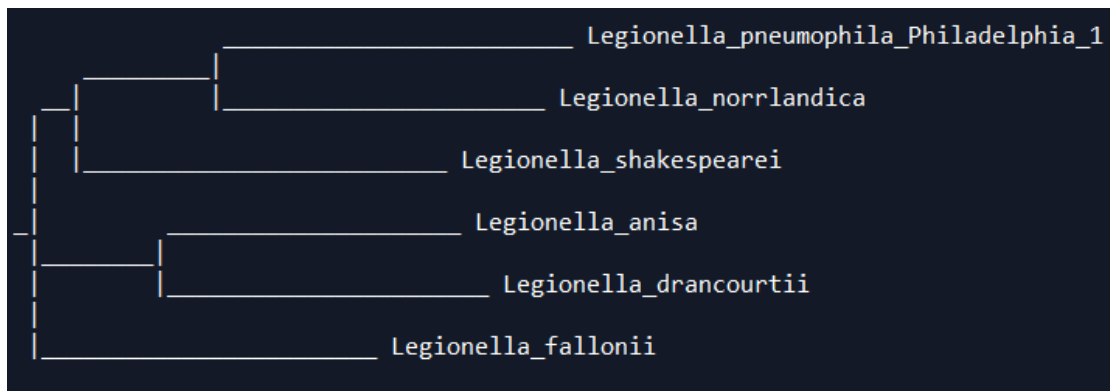


Figura 6. Árvore Filogenética obtida através de Biopython

Após a análise desta árvore filogenética podemos concluir que as proteínas da *Legionella pneumophila Philadelphia I* e a *Legionella norrlandica* são as mais próximas evolutivamente. Desta forma estas espécies serão as mais semelhantes estrutural e funcionalmente entre si para esta proteína.

Regulação da via metabólica

A riboflavina (ou vitamina B2) é um componente essencial da dieta humana e animal, uma vez que serve como precursor para dois cofatores de flavina bastante importantes e que estão envolvidos no metabolismo oxidativo:

- FMN – mononucleótido de flavina;
- FAD – dinucleótido de flavina-adenina.

A regulação da riboflavina ocorre ao nível da atividade enzimática e da sua síntese, ocorrendo principalmente ao nível da produção das enzimas da sua via metabólica.

Embora se saiba muito pouco em relação a estes mecanismos de regulação e não ocorram de igual forma em todos os organismos, um destes mecanismos é a inibição por feedback da enzima GTP Cyclohydrolase II e da enzima GTP Cyclohydrolase I. Relativamente à enzima GTP Cyclohydrolase II não se conhece o mecanismo exato através do qual o feedback acontece, no entanto há estudos que indicam que poderá ser a concentração de FMN que inibe a enzima, sendo que quando a concentração deste composto é elevada há inibição da enzima, podendo este ser o passo limitante da via metabólica. Outro mecanismo de feedback sugerido baseia-se na inibição da enzima 6,7-dimethyl-8-ribityllumazine, que é inibida pela concentração elevada de riboflavina, inibição esta que vai parar a via metabólica e parar a produção deste composto^{7,8}.

Outro importante mecanismo de regulação é através do “Riboswitch” presente no operão da Riboflavina, também conhecido como operão rib (constituído por 5 genes ribGBAHT), que é responsável pela síntese das enzimas envolvidas na produção de riboflavina. Este operão tem uma zona denominada RFN, elemento que consiste numa estrutura de RNA que altera a sua conformação tendo em conta os elementos que tem a si ligados. Um exemplo deste mecanismo consiste na ligação do FMN ao RFN o que vai alterar a sua conformação, alteração esta que vai pedir a transcrição normal

dos genes do operão e desta forma impedir a normal produção de riboflavina. Quando a concentração de FMN diminui esta deixa de estar ligada ao RFN, a conformação deste volta ao normal e desta forma ocorre a transcrição génica normal⁹.

Fatores de virulência

Para a pesquisa destes fatores procedemos a uma procura na base de dados “Virulence Factors of Pathogenic Bacteria”, tendo sido pesquisadas as proteínas relativamente à via que nos foi atribuída. Os fatores de virulência associados à via referida e ao microrganismo *Legionella pneumophila* foram os seguintes:

- **Dot/Icm:** permite a sobrevivência e crescimento em macrófagos prevenindo a acidificação dos fagossomas e a fusão lisossomal;
- **FeoAB:** versão mutante do gene feoB aumentando o crescimento intracelular;
- **Flagella:** aumenta a capacidade de invasão do microrganismo;
- **FrgA:** aumenta a capacidade de replicação extracelular, infeção e virulência;
- **LigA:** fator de virulência relacionado com a replicação em protozoários;
- **pmiA:** envolvido na sobrevivência e replicação dentro de macrófagos e protozoários;
- **Pht:** permite ao microrganismo ser nutrido dentro dos fagossomas.

Resistência a antibióticos

Relativamente à pesquisa da existência de resistência a antibióticos relacionada com as proteínas da nossa via metabólica efetuou-se a procura dos mesmos na base de dados “CARD”. Os dados obtidos foram de que não existe resistência a antibióticos relacionada com estas proteínas e com o microrganismo que estamos a estudar em específico.

Domínios de proteínas

De forma a completar informação acerca das funções das proteínas de interesse, recorreu-se à análise dos domínios dessas proteínas através da pesquisa em bases de dados de domínios de proteínas como a NCBI CDD (*conserved domain database*).

Assim, inseriu-se a sequência da proteína *riboflavin biosynthesis protein RibD* nesta base de dados, tendo-se obtido como resultado o domínio MafB18-deam super family, pelo que podemos inferir que a proteína tem uma função enzimática, tal como anteriormente verificado em outras bases de dados como a Uniprot. Mais especificamente, podemos concluir que esta proteína tem uma função de deaminase, estando por isso envolvida na remoção de grupos amina de determinadas moléculas. Podemos fazer esta inferência, uma vez que o *e-value* obtido é bastante baixo (2.29×10^{-69}) pelo que se podem considerar os dados obtidos como confiáveis.

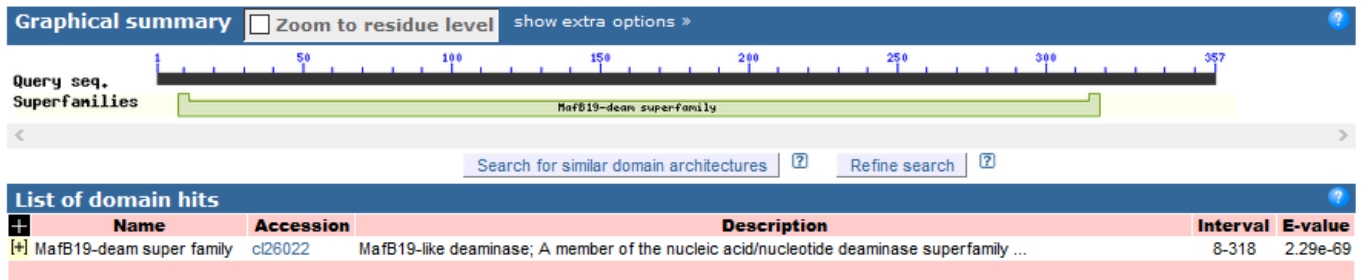


Figura 7. Domínio presente na proteína riboflavin biosynthesis protein RibD

De seguida, para análise da proteína *riboflavin synthase subunit alpha*, inseriu-se a sequência na base de dados, tendo-se obtido como resultado o domínio PRK09289, podendo-se concluir que a proteína terá a função de catalisar reações envolvidas na biossíntese de riboflavina, função já esperada e concordante com os resultados obtidos noutras bases de dados. Os resultados podem ser considerados fiáveis uma vez que o *e-value* é, mais uma vez, bastante reduzido ($2.09e-84$).

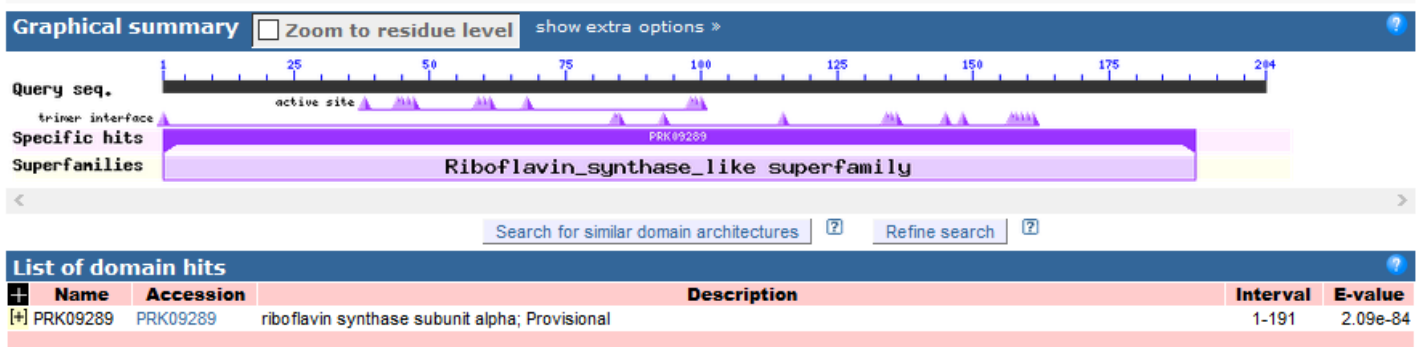


Figura 8. Domínio presente na proteína riboflavin synthase subunit alpha

Para análise da proteína *riboflavin biosynthesis protein RibA*, inseriu-se a sequência na base de dados, tendo-se obtido como resultado o domínio DHBP_synthase super family. Podemos, assim, inferir que esta proteína tem uma função enzimática de síntese, mais especificamente, serve como precursor biossintético do *xylene ring* da riboflavina. O *e-value* é 0, pelo que os resultados podem ser considerados fiáveis.

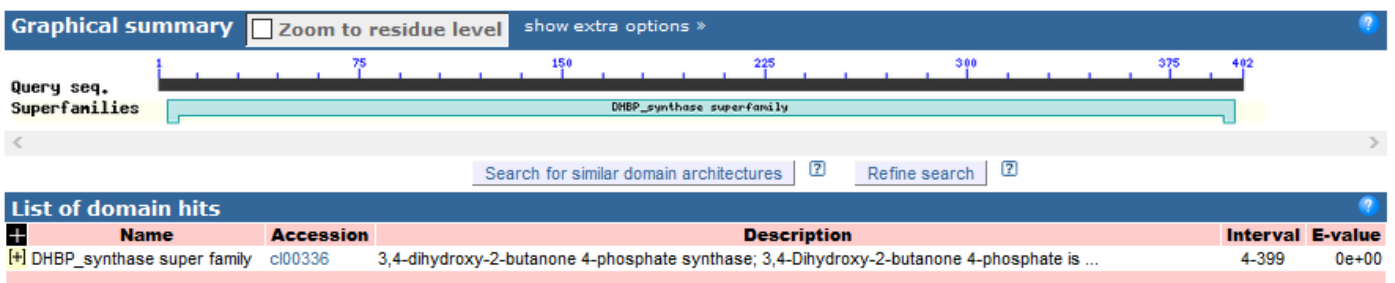


Figura 9. Domínio presente na proteína riboflavin biosynthesis protein RibA

Para análise da proteína 6,7-dimethyl-8-ribityllumazine synthase, inseriu-se a sequência na base de dados, tendo-se obtido como resultado o domínio Lumazine_synthase-I, podendo-se inferir que

esta proteína tem função de síntese, catalisando um dos últimos passos da biossíntese de riboflavina. Mais uma vez, o *e-value* é bastante baixo ($2.56e-71$), pelo que os resultados podem ser considerados confiáveis.

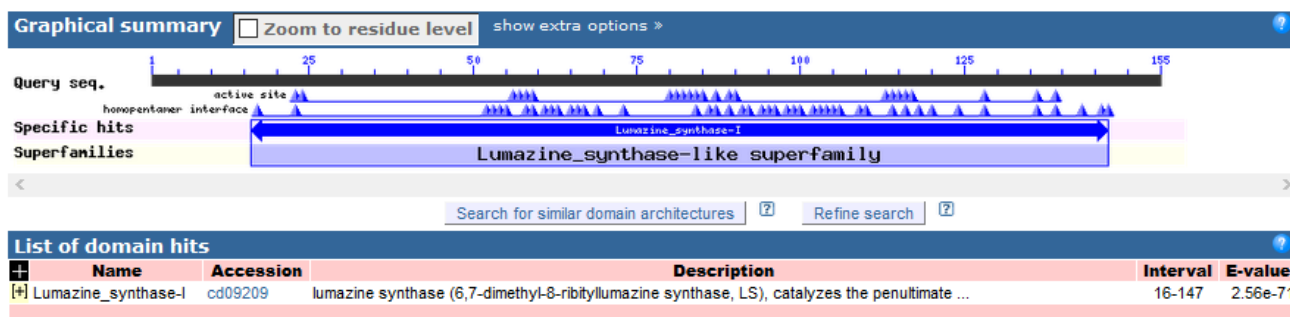


Figura 10. Domínio presente na proteína 6,7-dimethyl-8-ribitylumazine synthase

Para análise da proteína *riboflavin biosynthesis RibF*, inseriu-se a sequência na base de dados, tendo-se obtido como resultado o domínio PRK05627. Assim, podemos concluir que esta proteína terá a função de riboflavina cinase, responsável pela conversão da riboflavina nos cofatores FMN e FAD, tendo, assim, uma função importante no metabolismo de flavina. Mais uma vez, o *e-value* é bastante reduzido ($3.45e-142$) pelo que se pode considerar os resultados como confiáveis.

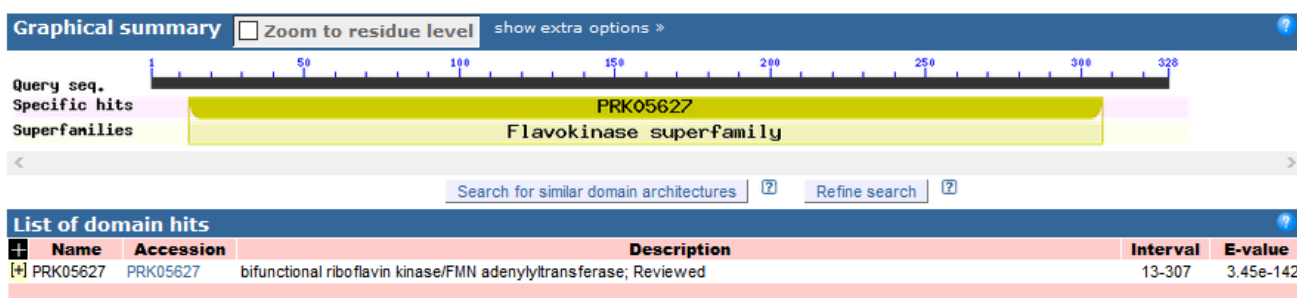


Figura 11. Riboflavin biosynthesis RibF

Análise dos *motifs* e estrutura das proteínas

Quanto à análise dos *motifs* das proteínas que se encontram na tabela após terem sido corridas individual e conjuntamente na base de dados da "ScanProsite" e após ter sido feito o seu BLAST podemos confirmar que os dados obtidos estão de acordo com os analisados anteriormente. Quanto à função esperada das proteínas também podemos inferir que está de acordo com as funções encontradas anteriormente, uma vez que nestas proteínas podemos encontrar domínios conservados, quer sejam centros catalíticos, como o CMP/dCMP-type deaminase zonas onde há a presença de elementos metálicos, nomeadamente o Zinco, ou zonas onde ocorre a ligação das proteínas em que a enzima vai atuar, como o Riboflavin synthase alpha chain lumazine-binding repeat profile (Figura 12).

PS51747 CYT_DCMP_DEAMINASES_2 Cytidine and deoxycytidylate deaminases domain profile :

1 - 123: score = 23.769

-MHEQFLLAALAEQAKLGRGFCapnPSVGAVAVQNGNIQAHR----GAGTPHAEQLLL
SQI-----PPQTPGVTLYVTLPCNHwgktpCVDAIINHGIEEVVFSYFDPNpivaKSN
SSEKLREHGIQVR

Predicted feature:

DOMAIN 1 123 CMP/dCMP-type deaminase [condition: none]

Absent features:

METAL	49	Zinc; catalytic	[condition not true: H and <54=E> and not <AnyFeature:MF_00972>]	[group: 1]
METAL	49	Zinc	[condition not true: H and not(<54=E> and not <AnyFeature:MF_00972>) and (not <AnyFeature:MF_00972>)]	[group: 2]
ACT_SITE	51	Proton donor	[condition not true: E and <54=E> and not <AnyFeature:MF_00972>]	[group: 1]
METAL	74	Zinc; catalytic	[condition not true: C and <54=E> and not <AnyFeature:MF_00972>]	[group: 1]
METAL	74	Zinc	[condition not true: C and not(<54=E> and not <AnyFeature:MF_00972>) and (not <AnyFeature:MF_00972>)]	[group: 2]
METAL	83	Zinc; catalytic	[condition not true: C and <54=E> and not <AnyFeature:MF_00972>]	[group: 1]
METAL	83	Zinc	[condition not true: C and not(<54=E> and not <AnyFeature:MF_00972>) and (not <AnyFeature:MF_00972>)]	[group: 2]

ruler:

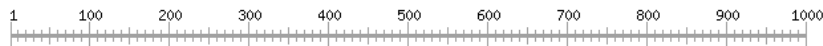


Figura 12. Motifs das proteínas

Esta informação é ainda posteriormente validada através da análise das proteínas na base de dados *RCSB PDB* onde se encontra informação relativa à estrutura, nomeadamente domínios e subunidades da proteína, e composição, não só da macromolécula no geral como de outras moléculas mais pequenas que se encontram na proteína e que podem funcionar por exemplo como ligandos. É importante referir que na base de dados PDB as proteínas encontradas são as por nós estudadas embora não sejam da mesma espécie, sendo por isso utilizados dados referentes à *E. coli*, desta forma os dados recolhidos desta base de dados poderão não ser os mais adequados. Após esta análise conseguimos confirmar a informação previamente recolhida de outras bases de dados e do BLAST podendo ser visível a presença dos centros catalíticos (relativos aos grupos de fosfato presentes na proteína) e elementos metálicos como o Zinco. Tendo em conta toda a análise feita nas bases de dados anteriormente referidas podemos concluir que a função enzimática da proteína está de acordo com o esperado e previamente inferido.

Relativamente à análise dos fatores de transcrição do nosso organismo e da nossa via metabólica, aquando da pesquisa na base de dados “DBD: Transcription factor prediction database” não nos foi possível encontrar dados relativamente à via metabólica da biossíntese da flavina.

Referências

1. Cho MS, Ahn T-Y, Joh K, Lee ES, Park DS. Improved PCR assay for the species-specific *identification* and quantitation of *Legionella pneumophila* in water. *Appl Microbiol Biotechnol* [Internet]. 2015 Nov 4 [cited 2018 Jan 26];99(21):9227–36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26142386>
2. *Legionella* Infection: Background, Pathophysiology, Epidemiology [Internet]. [cited 2018 Jan 26]. Available from: <https://emedicine.medscape.com/article/965492-overview>
3. Cao B, Tian Z, Wang S, Zhu Z, Sun Y, Feng L, et al. Structural comparison of O-antigen gene clusters of *Legionella pneumophila* and its application of a serogroup-specific multiplex PCR assay. *Antonie Van Leeuwenhoek* [Internet]. 2015 Dec 28 [cited 2018 Jan 26];108(6):1405–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26415652>
4. Haase I, Sarge S, Illarionov B, Laudert D, Hohmann H-P, Bacher A, et al. Enzymes from the Haloacid Dehalogenase (HAD) Superfamily Catalyze the Elusive Dephosphorylation Step of Riboflavin Biosynthesis. *ChemBioChem* [Internet]. 2013 Nov 25 [cited 2018 Jan 29];14(17):2272–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24123841>
5. Bacher A, Eberhardt S, Fischer M, Kis K, Richter G. BIOSYNTHESIS OF VITAMIN B₂ (RIBOFLAVIN). *Annu Rev Nutr* [Internet]. 2000 Jul [cited 2018 Jan 29];20(1):153–67. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10940330>
6. Bacher A, Richter G, Ritz H, Eberhardt S, Fischer M, Krieger C. Biosynthesis of riboflavin: GTP cyclohydrolase II, deaminase, and reductase. *Methods Enzymol* [Internet]. 1997 [cited 2018 Jan 29];280:382–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9211333>
7. WILSON AC, PARDEE AB. Regulation of Flavin Synthesis by *Escherichia coli*. *J Gen Microbiol* [Internet]. 1962 Jun 1 [cited 2018 Jan 26];28(2):283–303. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14007303>
8. Abbas CA, Sibirny AA. Genetic Control of Biosynthesis and Transport of Riboflavin and Flavin Nucleotides and Construction of Robust Biotechnological Producers. *Microbiol Mol Biol Rev* [Internet]. 2011 Jun 1 [cited 2018 Jan 26];75(2):321–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21646432>
9. Serganov A, Huang L, Patel DJ. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* [Internet]. 2009 Mar 25 [cited 2018 Jan 26];458(7235):233–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19169240>