

Modelado molecular y predicción funcional de la proteína bll4781 de *Bradyrhizobium diazoefficiens* (USDA 110)

Damián Brignoli

Cátedra de Bioinformática, Área de Biotecnología y Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina.

RESUMEN

La proteína denominada bll4781 pertenece a la bacteria de suelo *Bradyrhizobium diazoefficiens* (USDA 110), rizobio de importancia agronómica por su capacidad de fijar N atmosférico en las raíces de leguminosas. La secuencia de esta proteína presenta una longitud de 100 aa, no presenta regiones transmembrana y presenta regiones desordenadas entre las posiciones 1-15 y 83-100, respectivamente. Se realizó una búsqueda de homología por diversos métodos y programas, y se encontraron homólogos remotos con estructura conocida, y sólo uno de ellos presentó la mayor significancia estadística, una proteína piruvato dehidrogenasa perteneciente a la bacteria fotosintética *Rhodospseudomonas palustris* CGA009. Las relaciones filogenéticas muestran mayor cercanía evolutiva con varias especies de *Bradyrhizobium*, formando un clado de 8 especies donde se encuentra el representante de la proteína de estructura conocida. Para predecir la posible función biológica de la proteína se realizó una búsqueda exhaustiva en diferentes bases de datos realizando el seguimiento del nº EC, encontrando que la proteína de estudio participa en las rutas metabólicas de la fermentación del acetato y metabolismo del piruvato, realizando la posible función molecular de catalizar la reacción entre el piruvato y una molécula de ubiquinona, dando como resultado acetato, ubiquinol y dióxido de carbono.

PALABRAS CLAVE: piruvato deshidrogenasa, función, bioinformática

INTRODUCCIÓN

La proteína denominada bll4781 pertenece a la bacteria de suelo *Bradyrhizobium diazoefficiens* (USDA 110), rizobio Gram negativo capaz de fijar el N atmosférico en órganos especializados de las raíces de las leguminosas denominados nódulos. La simbiosis rizobio-leguminosa es de gran importancia agronómica y permite que la planta crezca con éxito en ausencia de fertilizantes nitrogenados suministrados de forma externa. Esta bacteria, junto a otras de su misma especie, como por ejemplo *Bradyrhizobium elkanii* y *Bradyrhizobium japonicum* (referidas conjuntamente como *Bradyrhizobium spp.*), forman parte del grupo de microorganismos conocidos como PGPR (Plant growth-promoting rhizobacteria) debido a la propiedad mencionada anteriormente. Gracias a esta propiedad, se utilizan como componentes principales de los inoculantes para soja, los cuales hoy en día son ampliamente utilizados y la práctica de la inoculación está ampliamente extendida.

El genoma de esta bacteria es un cromosoma circular de 9.105.828 bp de longitud que comprende 8317 posibles genes codificantes de proteínas, de los cuales el 52% muestra una secuencia similar a genes de función conocida y un 30% a genes hipotéticos. El restante 18% no muestra una similitud aparente a los restantes genes reportados.

El objetivo general de este trabajo es predecir la posible función biológica de la proteína en cuestión a partir de conocer su posible estructura utilizando algunas de las herramientas brindadas durante el curso.

MÉTODOS Y RESULTADOS

En particular, la secuencia de aminoácidos de la proteína bll4781 tiene una longitud de 100 aa. A continuación se muestra la secuencia de aa de dicha proteína en formato FASTA:

```
>tr|Q89KW9|Q89KW9_BRADU Bll4781 protein OS=Bradyrhizobium diazoefficiens (strain JCM 10833 / BCRC 13528 / IAM 13628 / NBRC 14792 / USDA 110) OX=224911 GN=bll4781 PE=4 SV=1  
MAGPKEQPLPPDVVTREDAVEILRVFVLDGGLSMAFQRAFEEDMWGLLLVDLARHAARAYARESEYTEEDALSRIEMFQA  
EIERPTDTGTTTPRGKGH
```

En la búsqueda de homólogos cercanos y remotos se hizo uso de los programas BLAST y PSI-BLAST, de la base de datos del NCBI (National Center for Biotechnology Information), así como de los programas HMMER y HHpred, del servidor Toolkit.

Se realizó una primera búsqueda con BLAST con un parámetro de 1000 secuencias iniciales, descartando aquellas secuencias con porcentajes de identidad menores al 30%, con lo cual quedaron establecidas 521 secuencias como posibles homólogos cercanos. A continuación, se realizó una búsqueda de homólogos remotos por PSI-BLAST, también con un parámetro inicial de 1000 secuencias, arrojando en la primera iteración un resultado de 765 secuencias, luego se realizaron 3 iteraciones más donde el programa encuentra y añade secuencias con menos del 30% de identidad a partir de la tercera iteración. Para un análisis de mayor sensibilidad, se utilizó el programa JACKHMMER (secuencia-profile) del EMBL-EBI, encontrando 463 "matches" con significancia estadística en la primera iteración. La búsqueda se realizó por base de datos de UniProt, con restricción por taxonomía (Eubacteria) y sin modificación de parámetros numéricos preestablecidos. Por último, se utilizó el servidor Toolkit para correr el programa HHpred seleccionando la base de datos del PDB, con el cual llegamos al grado de mayor sensibilidad (profile-profile) en la búsqueda de homología remota. El programa encontró 27 "hits", de los cuales el primero de ellos tuvo la mayor significancia estadística (E-value 7,9 e-35), Score (192,7) y porcentaje de Identidad (%76).

La distribución taxonómica de los resultados obtenidos por BLAST y PSI-BLAST muestra que el mayor número de microorganismos representados corresponden a proteobacterias, más precisamente a alfa proteobacterias, seguidas de gamma y betaproteobacterias, encontrándose como más distantes a organismos como *Calditrichaeta bacterium* y un nemátode (*Diploscapter pachys*) (Figura S1).

Por otra parte se llevó a cabo la predicción de la estructura secundaria, utilizando los programas Quick2D (Figura 1) y PSIPRED, de sus respectivos servidores online comparando sus resultados. Para visualizar si la proteína posee segmentos transmembrana, se utilizó el servidor Phobius (Figura 2.a). Así mismo para conocer las posibles regiones desordenadas, se utilizó el servidor IUpred2A (Figura 2.b). La búsqueda de dominios se llevó a cabo mediante la base de datos de Pfam. La proteína en estudio presenta un dominio que abarca desde la posición 6 a la posición 88 en la secuencia. El dominio se denomina DUF5076, y en la base de datos mencionada se pueden observar las secuencias (153), las arquitecturas (3), las especies (128) y las estructuras (4) asociadas al mismo.

Como se observa en la figura 2.a, la probabilidad asignada para conocer si la proteína posee segmentos transmembrana es muy baja, y ocupa aproximadamente desde la posición 20 a 40 de la secuencia de aa. El programa Quick2D al tener integrada la predicción de segmentos transmembrana, y al ser la probabilidad

búsqueda exhaustiva en diferentes bases de datos para tener una aproximación a conocer la respuesta a la pregunta de ¿qué función cumple esta proteína? La búsqueda se basó en el seguimiento del nº EC (Enzyme Commission), el cual para la proteína estudiada según el servidor ExPASy es el 1.2.2, que refiere a enzimas oxidorreductasas actuantes sobre donantes de grupos aldehídos u oxo con citocromo como aceptor.

Sin embargo, en otras bases de datos como BRENDA, IUBMB Enzyme Nomenclature, IntEnz y Gene Ontology (GO), este EC se muestra como nº 1.2.5.1, con el nombre aceptado de piruvato deshidrogenasa (quinona), donde el cambio se produce en que la enzima actúa sobre donantes de grupos aldehídos u oxo con una quinona o componente similar como aceptor, participando en las rutas metabólicas de la fermentación del acetato y metabolismo del piruvato, realizando la posible función molecular de catalizar la siguiente reacción:

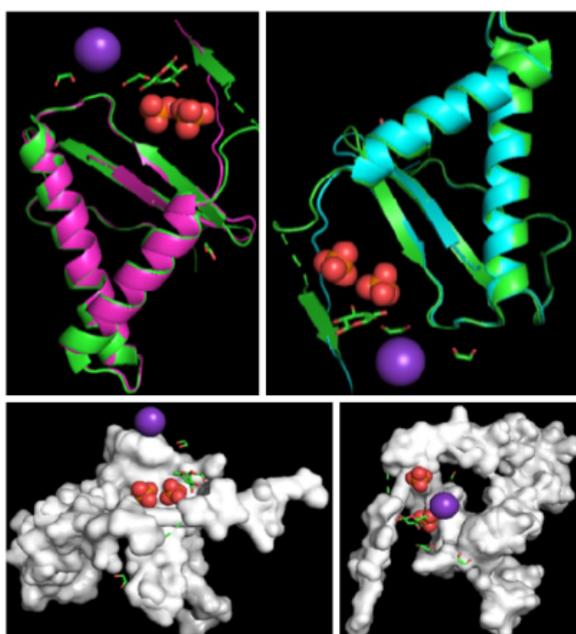
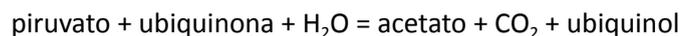


Figura 4. Modelo propuesto para la proteína bll4781 de *Bradyrhizobium diazoefficiens*. Se realizó el modelado por homología utilizando el programa Modeller y la proteína 3IC3 (PDB) como template. Las figuras se obtuvieron utilizando el programa PyMol.

Finalmente, se realizó un estudio filogenético con el objetivo de conocer las relaciones y distancias evolutivas entre las secuencias (especies) seleccionadas para tal fin. Como primer paso para la realización del árbol filogenético, se realizó un alineamiento múltiple entre la secuencia de la proteína de estudio y las secuencias homólogas seleccionadas en BLAST (16 secuencias en total). Se utilizó el servidor T-Coffee para la realización del mismo y para evaluar su calidad. En la figura S5 se observa el alineamiento múltiple obtenido en T-Coffee.

Para la construcción del árbol filogenético se hizo uso del programa PhyML 3.1, utilizando bootstrap para el soporte de los nodos (100 replicantes), y utilizando como árbol inicial un árbol construido por BioNJ (Neighbor-Joining). El mejor modelo evolutivo para la construcción del árbol fue WAG (Whelan & Goldman) según el programa ModelTest (paquete HYPHY) considerando un número de 4 categorías de velocidades de sustitución para la *gamma distribution*. En la siguiente figura (Figura 5) se observa el árbol filogenético obtenido, con los nombres de las respectivas especies, nodos, y longitud de ramas.

CONCLUSIONES Y DISCUSIÓN

La proteína bll4781 perteneciente a la bacteria de suelo Gram negativa *Bradyrhizobium diazoefficiens* presenta una longitud de 100 aa en su secuencia. Al realizar una búsqueda de homólogos, el mayor porcentaje de los mismos presentan longitudes de secuencia muy similares y el dominio DUF5076, de función desconocida. Con el programa HHpred, se encontró que el primer *hit* presentó la mayor significancia estadística y el porcentaje de identidad más alto (%76). Esta búsqueda se realizó contra la base de datos del PDB y la secuencia y estructura asociada demuestra que pertenece a la proteína homotetrámero piruvato dehidrogenasa de la bacteria fotosintética *Rhodopseudomonas palustris* CGA009. A partir del alineamiento realizado entre el *target* y la secuencia del *template* se puede observar un alto grado de conservación, lo que se demuestra en las imágenes del modelado por homología realizado en el programa Modeller.

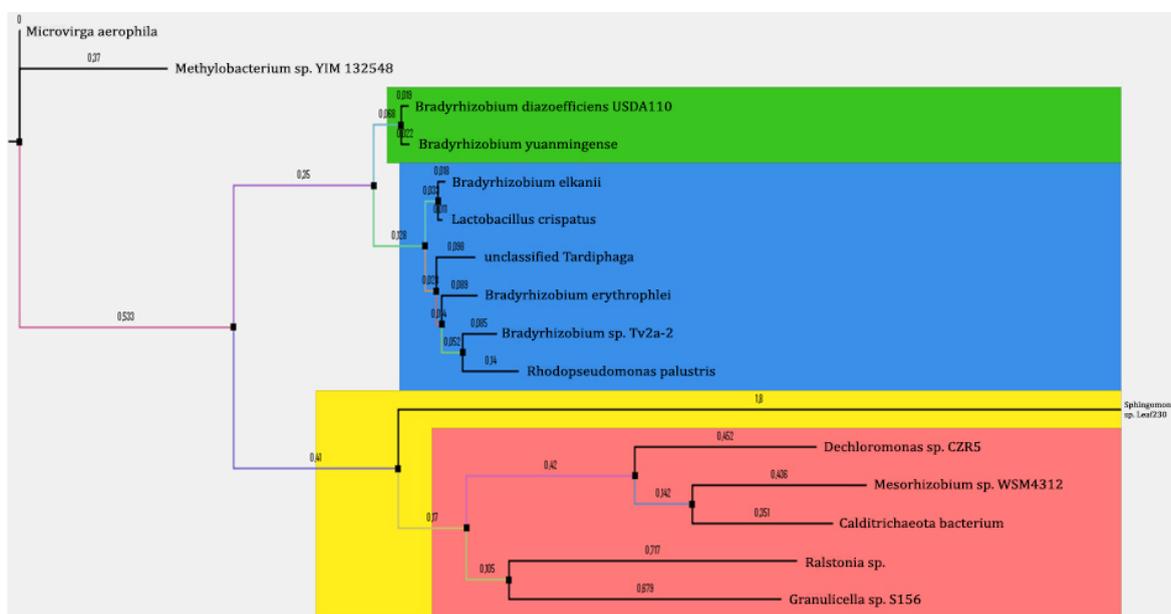


Figura 5. Árbol filogenético de 16 secuencias homólogas a la proteína bll4781. El árbol fue obtenido utilizando PhyML 3.1 y visualizado en FigTree.

La mayor porción de la proteína no presenta regiones desordenadas, las mismas se observan desde las posiciones 1 a 15 y 83 a 100, respectivamente. Así mismo, esta proteína carece de regiones transmembrana.

En las relaciones filogenéticas, se observa que *Bradyrhizobium diazoefficiens* comparte estrecha relación con *Bradyrhizobium yuanmingense* ya que divergieron en conjunto a partir de su ancestro común, formando parte de un clado más divergente y diverso junto a *Bradyrhizobium elkanii*, *Bradyrhizobium erythrophlei*, *Bradyrhizobium sp. Tv2a-2*, *Lactobacillus crispatus*, *Tardiphaga* y *Rhodopseudomonas palustris*.

Ante la ausencia de datos que puedan confirmar la función de la proteína estudiada, se le asignó la posible función molecular de catalizar la reacción química entre el piruvato y ubiquinona dando como producto final acetato, ubiquinol y dióxido de carbono.

BIBLIOGRAFÍA

A. Šali & T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.

[BRENDA, the Elixir core data resource in 2021: new developments and updates.](#) Chang A., Jeske L., Ulbrich S., Hofmann J., Koblitz J., Schomburg I., Neumann-Schaal M., Jahn D., Schomburg D., *Nucleic Acids Res.*, 49:D498-D508

Brooksbank, C., Camon, E., Harris, M. A., Magrane, M., Martin, M. J., Mulder, N., ... & Cameron, G. 2003. The European Bioinformatics Institute's data resources. *Nucleic acids research*, 31(1), 43-50.

CATH: increased structural coverage of functional space. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, Abbasian M, Le Cornu S, Lam SD, Berka K, Varekova IH, Svobodova R, Lees J, Orengo CA. *Nucleic Acids Res.* 2021. [Pubmed: 33237325](#). [doi: 10.1093/nar/gkaa1079](#)

Davis-Richardson, A. G., Russell, J. T., Dias, R., McKinlay, A. J., Canepa, R., Fagen, J. R., ... & Triplett, E. W. 2016. Integrating DNA methylation and gene expression data in the development of the soybean-Bradyrhizobium N₂-fixing symbiosis. *Frontiers in microbiology*, 7, 518.

Gábor Erdős & Zsuzsanna Dosztányi. 2020. [Analyzing Protein Disorder with IUPred2A](#). *Current Protocols in Bioinformatics* 2020; 70(1):e99

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31:3784-3788

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. 2000. The Protein Data Bank. [Nucleic Acids Research](#), 28: 235-242. Disponible en: <https://www.rcsb.org/>

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. 2005. FFAS03: a server for profile-profile sequence alignments. *Nucl. Acids Res.* 33, W284-W288

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., ... & Tabata, S. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA research*, 9(6), 189-197

Landau M., Mayrose I., Rosenberg Y., Glaser F., Martz E., Pupko T. and Ben-Tal N. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucl. Acids Res.* 33:W299-W302.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2021. <https://www.ncbi.nlm.nih.gov/>

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). 2021. School of Biological and Chemical Sciences, Queen Mary, University of London. <https://www.qmul.ac.uk/sbcs/iubmb/>

[Pfam: The protein families database in 2021](#): J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman. 2020. *Nucleic Acids Research*. doi: 10.1093/nar/gkaa913

Phobius: A combined transmembrane topology and signal peptide predictor. 2021. Centre Stockholm Bioinformatics. Disponible en: <https://phobius.sbc.su.se/>

Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. [Curr Protoc Bioinformatics](#). 2020 Dec;72(1)

[The InterPro protein families and domains database: 20 years on](#). Matthias Blum, Hsin-Yu Chang, Sara Chuguransky, Tiago Grego, Swaathi Kandasamy, Alex Mitchell, Gift Nuka, Typhaine Paysan-Lafosse, Matloob Qureshi, Shriya Raj, Lorna Richardson, Gustavo A Salazar, Lowri Williams, Peer Bork, Alan Bridge, Julian Gough, Daniel H Haft, Ivica Letunic, Aron Marchler-Bauer, Huaiyu Mi, Darren A Natale, Marco Necci, Christine A Orengo, Arun P Pandurangan, Catherine Rivoire, Christian J A Sigrist, Ian Sillitoe, Narmada Thanki, Paul D Thomas, Silvio C E Tosatto, Cathy H Wu, Alex Bateman, Robert D Finn *Nucleic Acids Research*. 2020, gkaa977, PMID: [33156333](#)

The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. <https://www.uniprot.org/>

Wiederstein, M. & Sippl, M. J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 35(suppl_2), W407-W410.