

# Análisis bioinformático de la enzima Timidilato Sintasa de *Candidatus Poseidoniales archaeon*

Bernarda Pschunder

*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

En el presente trabajo se genera un modelo estructural tridimensional de la Timidilato Sintasa de *Candidatus Poseidoniales archaeon* utilizando herramientas bioinformáticas de carácter secuencial, estructural, evolutivo y funcional. El modelo encontrado demuestra tener similitud secuencial y estructural con Timidilato Sintasa provenientes de distintos organismos.

**Palabras clave:** timidilato sintasa, Archea, *Candidatus Poseidoniales*, modelado, bioinformática.

## INTRODUCCIÓN

Las arqueas del Grupo Marino II (MGII) representan el grupo de arqueas planctónicas más abundante de las aguas superficiales del océano pero su descripción se encuentra limitada debido a la falta de organismos cultivables representativos y la falta de secuencias genómicas disponibles. A partir del análisis filogenético comparativo de 270 secuencias se propuso que el MGII es un linaje a nivel de orden para el cual se propuso el nombre *Candidatus Poseidoniales* y que comprende a las familias *Candidatus Poseidonaceae* y *Candidatus Thalassarchaeaceae*, dentro de las cuales se pueden distinguir 21 géneros (Rinke et al. 2019).

La proteína que se analizó en este trabajo (UniProt ID: A0A6V8D8A1) corresponde a la enzima Timidilato Sintasa perteneciente a la especie *Candidatus Poseidoniales archaeon* que aún no se clasifica dentro de ninguna de las dos familias mencionadas anteriormente. Es una enzima metiltransferasa cuya función se encuentra relacionada a la biosíntesis de nucleótidos y específicamente cataliza la reacción de 5,10-metileno-tetrahidrofolato con dUMP para dar 7, 8-dihidrofolato y timidilato. Esta enzima se caracteriza por presentar un único dominio, con una longitud de 264 aminoácidos, de la cual se conoce su secuencia y homólogos con 90% y 50% de identidad como lo muestra la base de datos de UniProt. Por el contrario, no se tiene información sobre su estructura secundaria ni tampoco un modelo tridimensional de la proteína.

A partir de esta breve descripción nos propusimos como objetivos principales confirmar lo que se conoce hasta el momento sobre esta enzima y posteriormente realizar un modelado tridimensional con la intención de inferir los posibles sitios de mayor importancia biológica. Para el cumplimiento de estos objetivos se utilizarán herramientas bioinformáticas de carácter secuencial, estructural y evolutivo

## MÉTODOS Y RESULTADOS

### Búsqueda de homólogos cercanos y remotos utilizando diferentes metodologías

Para el análisis secuencial se utilizan 3 herramientas con fundamentos distintos: en primer instancia se realiza un BLAST, donde la metodología se basa en un análisis secuencia-secuencia, luego se realiza un PsiBlast siguiendo un análisis de tipo secuencia-profile y por último a través de la herramienta HMMER se obtienen resultados con base secuencia-HMM. A continuación, se detallan los parámetros utilizados en cada programa y en la *Tabla 1* se muestra el resumen de los datos más relevantes obtenidos utilizando las diferentes herramientas.

- *Alineamiento secuencia- secuencia con BLAST* utilizando <https://blast.ncbi.nlm.nih.gov/>

Número de secuencias: 1000  
 Base de datos: non-redundant protein sequences (nr)  
 Umbral de E-value: 0.05  
 Tamaño de palabra: 6

- *Alineamiento secuencia-profile con PSI-BLAST:* <https://toolkit.tuebingen.mpg.de/>

Número de secuencias: 1000  
 Base de datos: nr50\_12\_Apr (Default)  
 Valor de corte de E-value para inclusión en el profile: 1E-6

- *Alineamiento secuencia-HMM utilizando Hmmer:* <https://www.ebi.ac.uk/Tools/hmmer>

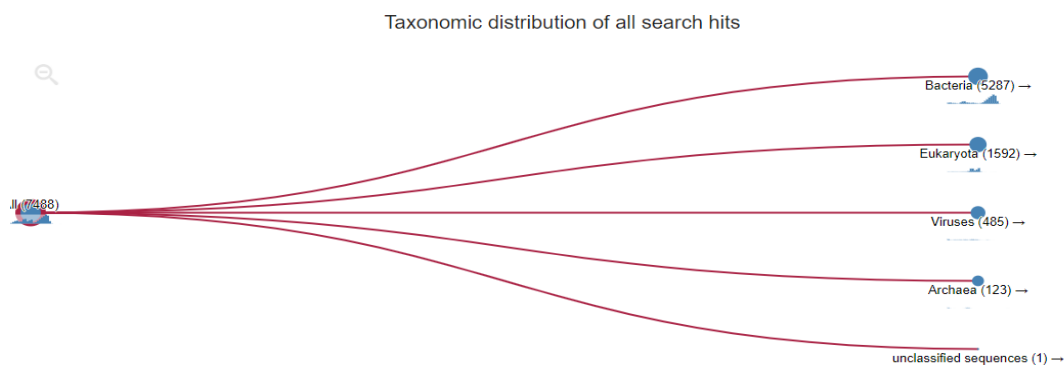
Metodología	BLAST		PsiBlast		HMMER		
	Homólogo más cercano	Homólogo más lejano	Homólogo más lejano (IT1)	Homólogo más lejano (IT2)	Homólogo más lejano (IT1)	Homólogo más lejano (IT2)	Homólogo más lejano (IT3)
<b>%Identidad</b>	100	67,8	15	14	26,9	23,7	16,9
<b>%Similitud</b>	100	86	35	36	77	63,4	57,1
<b>%Gaps</b>	0	0	-	-	-	-	-
<b>E-value</b>	0	3,00E-144	3,00E-08	3,00E-06	1,00E-02	1,00E-02	1,10E-02
<b>Score</b>	555	414	63,2	54	26,6	26,4	26,3

**Tabla 1.** Resultados más relevantes obtenidos en las búsquedas realizadas en BLAST, PsiBlast y HMMER

En primer lugar, mediante el uso de BLAST se encontró que la secuencia con 100% de identidad y 0% de Gaps corresponde a la secuencia de la Timidilato Sintasa de *Candidatus Poseidoniales archaeon*, información que se corresponde con lo encontrado en la base de datos de UniProt. En segundo lugar, si bien el BLAST es un buen comienzo para encontrar los homólogos cercanos a nuestra secuencia de interés, los homólogos más lejanos tienen un %Identidad del 67% con buenos scores y E-values bajos, por lo tanto con esta herramienta y con la limitante de que sólo se pueden analizar 5000 secuencias, se complica la búsqueda de homólogos más lejanos. Para cumplir este objetivo es necesario utilizar otros algoritmos y otras bases de datos.

Mediante el uso de PsiBlast o HMMER encontramos secuencias con un porcentaje de identidad mucho menor y con E-values mayores. En estos casos donde los %Identidad se encuentran por debajo del 20-30%, no se puede deducir a simple vista si las secuencias alineadas son homólogas o no, por lo que sería conveniente realizar otro tipo de estudios como un estudio estructural.

Por otra parte, la herramienta HMMER nos da como resultado una distribución taxonómica de los hits como se muestra en la Figura 1. Esta representación nos muestra la distribución de los hits donde como se puede observar, la mayor proporción corresponde al dominio Bacteria, siguiendo con el dominio Eukaryota y por último el dominio Archaea. Sabemos que la secuencia analizada corresponde a un organismo del dominio Archaea y el bajo número de hits puede corresponder no a diferencias secuenciales, sino que tal vez las bases de datos no contienen tanta información secuencial proveniente de arqueas. Por el contrario, sí se tiene información del dominio Bacteria y Eukarota y por eso el número de hits podría ser mayor. A su vez, teniendo en cuenta que la timidilato sintasa cumple una función que podría considerarse universal (participa en la biosíntesis de nucleótidos), tiene sentido que sea una enzima que se encuentre en organismos tan diversos que abarque desde bacterias hasta eucariontes.



**Figura 1.** Árbol filogenético que se obtiene tras realizar la búsqueda secuencial utilizando HMMER.

A partir de los resultados obtenidos y teniendo en cuenta las distintas metodologías utilizadas podemos observar cómo se complementan las tres herramientas. Por un lado el BLAST nos provee de información correspondiente a los organismos más cercanos evolutivamente mientras que las herramientas como PsiBlast o HMMER permiten obtener homólogos más distantes con porcentajes de identidad menores pero con buenos E-values.

## **Predicción de estructura secundaria, segmentos transmembrana, regiones desordenadas y dominios**

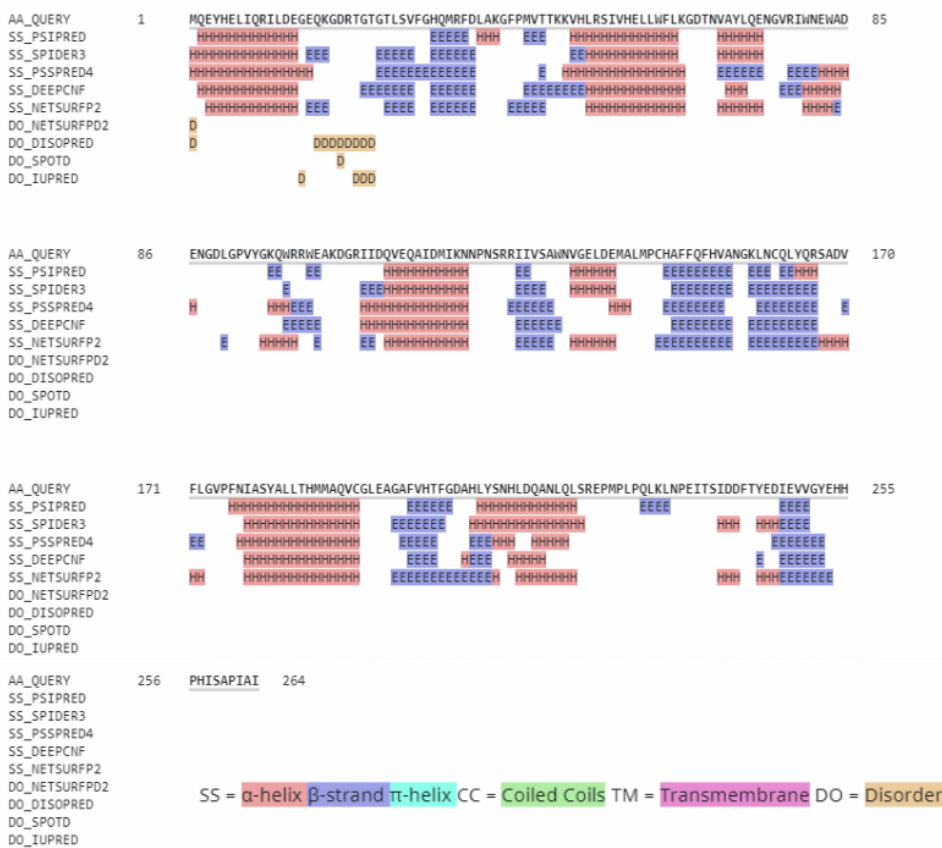
### **Dominios**

A partir de la base de datos de Pfam se encontró que toda la secuencia de la proteína corresponde a un único dominio correspondiente al dominio de la enzima timidilato sintasa y usando CDART encontramos que ese dominio sólo se encuentra en un tipo de arquitectura (compuesta sólo por el dominio Timidilato Sintasa) y en una única especie (*Candidatus Poseidoniales archaeon*). Predicción de estructura secundaria utilizando la herramienta QUick2D (<https://toolkit.tuebingen.mpg.de/tools/quick2d>).

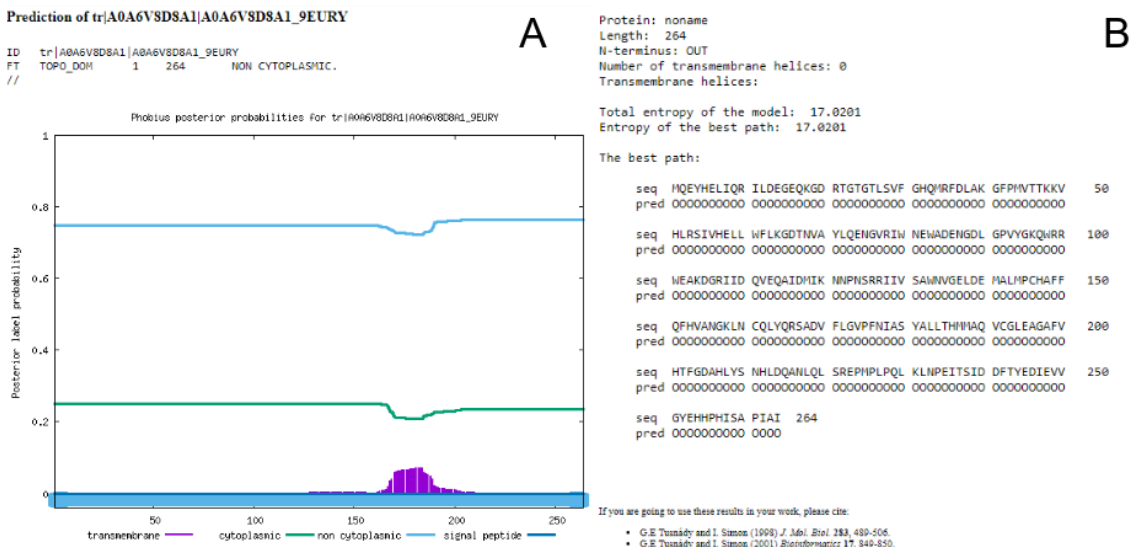
Como se muestra en la *Figura 2* podemos observar que la estructura del dominio se basa en una combinación de alfa hélices y hojas beta y no contiene regiones coiled-coils ni transmembrana. A su vez, si bien esta herramienta muestra posibles coincidencias de regiones desordenadas cercanas al N-terminal, más adelante en el informe se analizará con otros programas las regiones desordenadas al igual que los segmentos transmembrana.

### ***Segmentos transmembrana***

Para el análisis de los segmentos transmembrana se utilizaron dos herramientas distintas: Phobius y HMMTOP2. Como se observa en la *Figura 3A*, no hay predicción de segmentos transmembrana y gráficamente se puede observar que la probabilidad de que el segmento comprendido entre los aminoácidos 150-200 no supera el umbral correspondiente al citoplasmático ni al no-citoplasmático (extracelular). Los resultados de la *Figura 3B* son resultados complementarios para corroborar que la Timidilato Sintasa de *Candidatus Poseidoniales* no posee regiones transmembrana. Además, los resultados obtenidos se condicen con lo encontrado en el servidor Toolkit para predicción de estructura secundaria. Con estos resultados podemos decir que la Timidilato Sintasa es una proteína citoplasmática.



**Figura 2. Resultados que se obtienen luego de correr el Quick2D.** Las porciones rosas corresponden a las regiones alfa hélice, las azules a Hojas beta y las naranjas a las regiones desordenadas.

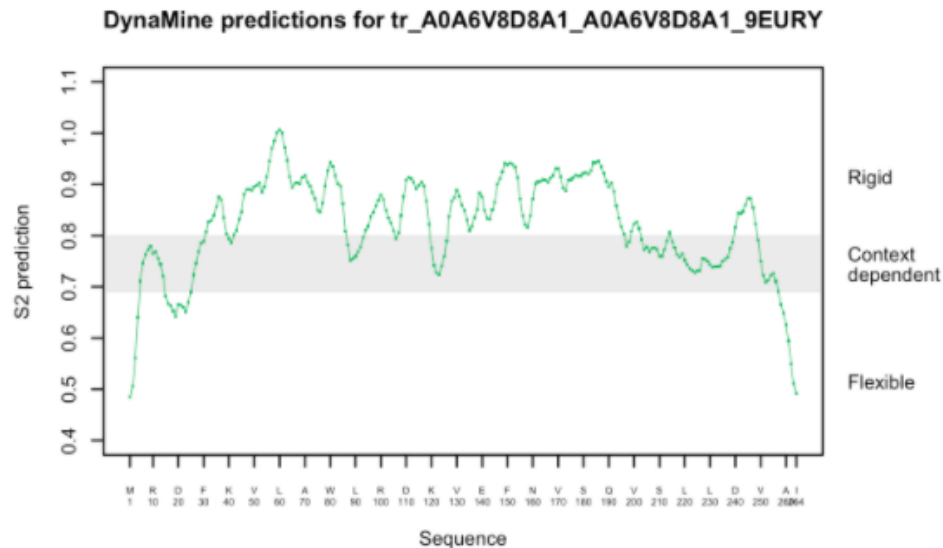


**Figura 3. Resultados de la evaluación de segmentos transmembrana.** A: resultado obtenido utilizando la herramienta Phobius. B: resultados obtenidos a través de HMMTOP2

**Regiones desordenadas**

A partir de la herramienta Dynamine se analizó la secuencia en búsqueda de regiones desordenadas y los resultados se muestran en la *Figura 4*. Se puede observar que la mayoría de las predicciones para cada aminoácido se encuentran por encima del umbral de flexibilidad, en otras palabras, gran parte de la

secuencia se considera rígida. Teniendo en cuenta que Dynamine asocia la flexibilidad con la presencia de regiones desordenadas, la presencia de rigidez entonces se corresponde con estructuras ordenadas. En conclusión, el dominio Timidilato sintasa es un dominio mayormente ordenado, con presencia de “desorden” en sus extremos N-terminal y C-terminal. Esto último tiene sentido si tenemos en cuenta que la mayoría de las proteínas tienen regiones más flexibles en sus extremos.



**Figura 4.** Resultados obtenidos utilizando la herramienta DynaMine  
<http://dynamine.ibsquare.be/tag/disorder/>

## Clasificación estructural de la proteína

Con el objetivo de seguir con la caracterización de la enzima Timidilato Sintasa de *Candidatus Poseidoniales archaeon* se evaluó la presencia de dominios estructurales pero a partir de herramientas que se basan en la aproximación de la estructura por homología de secuencia. Para ello se utilizó la base de datos de dominios estructurales CATH de la cual se obtuvieron los siguientes resultados: 231 dominios alineados y 32 matching Fun Fams.

Dentro de los 231 dominios, la clasificación CATH encontrada es la siguiente:

- C (clase): 3. Alpha- Beta
- A (arquitectura): 3.30. 2 Layer-Sandwich
- T (topología): 3.30.572. Thymidylate Synthase; Chain A
- H (Superfamilia de homólogos): 3.30.572.10. Thymidylate synthase/dCMP hydroxymethylase domain

Estos resultados concuerdan con los encontrados utilizando herramientas de predicción de estructura secundaria, pero a su vez da información de cómo se relacionan esas estructuras secundarias entre ellas y, de tenerlos, con otros dominios. En este caso, la información de proteínas con estructura conocida refuerza la idea de que la Timidilato Sintasa de *Candidatus Poseidoniales archaeon*, al igual que muchas de las Timidilato Sintasas, están compuestas por un único dominio.

## Obtención de un modelo molecular para la proteína

### Elección del Template

Se utilizan varias herramientas para seleccionar el mejor template posible, con un alto %Identidad, bajo E-value, buena cobertura, con estructura conocida y en lo posible una resolución menor a 2 amgstroms. El resultado que cumple con los requisitos mencionados se encontró utilizando HHPred y posee las siguientes características:

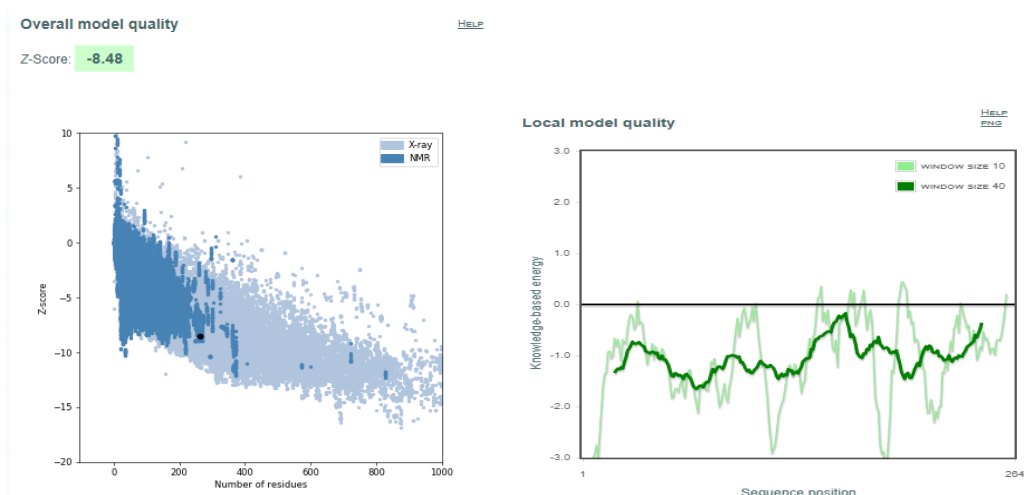
- Nombre: Thymidylate synthase (Escherichia coli (strain K12))
- Código PDB: 6NNR
- E-value: 3.4e-57
- %Identidad: 71%
- Similitud: 1.241
- Score: 393.2
- Posiciones alineadas: 264
- Resolución: 1.05 Å

**Alineamiento:** se lleva a cabo utilizando Clustal y se convirtió de manera manual al formato PIR.

### **Determinación de regiones conservadas y variables**

Para generar el modelo tridimensional de la proteína se utilizó el programa Modeller que construye el backbone de carbonos alfa, los loops y cadenas laterales que luego podrán ser optimizadas. Para poder obtener un modelo tridimensional de la proteína de interés, al programa se le debe ingresar un archivo de comandos que indique el alineamiento y el template que debe utilizar para el modelado. El resultado son una serie de modelos con distintos valores de función objetivo. En este trabajo se eligió el modelo trA0A6V8D8A1.B99990003.pdb, que pondera entre el valor más bajo de la función objetivo y el DOPEscore. Cabe aclarar que el Modeller, a diferencia de otros programas, no utiliza una función objetivo con base energética, sino que a partir de la información del template genera una matriz con información estructural y a partir de la misma obtiene los “restrains”. Los “restrains” son parámetros estructurales que poseen un rango en el que pueden moverse. El mejor modelo es el que infringe la menor cantidad de restrains y por eso se trata de elegir el menor valor de función objetivo.

### **Evaluación del modelo en función del criterio energético utilizando ProsaII**



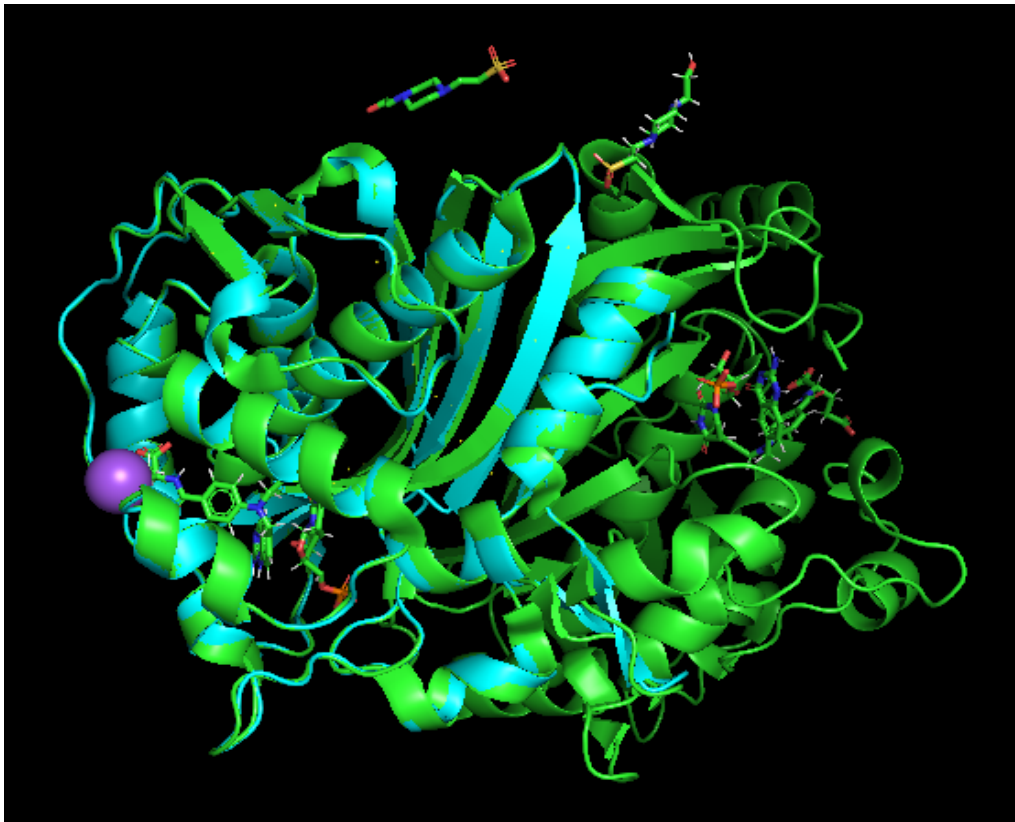
**Figura 5. Izquierda:** Resultados de Z-score en función del número de residuos de las proteínas con estructura conocida que se obtiene en ProsaII. El punto negro indica la ubicación del modelo. **Derecha:** Energía de cada posición dentro de la secuencia de interés.

La evaluación del modelo se puede llevar a cabo siguiendo criterios estructurales y/o energéticos. En este caso sólo se evaluó a partir de criterios energéticos y para ello se utilizó la herramienta ProsaII. En la *Figura 5* se observa la evaluación del modelo a nivel global. El gráfico refleja la relación entre los valores de

Z-score, que se obtienen para proteínas con estructura conocida, en función de la longitud de cada proteína. Como el modelo se encuentra dentro de la distribución de posibles estructuras, podemos decir que es un buen modelo. Por otro lado, se muestra la distribución energética de cada posición de la secuencia del modelo (*Figura 5, derecha*), dando información de las regiones que pueden ser optimizadas, teniendo en cuenta que las posiciones de mayor energía corresponden generalmente a los loops. Como en este caso, las posiciones de mayor energía ni siquiera llegan a valores positivos y cómo esas regiones no se corresponden en loops, no se realiza una optimización de estos.

### **Visualización del modelo y comparación con el template en Pymol**

En la *Figura 6* se observa el modelo obtenido alineado con el template 6NNR (Timidilato Sintasa de *E.coli*), que a simple vista, no demuestra tener grandes diferencias uno de otro y en la consola de Pymol se puede leer que el RMSD del alineamiento es de 0.18. Teniendo en cuenta que el error experimental asociado a la difracción de rayos X y al alineamiento de una misma proteína es de 0.5, podemos decir que el modelo obtenido a través de Modeller es muy similar al template si sólo tenemos en cuenta la información estructural.



**Figura 6.** Superposición 3D del modelo (cian) y el template 6NNR (verde) que se obtiene a partir de Pymol.

### **Selección y alineamiento múltiple de secuencias**

Con el objetivo de analizar las regiones secuenciales conservadas de la enzima Timidilato Sintasa y posterior análisis filogenético, se seleccionaron al azar 15 secuencias del BLAST con diferentes porcentajes de identidad que sean lo más heterogéneos posibles (Ver *Tabla 1* en el anexo). Se buscaron las secuencias de cada organismo en la base de datos de NCBI y a partir de las mismas en formato FASTA se procesaron utilizando la herramienta CLUSTAL <https://www.ebi.ac.uk/Tools/msa/clustalo/>.

A partir de las secuencias se observó que la longitud difiere entre los diferentes organismos. Esto puede deberse a que como la Timidilato Sintasa es un dominio completo, una secuencia más larga puede corresponder a una proteína con arquitectura compuesta por varios dominios donde uno de ellos es el



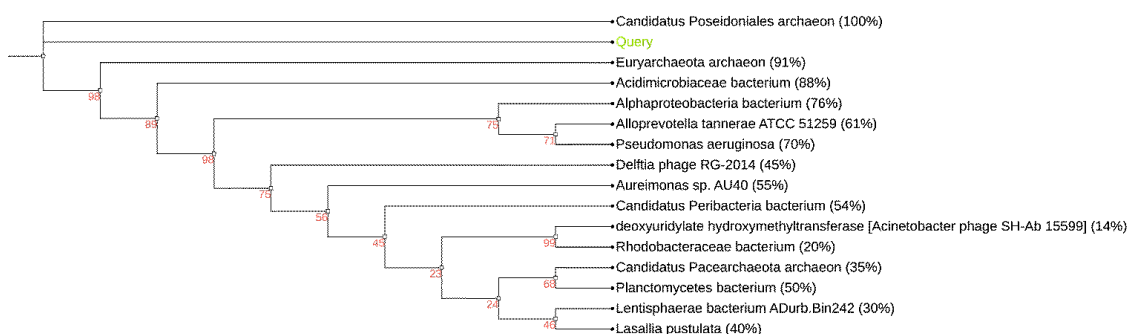
correspondiente a la Timidilato Sintasa. Dicho esto, y si bien el alineamiento no es de los mejores ya que presenta muchos gaps, donde comienza el alineamiento de la secuencia de interés se observan regiones con un grado de conservación considerable, teniendo en cuenta que se tomaron secuencias desde un 95% de identidad hasta un 14% de identidad. (Ver Figura 1 del anexo)

Si consideramos que la conservación secuencial conlleva a la conservación estructural y funcional, tiene sentido que la mayor parte del dominio se encuentre conservado, ya que de esa forma se asegura que la función se mantiene a lo largo del tiempo. Además, se observan algunas regiones que se encuentran más conservadas que otras (entre los aminoácidos 174-255) que con un futuro análisis estructural y funcional se podría llegar a dilucidar el rol que cumple esa región para que la secuencia se conserve más que en otras regiones.

### Obtención de un árbol filogenético

Con el objetivo de obtener un árbol filogenético se utilizó la herramienta IQtree, que en primera instancia realiza una evaluación de modelos para encontrar el modelo de evolución que mejor ajuste al alineamiento múltiple presentado anteriormente. El modelo de evolución encontrado es LG y es el que se utiliza para estimar la topología del árbol filogenético. Además, se realizó un bootstrapping ultrarápido con 1000 iteraciones (alineamientos realizados al azar). Se obtuvo un árbol filogenético en formato newick que se usó como input en el programa iTOL que nos permite una mejor visualización.

Como se puede observar en la *Figura 7*, a medida que disminuye la similitud secuencial, es decir el porcentaje de identidad baja, las diferencias evolutivas aumentan. Por otro lado, se observa que a medida que disminuye la similitud secuencial, los valores de bootstrapping disminuyen, demostrando que la información que posee el alineamiento no es suficiente para soportar las inferencias filogenéticas en torno a la enzima Timidilato Sintasa cuando la similitud secuencial disminuye. Como era de esperarse, la secuencia Query corresponde a la misma Timidilato Sintasa de *Candidatus Poseidoniales archaeon*, y además la mayor cercanía evolutiva se da con proteínas con alto porcentaje de identidad. Salvo *Lasallia pustulata* que es un hongo y *Delftia phage RG-2014* y *deoxyuridylate hydroxymethyltransferase [Acinetobacter phage SH-Ab 15599]* que son fagos, el resto de las secuencias analizadas corresponden a proteínas de bacterias. Esta información se ve reflejada tanto en la similitud secuencial como en la divergencia evolutiva. La Timidilato Sintasa de *Lasallia pustulata* se encuentra más lejana evolutivamente que el resto de las enzimas de sus respectivos organismos.



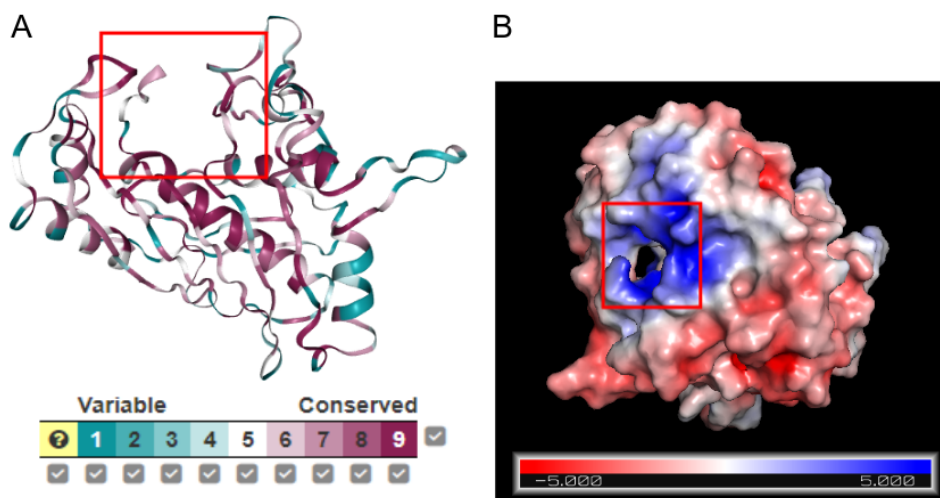
**Figura 7.** Árbol filogenético obtenido a partir de iTOL. Los números rosas corresponden a los valores de bootstrapping y los números entre paréntesis hacen referencia al porcentaje de identidad que comparten la secuencia de Timidilato Sintasa de ese organismo alineada con la secuencia Query.

Si bien no se pueden realizar inferencias filogenéticas a nivel de organismo ya que sólo se está analizando una única secuencia correspondiente a una única proteína, encontramos una mayor distancia evolutiva entre Timidilato Sintasas de organismos que presentan mayores diferencias (hongos-arqueas).



## Análisis funcional de la Timidilato Sintasa

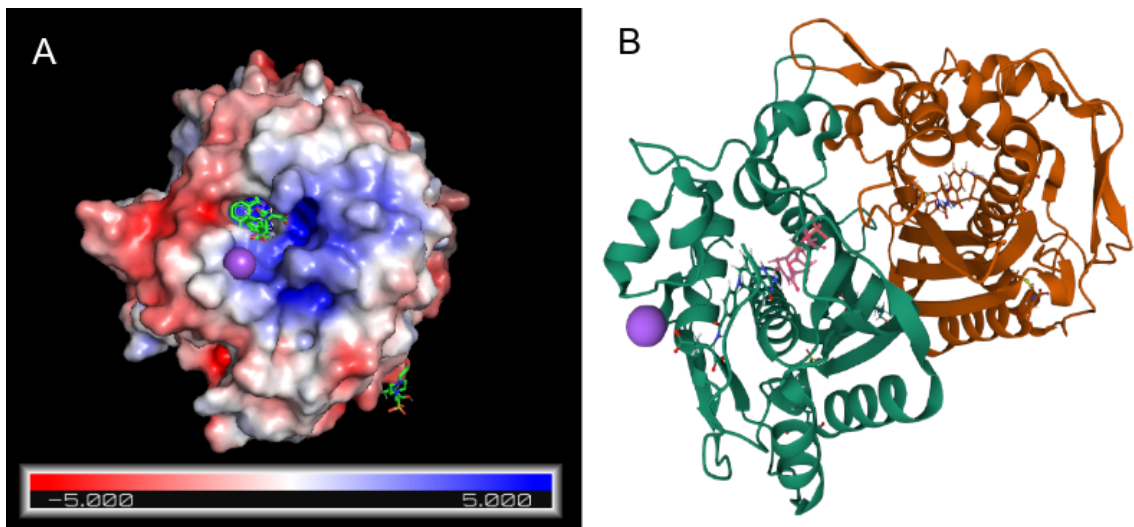
La estimación de la función de una proteína no es tarea sencilla y deben utilizarse la mayor cantidad de criterios para llegar al mejor resultado posible, que luego deberá ser confirmado de manera experimental. Desde el punto de vista evolutivo, resulta interesante evaluar las posiciones conservadas ya que estas posiblemente estén relacionadas con la función biológica de la enzima. Para ello se utilizó la herramienta ConSurf que, mediante Evolutionary Trace, da información de las posiciones más conservadas y menos conservadas.



**Figura 8.** Representaciones espaciales del modelo de Timidilato Sintasa. **A.** Posiciones conservadas dentro del modelo. **B.** Carga superficial del modelo. Los recuadros rojos hacen referencia a la misma región y muestra un posible sitio de unión o entrada de sustratos.

Como se observa en la *Figura 8A*, las posiciones más conservadas se observan en la alfa hélice del centro de la cadena y en algunos loops que se encuentran encuadrados en rojo. Esa misma región en la *Figura 8B* muestra la presencia de un canal que puede llegar a ser importante en la funcionalidad de la proteína.

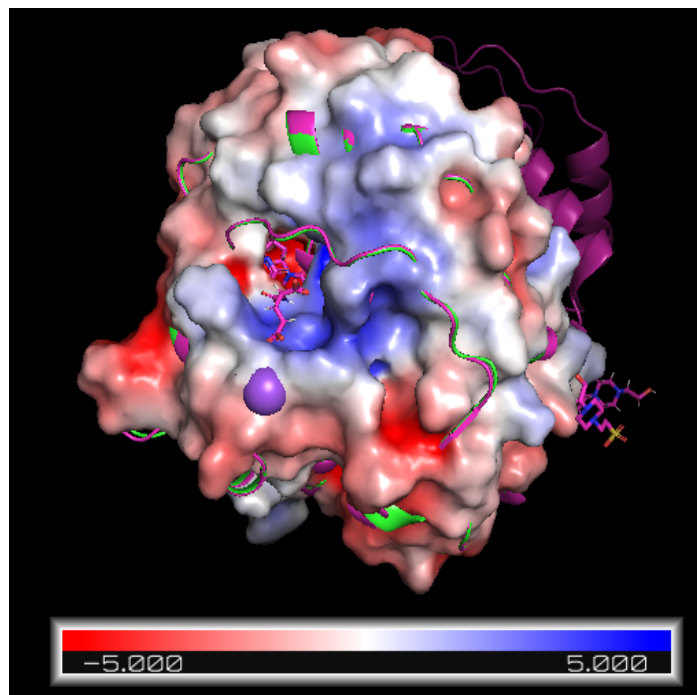
Al comparar la información del template utilizado en el modelado con el modelo obtenido a través de Modeller se puede llegar a una aproximación de cuál es el sitio activo de la proteína. Por un lado, se analiza la carga superficial del template utilizando Pymol (*Figura 9A*), donde se puede ver que los ligandos se encuentran inmersos en un canal conformado por aminoácidos de carga positiva. A su vez, la *Figura 9B* muestra la disposición estructural de la Timidilato Sintasa de *E.coli* (6NNR). Como puede observarse, la cavidad formada por aminoácidos de carga positiva (*9A*) direcciona hacia el sitio activo de la proteína, donde va a ingresar el (6R)-5,10-methylene-5,6,7,8-tetrahydrofolate y el dUMP (ambos con carga negativa), sustratos necesarios para que ocurra la reacción.



**Figura 9.** Estructura tridimensional de la Timidilato Sintasa de *E.coli*. **A:** carga superficial obtenida a partir de Pymol. **B:** Estructura 3D que se puede encontrar en la base de datos PDB. En rosa el dUMP y en azul (6R)-5,10-methylene-5,6,7,8-tetrahydrofolate, ambos sustratos de la reacción catalizada por la enzima.

El mismo procedimiento se realizó con el modelo obtenido en Modeller. Como se puede observar en la *Figura 10*, la carga superficial del canal varía levemente en el interior. En lugar de estar conformado en su totalidad por aminoácidos con carga positiva, se ve que hacia el interior aumenta la proporción de aminoácidos con carga negativa. Esto podría afectar a la función de la Timidilato Sintasa o simplemente puede estar reflejando que al ser de organismos distintos (el template proviene de *E.coli* mientras que nuestra proteína es de un organismo del reino Archaea), su composición aminoacídica refleja una divergencia evolutiva correspondiente al hábitat diferencial en que normalmente se encuentran esos organismos.

Por otro lado, retomando la *Figura 8*, podemos confirmar que las posiciones más conservadas tienen una funcionalidad enzimática, ya sea que da lugar al canal de acceso al sitio activo o el sitio activo en sí.



**Figura 10.** Carga superficial del modelo. El modelo(verde) está alineado con el template 6NNR (magenta)

Por último, en la base de datos de UniProt vemos que la proteína de interés tiene las siguientes anotaciones GO:

- Función Molecular: actividad metiltransferasa, actividad transferasa, actividad timidilato sintasa, actividad de transferencia de un grupo carbono
- Proceso Biológico: metilación, proceso de biosíntesis de nucleótidos, biosíntesis de dTMP.

## CONCLUSIONES Y DISCUSIÓN

El uso de diversas herramientas bioinformáticas nos permitieron identificar y confirmar que la secuencia de interés corresponde a la enzima Timidilato Sintasa presente en una especie de archeas conocida como *Candidatus Poseidoniales archaeon*. A partir de BLAST encontramos secuencias homólogas evolutivamente cercanas mientras que PsiBlast y HMMER que utilizan otros algoritmos basados en la generación de perfiles y HMM, nos permitieron obtener secuencias homólogas más lejanas. A partir de estos resultados se puede buscar un gran número de propiedades y características estructurales que intervienen en la funcionalidad y es así como encontramos que la proteína de interés se encuentra en el citoplasma y está conformada por un único dominio. A su vez ese dominio posee una estructura que se basa en la combinación de alfa hélices y hojas beta donde las regiones desordenadas se encuentran en las regiones terminales de la proteína.

La información secuencial nos permitió obtener la relación evolutiva entre secuencias proteicas provenientes de distintos organismos y de esa manera pudimos realizar inferencias funcionales de los canales por los que ingresan los metabolitos necesarios para que ocurra la reacción. También encontramos que si bien las regiones funcionales de la Timidilato Sintasa se conservan a lo largo de varias especies, el cambio de aminoácidos de carga positiva por aminoácidos neutros o de carga negativa pueden correlacionarse con el ambiente en el que se encuentran y desarrollan los diferentes organismos.

Resulta interesante evaluar la diversidad y la flexibilidad que presentan las herramientas utilizadas a lo largo del trabajo. Tanto la hipótesis como la pregunta biológica que se desea responder determinan los programas o herramientas a utilizar y qué información se utilizará como input. En este trabajo nos planteamos como objetivos corroborar la información brindada por las diferentes bases de datos sobre la Timidilato Sintasa de *Candidatus Poseidoniales archaeon* y a su vez explorar el modelado tridimensional de dicha enzima y la importancia de las secuencias conservadas sobre la funcionalidad de la misma y para ello se utilizaron herramientas que tienen un enfoque evolutivo que nos permitieron inferir o hipotetizar sobre los objetivos planteados. Es importante tener en cuenta que la inferencia bioinformática no reemplaza la necesidad de corroborar las hipótesis con medidas experimentales.

Por último, en el caso particular de la Timidilato Sintasa de *Candidatus Poseidoniales* donde existe bastante información al respecto de su función y disposición estructural, con el uso de las herramientas presentadas se pudo obtener un modelo estructural de la proteína que hasta el momento no se tenía. Además, a través de información evolutiva y estructural se pudo reconocer el sitio activo de la proteína.

## BIBLIOGRAFÍA

Rinke, C., Rubino, F., Messer, L.F. et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J* 13, 663–675 (2019). <https://doi.org/10.1038/s41396-018-0282-y>