



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## Maestría y Doctorado en Ciencias Bioquímicas

RECONSTRUCCIÓN DE SECUENCIAS ANCESTRALES DE PROTEÍNAS: GENERACIÓN DE SECUENCIAS Y CARACTERIZACIÓN DE ANCESTROS DE LA TRIOSAFOSOFATO ISOMERASA.

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias

PRESENTA:  
JORGE ALEJANDRO PÉREZ NIÑO

TUTOR PRINCIPAL  
DANIEL ALEJANDRO FERNÁNDEZ VELASCO  
[Facultad de Medicina, UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. Miguel Antonio Costas Basín  
[Facultad de Química, UNAM](#)

Dr. Lorenzo Patrick Segovia Forcella  
[Instituto de Biotecnología, UNAM](#)

Ciudad de México. Abril, 2019



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



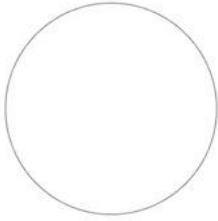
**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

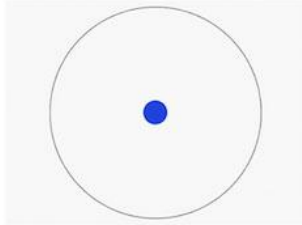
Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

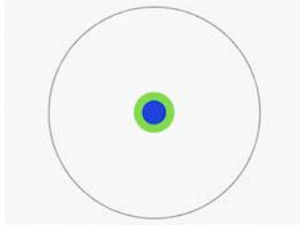
Imagine a circle that contains all of human knowledge:



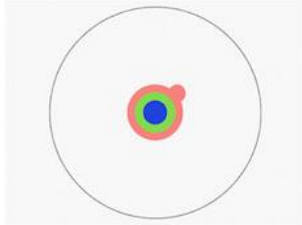
By the time you finish elementary school, you know a little:



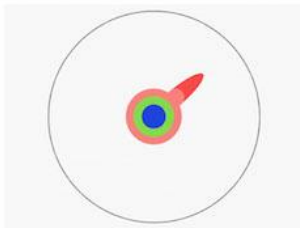
By the time you finish high school, you know a bit more:



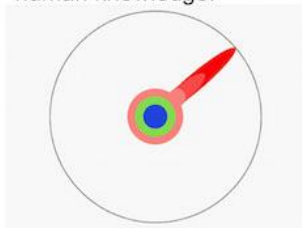
With a bachelor's degree, you gain a specialty:



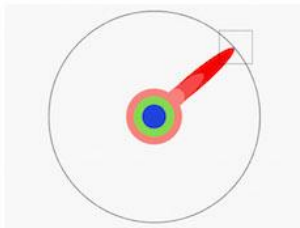
A master's degree deepens that specialty:



Reading research papers takes you to the edge of human knowledge:



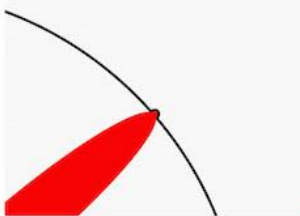
Once you're at the boundary, you focus:



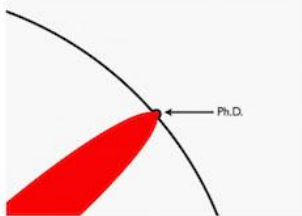
You push at the boundary for a few years:



Until one day, the boundary gives way:



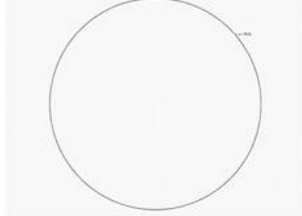
And, that dent you've made is called a Ph.D.:



Of course, the world looks different to you now:



So, don't forget the bigger picture:



Keep pushing.

"Yo solo sé que no sé nada"  
**Sócrates**

"Si oyes una voz dentro de ti decir que no puedes pintar, entonces por todos los medios pintarás y esa voz será silenciada"  
**Vincent van Gogh**

## **Agradecimientos académicos**

Este trabajo se realizó bajo la tutoría del Doctor Daniel Alejandro Fernández Velasco en el Laboratorio de Físicoquímica e Ingeniería de Proteínas (LFQIP), de la Facultad de Medicina.

Durante el desarrollo de este trabajo se contó con apoyo económico de beca de maestría por parte del Consejo Nacional de Ciencia y Tecnología (CONACyT), número de CVU 817384.

También se recibió apoyo por parte de CONACyT, número de proyecto 254514 y del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), número de proyecto IN220516.

Un agradecimiento a la Red, Estructura, Función y Evolución de Proteínas (REFEP) por proporcionar apoyo económico para el 6° Congreso de la REFEP en Durango, Durango.

Un agradecimiento a la REFEP, por proporcionar apoyo económico para el Taller de Cristalografía de Proteínas en el Cinvestav, Langedio. También quiero agradecer a la REFEP por crear cursos como éste, que ayudan a la comunidad científica a la formación de investigadores y a crear vínculos entre nosotros, una ventaja que a menudo es subestimada, pienso yo.

Al Doctor Miguel Antonio Costas Basín y la M. en C. Jessica Díaz por su incondicional apoyo en los experimentos de Calorimetría Diferencial de Barrido en el Laboratorio de Biofísicoquímica de la Facultad de Química.

Aunque no se muestren los resultados de cristalografía y difracción quiero agradecer al Doctor Yasel Guerra Borrego por su apoyo en las pruebas de cristalización y a la Doctora Adela Rodríguez Romero, la M. en C. Georgina Espinosa Pérez y el Laboratorio Nacional de Estructuras de Macromoléculas (LANEM) por su colaboración con la difracción de los cristales de proteína.

A mi Comité Tutorial conformado por el Dr. Alejandro Fernández Velasco, el Dr. Miguel Antonio Costas Basín y el Dr. Lorenzo Patrick Segovia Forcella, por sus acertados comentarios a lo largo del trabajo y que ayudaron a guiarlo en la dirección que se encuentra ahora.

A mi jurado de examen conformado por el Dr. Diego González Halphen, el Dr. Horacio Reyes Vivas, el Dr. Enrique García Hernández, el Dr. Arturo Carlos II Becerra Bracho y el Dr. Rogelio Rodríguez Sotres, por sus correcciones al trabajo aquí presentado que permitieron moldearlo a un trabajo de mayor calidad y me demostraron que presentación un proyecto es tan importante como el mismo.

A los técnicos del LFQIP la M. en C. Laura Illeana Álvarez Añorve y la Dra. María Isabel Velázquez López.

A los Doctores del LFQIP: la Doctora Georgina Garza Ramos, el Doctor Ismael Bustos Jaimes y el Doctor Alejandro Sosa Peinado por sus comentarios durante los seminarios interinos.

## **Agradecimientos personales**

A mis padres, primero decirles a los dos, sé que no es fácil ser padre, me sé poner en los zapatos de los demás y he visto lo que hacen y dejan de hacer por mí y mis hermanos. Espero este trabajo sea una muestra de lo que sus esfuerzos han logrado. A mi papá, hasta después de crecer me he dado cuenta de el maravilloso trabajo que has hecho como padre, me siento muy feliz de ser tu hijo. A mi mamá, eres el cariño personificado, la persona que siempre me prestará un oído amigo y siempre estará ahí conmigo. Los amo incondicionalmente.

A mi abuela, por ser lo que cualquier persona debería aspirar a ser, un ser de amor incondicional, de perdón, de recuerdos felices, un ser sin arrepentimientos.

A mi jefe, el Doctor Alejandro Fernández, por... todo. Por darme esa entrevista por Skype. Por incorporarme a las clases de Físicoquímica. Por permitirme ingresar a su equipo de trabajo, aunque estaba lleno. Por ver y compartir la pasión inicial que nos llevó a retomar este proyecto. Por contemplarme para cursos, libros, artículos y más, que sabe me harán crecer como persona e investigador. Por no dudar en atender mis dudas por más banales que sean. Por tener siempre su puerta abierta a quien lo necesite. Por ser un ejemplo a seguir como investigador y como ser humano. Por ser.

A Carlos Everardo, todo lo que soy como investigador te lo debo a ti, muchas gracias por darme la oportunidad de enseñarme todo lo que sabías, aunque en un principio no te animo la idea yo doy gracias que no te echaras para atrás.

A Renan, Sergio Romero, Banda, Saira y Eloy. Porque sin yo tener conocimiento absoluto en bioinformática me introdujeron, me enseñaron, me guiaron y me apoyaron en este tema tan escabroso.

A Karla, por compartirme tu pasión por la ciencia, por la biología, por la música, por la vida.

A Octavio, mi amiguito del alma, que caminamos juntos por el valle de la muerte, pero triunfamos al final.

A Dana, mi amiguita, ojalá te encontraras acá para compartir esto contigo.

A mis Galletines, SAMAGA, Victor, Lily, Luis y Jesús, me enseñaron tanto, no solo profesionalmente, y me dieron el regalo de la amistad incondicional.

A mis compañeros del LFQIP por soportar mis seminarios de dos horas y media y aun así estar atentos y aportarme puntos de vista que nunca contemple y sugerencias que nunca desechara.

A mis amigos que hice durante la licenciatura y aún están conmigo, que sepan que han marcado mi vida, me alegra que se encuentran en ella, que me consideren su amigo y ojalá estén en mi vida por mucho tiempo.

A Sandybell, mi amiga del alma, me demuestras que la sinceridad es el componente de la verdadera amistad y con ella me haces ser un mejor hombre. Tu eres mi amiga fiel.

A Pamela, tú que fuiste mi compañera por tanto tiempo te debo tanto que no creo podértelo pagar, espero la vida te regrese tanto que has dado.

A Lucy, que tras un fortuito encuentro (o tal vez un destinado encuentro) te convertiste en mi amiguita indispensable, mi confidente y quiero que estés en mi vida por años y años para ver como te llega la felicidad que mereces.

A mi hermanota Aleta, simplemente quiero ser como tú, eres mi ejemplo a seguir como persona y como investigadora. Me alegra mucho haberte conocido.

A Rogelio, Gaby y a los pequeños, Jiaru y Blanquita, no sólo por proporcionarme una casa sino un hogar, durante la licenciatura y parte de la maestría.

A mi tío Audon, a Erick y a toda su familia, por acogerme en tiempos de incertidumbre y demostrarme la generosidad a niveles inimaginables. Ojalá pueda llegar a dar como ustedes.

A Deheny, que sepas que cambiaste mi vida, cada que nos encontrábamos era una aventura y ojalá nos volvamos a encontrar.

De nuevo quiero agradecer a mi jurado de examen, para que sepan que sus comentarios no cayeron en saco roto. Que analice cada uno, por más insignificante que pareciera, pero me ayudaron a mejorar mi habilidad para la comunicación científica. Espero tener algún día la destreza y la sensibilidad para ver detalles que conviertan un trabajo de normal a trascendental.

También agradecer de nuevo a mi Comité Tutorial, sus comentarios me encantaron, no hubiera podido escoger un mejor Comité Tutorial. Agradecimiento especial al Dr. Lorenzo, tengo muy presente que hacer un viaje para un Tutorial desde Cuernavaca no es algo trivial. Lo aprecio mucho.

Y me di cuenta de que si me tomo el tiempo en agradecer a todas las personas que han estado en mi vida esta tesis sería de puros agradecimientos. “Que sepan que me han tocado mi corazón, y les aseguro que también los toque”.

## INDICE

|  |           |
|--|-----------|
| <b>Abreviaturas</b>  | <b>VI</b> |
| <b>Resumen</b>   | <b>1</b>  |
| <b>Introducción</b>  | <b>2</b>  |
| Reconstrucción de secuencias ancestrales                       | 2         |
| Triosafosfato isomerasa  | 6         |
| <b>Antecedentes</b>  | <b>9</b>  |
| <b>Justificación</b>   | <b>11</b> |
| <b>Hipótesis</b>   | <b>11</b> |
| <b>Objetivos</b>   | <b>11</b> |
| <b>Metodología</b>   | <b>11</b> |
| <b>Metodología de reconstrucción de secuencias ancestrales</b> | <b>11</b> |
| Selección de secuencias  | 12        |
| Alineamiento múltiple de secuencias                            | 13        |
| Selección del modelo de sustitución                            | 13        |
| Árbol filogenético   | 13        |
| Reconstrucción de secuencias ancestrales                       | 15        |
| <b>Procedimiento experimental</b>                              | <b>17</b> |
| Sobreexpresión de las proteínas ancestrales                    | 18        |
| Purificación por columna de afinidad                           | 18        |
| Dicroísmo circular   | 18        |
| Fluorescencia  | 18        |
| Calorimetría diferencia de barrido                             | 18        |
| Actividad enzimática   | 19        |
| Estado oligomérico y estabilidad a la dilución                 | 19        |
| <b>Resultados</b>  | <b>20</b> |
| <b>Metodología de reconstrucción de secuencias ancestrales</b> | <b>20</b> |
| Selección de secuencias  | 20        |
| Alineamiento múltiple de secuencias                            | 20        |
| Selección del modelo de sustitución                            | 22        |
| Árbol filogenético   | 22        |
| Reconstrucción de secuencias ancestrales                       | 25        |
| <b>Resultados experimentales</b>                               | <b>30</b> |
| Sobreexpresión de las proteínas ancestrales                    | 30        |
| Purificación por columna de afinidad                           | 30        |
| Dicroísmo circular   | 32        |
| Fluorescencia  | 33        |
| Calorimetría diferencia de barrido                             | 34        |
| Actividad enzimática   | 35        |
| Estado oligomérico y estabilidad a la dilución                 | 36        |
| <b>Discusión</b>   | <b>40</b> |
| <b>Conclusiones</b>  | <b>56</b> |
| <b>Perspectivas</b>  | <b>57</b> |
| <b>Referencias</b>   | <b>58</b> |
| <b>Anexos</b>  | <b>65</b> |

## Abreviaturas

|                                   |  |
|-----------------------------------|--|
| °C                                | Grados Celsius   |
| µg                                | Microgramo   |
| µL                                | Microlitro   |
| µm                                | Micrómetro   |
| µM                                | Micromolar   |
| Å                                 | Angstroms  |
| AB                                | TIM del nodo perteneciente al último ancestro de los opistocontes obtenido por inferencia bayesiana y el método de peor caso plausible |
| ADN                               | Ácido desoxirribonucleico  |
| AIC                               | Criterio de Información de Akaike  |
| AM                                | TIM del nodo perteneciente al último ancestro de los opistocontes obtenido por máxima verosimilitud y el método de peor caso plausible |
| ASR                               | Reconstrucción de secuencias ancestrales   |
| BAY                               | TIM del nodo perteneciente al último ancestro de los opistocontes obtenido por Inferencia Bayesiana                                    |
| CDNN                              | <i>Circular Dichroism analysis using Neural Network</i>  |
| C <sub>p</sub>                    | Capacidad calorífica   |
| DC                                | Dicroísmo circular   |
| DHAP                              | Dihidroxiacetona fosfato   |
| DSC                               | Calorimetría diferencial de barrido  |
| DTT                               | Ditiotreitol   |
| <i>E. coli</i>                    | <i>Escherichia coli</i>  |
| EDTA                              | Ácido etilendiaminotetraacético  |
| FPLC                              | Cromatografía rápida líquida de proteínas  |
| g                                 | gramo  |
| G3P                               | Gliceraldehido-3-fosfato   |
| In-del                            | Inserción-delección  |
| IPTG                              | Isopropil-β-D-1-tiogalactopiranosido   |
| k <sub>cat</sub> / K <sub>m</sub> | Eficiencia catalítica  |
| k <sub>cat</sub>                  | Constante catalítica   |
| kDa                               | Kilodalton   |
| K <sub>0.5</sub>                  | Concentración de enzima donde se pierde la mitad de la actividad   |
| K <sub>diss</sub>                 | Constante de disociación   |
| K <sub>m</sub>                    | Constante de afinidad  |
| LB                                | Medio Luria Bertani  |
| M                                 | Molar  |
| MAFFT                             | <i>Multiple Alignment using Fast Fourier Transform</i>   |
| MCMC                              | Método de cadena de Markov Monte Carlo   |
| mg                                | Miligramo  |
| min                               | Minuto   |
| ML                                | TIM del nodo perteneciente al último ancestro de los opistocontes obtenido por máxima verosimilitud                                    |
| mL                                | Mililitro  |
| mM                                | Milimolar  |



|                  |  |
|------------------|--|
| NAD <sup>+</sup> | Dinucleótido de nicotinamida y adenina oxidado           |
| NADH             | Dinucleótido de nicotinamida y adenina reducido          |
| NCBI             | <i>National Center for Biotechnology Information</i>     |
| ng               | Nanogramo  |
| nm               | Nanómetros   |
| PAGE             | Electroforesis en gel de poliacrilamida                  |
| PAML             | <i>Phylogenetic Analysis by Maximum Likelihood</i>       |
| Pc               | Concentración total de proteína                          |
| PDB              | Base de datos de proteínas                               |
| PGH              | Fosfoglicohidroxamato                                    |
| PM               | Peso molecular   |
| PMSF             | Fluoruro de fenilmetilsulfonilo                          |
| R                | Constante universal de los gases                         |
| R <sub>s</sub>   | Radio de Stokes  |
| SDS              | Dodecilsulfato sódico                                    |
| SMS              | <i>Smart model selection</i>                             |
| TIM              | Triosafosfato Isomerasa                                  |
| T <sub>m</sub>   | Temperatura media de desnaturalización                   |
| T <sub>s</sub>   | Temperatura de máxima estabilidad                        |
| UPGMA            | <i>Unweighted pair group method with arithmetic mean</i> |
| αGDH             | α-glicerol 3-fosfato deshidrogenasa                      |
| ΔC <sub>p</sub>  | Cambio de capacidad calorífica                           |
| ΔG               | Cambio en la energía libre de Gibbs                      |
| ΔH               | Cambio de entalpía                                       |
| λ <sub>max</sub> | Longitud de onda de máxima emisión                       |

## Resumen

La reconstrucción de secuencias ancestrales permite inferir la secuencia de proteínas que existieron hace mucho tiempo y caracterizarlas en el laboratorio. Gracias a esta técnica se han logrado realizar inferencias sobre las proteínas ancestrales, como su especificidad y actividad, y sobre las células ancestrales, como el ambiente en el que vivieron. Sin embargo, la identidad de la secuencia inferida no puede ser comparada con la verdadera secuencia ancestral, por lo que un punto importante de la técnica es asegurar que las características ancestrales inferidas se mantengan sin importar las variables del método de reconstrucción.

En uno de los trabajos anteriores en el laboratorio, se realizó una reconstrucción de secuencias ancestrales sobre la enzima triosafosfato isomerasa; una enzima con estructura de barril TIM, enzima partícipe en la glucólisis, oligómero obligado y con una eficiencia catalítica limitada por difusión. La reconstrucción se realizó con 44 secuencias de TIMs de eucariontes, un árbol de especies y el método de máxima verosimilitud para la inferencia de la secuencia. En este trabajo previo, se observaron características inusuales en la TIM del ancestro de los opistocontes. Entre estas características se observa una eficiencia catalítica reducida ( $8 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ), un alta constante de disociación ( $0.2 \mu\text{M}$ ) y una estabilidad mayor de los monómeros aislados. Ya que no se puede comparar la TIM ancestral predicha con la verdadera, es necesario aumentar la confianza en los resultados obtenidos mediante diferentes metodologías.

Se decidió realizar una segunda reconstrucción con un mejor manejo de la redundancia de secuencias y utilizando dos metodologías de reconstrucción; la máxima verosimilitud y la inferencia bayesiana. Además, para aumentar la robustez de las características frente a los sitios ambiguamente reconstruidos, se recurrió a la técnica de peor caso plausible, donde se cambian todos los sitios con ambigüedad por el segundo mejor caso predicho en la reconstrucción. Esto, con el fin de incorporar el mayor número de errores posibles y determinar que las características que se desea medir son invariantes frente a la ambigüedad de los sitios.

Se trabajó con cuatro proteínas ancestrales. Se determinó la TIM del último ancestro de los opistocontes utilizando dos métodos de reconstrucción y la técnica de peor caso plausible para cada metodología, posteriormente las proteínas correspondientes se sobreexpresaron y estudiaron. La caracterización se realizó a nivel funcional y estructural con el fin de determinar qué características se conservan a pesar de variar la metodología o modificando los sitios ambiguos, y cuáles son robustos sin importar la incertidumbre de la propia metodología de la reconstrucción ancestral.

Las propiedades más generales como topología, estado oligomérico y catálisis se conservan invariables sin importar el método o los aminoácidos ambiguos. La temperatura media de desnaturalización y el  $\Delta H$  de desplegamiento variaron dependiendo del método de reconstrucción. La estabilidad a la dilución es muy variable entre las cuatro reconstrucciones y es dependiente del método que se utilice y los aminoácidos ambiguos que se incorporen.

## Introducción

### Reconstrucción de secuencias ancestrales

La evolución de los genes y las proteínas es central en el campo de la evolución molecular. Cómo las mutaciones que se dan sobre el ADN afectan el funcionamiento de las proteínas dentro de la célula ha sido tema de discusión de centenares de artículos. Así mismo, los cambios en los genes que han ocurrido a lo largo de millones de años de evolución han dado como resultado una gama de organismos con cualidades extraordinarias y tan diversos que parecerían no tener relación entre sí. Establecer las relaciones entre las especies, así como determinar las mutaciones que cambian la función de las proteínas, resulta tarea difícil debido al tiempo que pasa entre cambios apreciables. La evidencia fósil ha ayudado a dilucidar la evolución de caracteres que se presentan matizados a lo largo de esta línea evolutiva. Sin embargo, no siempre se tiene disponibilidad de esta evidencia fósil, ya que, obtener muestras de ADN ancestral para su estudio resulta posible sólo para ejemplares de unos cuantos miles de años atrás (Gumulya & Gillam, 2017).

Sin la posibilidad de tomar las muestras de tiempos ancestrales, se ha desarrollado una técnica para inferir precursores de proteínas existentes llamada reconstrucción de secuencias ancestrales (ASR por sus siglas en inglés). Esta técnica, es capaz de inferir secuencias de proteínas que pudieron existir hace mucho tiempo y evaluar sus características en el laboratorio. Esta comparativa entre la proteína actual y la ancestral (análisis vertical) proporciona una investigación de la evolución molecular diferente a lo que se realiza usualmente, es decir, la comparación de genes ortólogos y parálogos (análisis horizontal). Esta técnica representa una ventaja para determinar las mutaciones causales de un cambio de función, ya que al comparar el ancestro con la proteína actual se observan menos mutaciones neutrales, permitiendo detectar mutaciones que proporcionen cambios de función fácilmente (Field & Matz, 2009). De manera simplificada, la ASR funciona de la siguiente manera: si quisiéramos saber cómo fue un carácter morfológico en el ancestro de un grupo podemos inferirlo a partir de los caracteres de las especies que podemos encontrar actualmente (Thornton, 2004). Se requieren secuencias de nucleótidos o de proteína que existen actualmente y que sean descendientes del ancestro que se desea conocer. Se propone cómo fue la evolución de los aminoácidos o bases nucleotídicas a lo largo del tiempo con un modelo de sustitución. Se determinan las relaciones filogenéticas de las secuencias mediante un árbol filogenético y, por último, se infiere con cierta probabilidad la secuencia del ancestro basándose en el árbol filogenético y el modelo de sustitución. Finalmente, con las técnicas modernas de biología molecular es posible sintetizar la secuencia ancestral predicha y estudiar las propiedades de la proteína asociada a esa secuencia.

La técnica de reconstrucción de secuencias ancestrales fue planteada desde 1963 (Pauling & Zuckerkandl, 1963), pero no fue sino hasta 1990 que se pudieron realizar las primeras reconstrucciones con el método de máxima parsimonia. La reconstrucción por el método de máxima parsimonia resulta inexacta cuando se realizan reconstrucciones muy antiguas y cuando las secuencias comienzan a perder similitud. Por lo tanto, no se recurre a la reconstrucción de secuencias por este método a menos que las reconstrucciones sean recientes,

como en reconstrucciones de genomas virales (Zhang & Nei, 1997). Posteriormente, Yang (1995) propuso un nuevo método estadístico para su estimación basándose en la máxima verosimilitud, que ofrece mejores resultados y un modelo más realista de evolución que el método de máxima parsimonia.

La inferencia por máxima verosimilitud evalúa cada posible estado ancestral y calcula su verosimilitud, es decir, la probabilidad de obtener un conjunto de secuencias actuales dado un estado ancestral, un árbol filogenético y un modelo de evolución molecular (Yang, Kumar & Nei, 1995). La máxima verosimilitud presenta ventajas sobre la técnica de máxima parsimonia, ya que proporciona información sobre el proceso de evolución molecular que incluye la longitud de las ramas. Además, se puede escoger el modelo evolutivo que mejor se ajuste al conjunto de datos en lugar del modelo fijo de la máxima parsimonia, es decir, aquel que implique el menor número de cambios (Thornton, 2004).

La reconstrucción de secuencias por el método de máxima verosimilitud sigue la siguiente ecuación:

$$f(x_i | y_i, \tau_j, \nu_j, \theta) = \pi_{x_{is}} p_{y_i(2s-2)x_{is}}(\nu_{2s-2}) \times \left( \prod_{k=1}^{s-1} p_{y_{i\sigma(k)}x_{ik}}(\nu_k) \right) \left( \prod_{k=s+1}^{2s-3} p_{y_{i\sigma(k)}y_{ik}}(\nu_k) \right) \quad (1)$$

La cual indica que la probabilidad de observar los datos en las puntas de los arboles ( $x_i$ ) es condicional a los estados asignados en el interior de los nodos del árbol ( $y_i$ ), la topología del árbol ( $\tau_j$ ), la longitud de las ramas ( $\nu_j$ ) y los parámetros del modelo de sustitución ( $\theta$ ). En la ecuación también se muestra el radio de transiciones/transversiones ( $\kappa$ ) y las frecuencias base al equilibrio ( $\pi$ ), propios del modelo de sustitución de nucleótidos que se utilice. Por último,  $s$  se refiere al estado que se está analizando.

Además de la técnica de máxima verosimilitud, existe otro método estadístico utilizado para la ASR denominado inferencia bayesiana. Esta inferencia, a diferencia del método de máxima verosimilitud, tiene por objetivo calcular la probabilidad posterior de que un nodo ancestral presente cierto estado, condicionado por los datos (Huelsenbeck & Bollback, 2001). Para esta metodología se utiliza, al igual que en la máxima verosimilitud, un conjunto de secuencias. Sin embargo, existe incertidumbre para la integración de valores que se muestran en la ecuación 2. Para solucionarlo, se utiliza la metodología de cadenas de Markov Monte Carlo (MCMC) para la integración sobre la incertidumbre del árbol filogenético, la longitud de las ramas y los parámetros de sustitución para el problema de la estimación del estado ancestral. Posteriormente, se corren cadenas sobre la topología, la longitud de las ramas y los parámetros del modelo de sustitución.

La ecuación de la inferencia bayesiana jerárquica tiene la siguiente forma:

$$f(y_{ij}|\mathbf{X}) = \frac{f(\mathbf{X}|y_{ij})\pi_{y_{ij}}}{\sum_{y_{ij} \in \{A,C,G,T\}} f(\mathbf{X}|y_{ij})\pi_{y_{ij}}} \quad (2)$$

donde:

$$f(\mathbf{X}|y_{ij}) = \sum_{k=1}^{B_c(s)} \int_{\mathbf{v}_k} \int_{\boldsymbol{\theta}} f(\mathbf{X}|y_{ij}, \tau_k, \mathbf{v}_k, \boldsymbol{\theta}) \times f(\tau_k) f(\mathbf{v}_k) f(\boldsymbol{\theta}) d\mathbf{v}_k d\boldsymbol{\theta} \quad (3)$$

Aquí vemos que la probabilidad del nodo ancestral ( $y_{ij}$ ) dado los datos ( $\mathbf{X}$ ), es igual a una función de los datos dado el estado ancestral, éste a su vez se refiere a la suma de la función de los datos dado el estado ancestral ( $y_{ij}$ ), la topología del árbol ( $\tau_k$ ), la longitud de las ramas ( $\mathbf{v}_k$ ) y los parámetros del modelo de sustitución ( $\boldsymbol{\theta}$ ), es decir, la suma de todos los posibles árboles consistentes con la restricción.  $\pi$  también se refiere a las frecuencias base al equilibrio.

La ASR ha generado resultados interesantes en el campo de la evolución molecular. Se tienen varios ejemplos de esta técnica aplicada a problemas de diversa índole, como en la caracterización de ancestros de fluoróforos de especies de corales (Ugalde *et al.*, 2003), una aproximación a la temperatura de la tierra ancestral mediante la correlación entre temperatura de desnaturalización de factores de elongación y la temperatura a la que vive el organismo (Gaucher *et al.*, 2008), la caracterización de un ancestro de un receptor esteroideo que cambió la preferencia de uso de aldosterona a cortisol (Ortlund *et al.*, 2007) y la determinación de la promiscuidad y termoestabilidad de las enzimas beta-lactamasas ancestrales (Risso *et al.*, 2013).

Al ser un modelo basado en probabilidad, a cada sitio reconstruido de la secuencia ancestral se le asocia una probabilidad, que en estadística bayesiana es llamada como probabilidad posterior. Ésta se define como la probabilidad de que una hipótesis es verdadera después de que los datos son incorporados a la experiencia (Thornton, 2004) y la probabilidad posterior de un sitio se calcula a partir de la verosimilitud de ese estado (ponderado por su frecuencia al equilibrio, que sirven como probabilidad previa) dividido por la suma de las verosimilitudes previas ponderadas de los posibles 20 estados en ese sitio (Eick *et al.*, 2016).

Debido a este rasgo probabilístico, es muy improbable contar con la secuencia correcta del ancestro. Por ejemplo, una proteína de 250 aminoácidos, con 90% de sitios se reconstruidos con una alta probabilidad posterior, digamos 0.9 y el 10% con una probabilidad un poco menor, digamos 0.8. Para obtener la probabilidad posterior de la secuencia, se realiza un promedio de la probabilidad posterior de todos los sitios, por lo tanto, la probabilidad posterior de la secuencia es de 0.89. Se considera que la probabilidad posterior de una secuencia es buena al ser mayor a 0.8. Para determinar si la probabilidad de que la secuencia que hemos reconstruido sea la correcta es necesario multiplicar cada una de las probabilidades posteriores. Al realizar

el cálculo para el ejemplo anterior, se obtiene una probabilidad  $1.91 \times 10^{-13}$ , es decir, una probabilidad de 1 en 5 billones de que la secuencia que se propone haya sido en realidad la secuencia ancestral.

A pesar de que se tiene una probabilidad muy baja de obtener la secuencia correcta del ancestro, esto no quiere decir que no se pueda obtener el fenotipo de la proteína ancestral, aunque no se conozca con exactitud el genotipo. Sin embargo, ¿Cómo aumentamos la certidumbre sobre el fenotipo ancestral sin importar la incertidumbre del genotipo asociados al método?

Para esto, se debe disminuir la incertidumbre ante las características del propio método, por lo que puede recurrirse a diferentes estrategias como aumentar la probabilidad posterior de la reconstrucción. Se puede trabajar con diferentes metodologías de reconstrucción que cuenten con distintos acercamientos, ya que los métodos de reconstrucción sobrestiman o subestiman ciertas características y finalmente disminuir la incertidumbre de ciertos sitios con altas probabilidades de reconstrucción para dos o más aminoácidos, denominados sitios ambiguos.

La verosimilitud del ancestro se define como la probabilidad de que los estados que se encuentran presente hoy en día en las puntas de los arboles filogenéticos hayan evolucionado a partir de este estado ancestral. Como se observa en la ecuación (1), la verosimilitud del estado actual depende del estado ancestral (lo que se está buscando), la longitud de las ramas, la topología, y el modelo de sustitución. Por lo tanto, la verosimilitud de la secuencia ancestral, y así mismo su probabilidad posterior, están condicionadas por los parámetros que definen al árbol filogenético, es decir, las secuencias y el alineamiento.

Hay estudios que indican que un mayor muestreo aumenta la calidad y precisión de un árbol filogenético (Pollock *et al.*, 2002; Hills, 1998), mientras que también los hay indicando lo contrario o que simplemente no afecta (Kim, 1998; Rosemberg & Kumar, 2001). Si bien es necesario alimentar al modelo con la información necesaria para realizar un análisis más detallado de los procesos ocurridos en la evolución de una proteína, el aumento innecesario de información puede llevar a un sesgo por parte de las secuencias empleadas y además aumentar de manera innecesaria el tiempo de cómputo. Esto queda demostrado con un ejemplo por parte de Benner (2007), en donde seis proteínas tienen una metionina en el sitio 39, mientras que 18 secuencias poseen una leucina en el mismo sitio. Sin embargo, es un sesgo si determinamos que las 18 secuencias pertenecen a la misma raza de buey, mientras que las otras 6 secuencias pertenecen a las especies de búfalo, antílope, ciervo, oveja, impala y camello. Por lo tanto, una estrategia para una selección de secuencias útiles y suficientes es tratar de disminuir la redundancia de las secuencias.

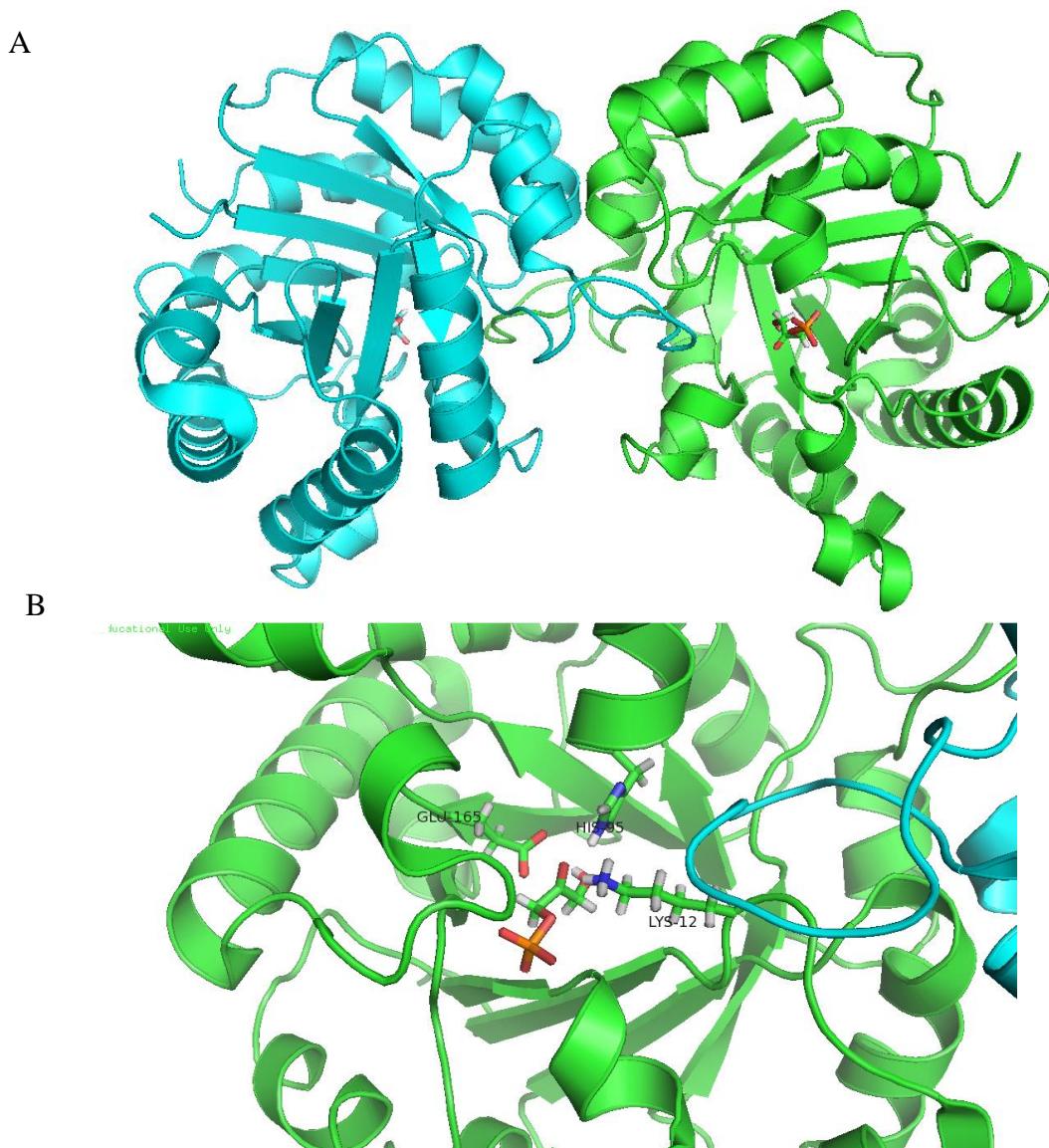
Hay que tomar en cuenta que no todos los sitios se reconstruyen con una alta probabilidad posterior, incluso se llega a observar que existen sitios donde se seleccionan dos o más aminoácidos con una alta probabilidad posterior. La incertidumbre de la secuencia ancestral aumenta exponencialmente con el número de sitios, denominados sitios ambiguamente reconstruidos.

Existen ciertos sitios que se reconstruyen ambiguamente, es decir, que tienen altas probabilidades de poseer tanto un aminoácido como otro. Por lo tanto, la certidumbre de estos sitios no es robusta y pueden no representar con alta probabilidad la secuencia ancestral (Akanuma *et al.*, 2013). La mejor estrategia para evaluar estos sitios ambiguos es sintetizar el gen variando uno a uno dichos sitios y determinar las características de la proteína reconstruida, determinando que tanto afecta cada uno de los sitios ambiguos al fenotipo de la secuencia. Sin embargo, esta estrategia es muy costosa tanto por la síntesis de genes, como por el trabajo experimental requerido.

Para evaluar la robustez de la reconstrucción que concierne a estos sitios ambiguos se puede recurrir a otra estrategia llamada peor caso plausible (Akanuma *et al.*, 2013). La construcción de las peores secuencias plausibles (*worst plausible case* en inglés) se realiza de la siguiente manera. Si algún sitio posee más de un aminoácido con una probabilidad posterior mayor a 0.2, se considera como sitio ambiguamente reconstruido y para obtener la peor secuencia plausible se toma el segundo mejor aminoácido (el aminoácido con la segunda mejor probabilidad posterior) para todos los sitios ambiguamente reconstruidos. Esta secuencia tiene más errores y puede ser diferente a la secuencia más verosímil. Por lo tanto, su caracterización representa una prueba de robustez funcional de la incertidumbre de la secuencia; es decir, si la secuencia más verosímil y la peor secuencia plausible tienen la misma función, es más probable que la secuencia correcta haya compartido esta misma función (Eick *et al.*, 2016).

### **Triosafosfato isomerasa**

La triosafosfato isomerasa (TIM) (Figura 1A) es una enzima con estructura barril TIM ( $\beta/\alpha$ )<sub>8</sub>, una de las topologías más frecuentes encontradas en la naturaleza (Farber & Petsko, 1990). La TIM participa en la glucólisis, la gluconeogénesis y síntesis de triglicéridos. La TIM cataliza, sin cofactores, la interconversión entre la dihidroxiacetona fosfato y el gliceraldehido-3-fosfato. La TIM posee tres aminoácidos (Figura 1B), responsables de la catálisis de isomerización. Se trata de los residuos Lys12, His95 y Glu165 (según la numeración de la TIM de levadura) ubicados en las asas superiores 1, 4 y 6 respectivamente.



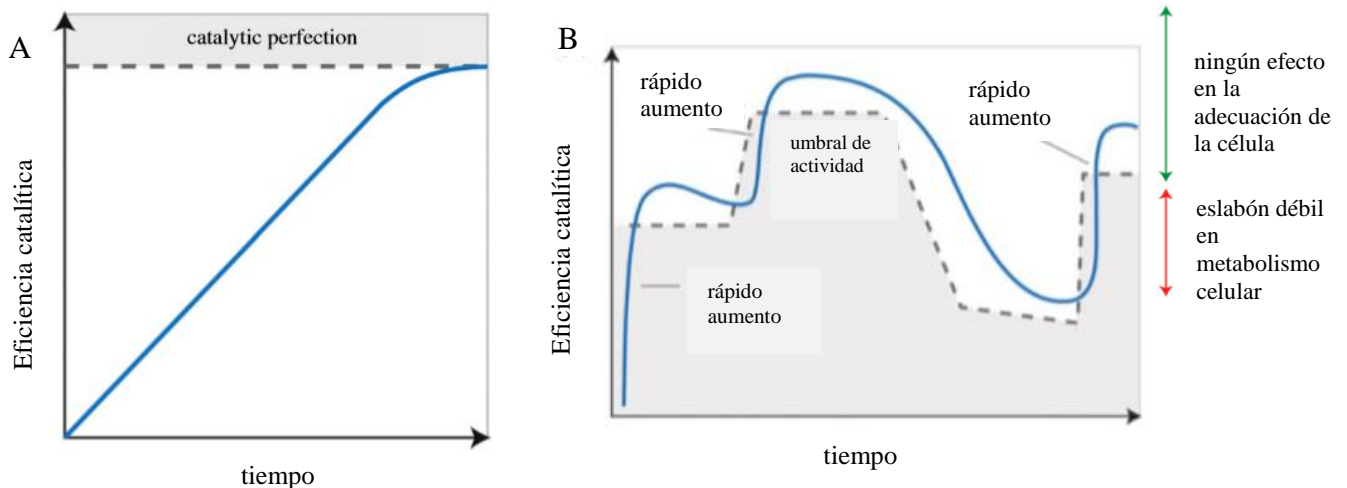
**Figura 1.-** (A) Estructura de la TIM de *Saccharomyces cerevisiae* (1NEY) con sus dos monómeros en azul y verde. (B) Sitio activo de la proteína donde se muestran los 3 aminoácidos catalíticos K12, H95 y E165 interaccionando con el análogo de sustrato fosfoglicohidroxamato (PGH) .

Esta enzima funciona únicamente como oligómero, de forma dimérica en el caso de eucariontes y procariontes y de manera dimérica o tetramérica en el caso de algunas arqueas. La oligomerización obligada ocurre a pesar de que el monómero posee por sí mismo todos los aminoácidos del sitio catalítico. La razón por la que se requiera obligatoriamente la oligomerización de la TIM no se ha esclarecido por completo, sin embargo, se ha propuesto que el monómero posee una estabilidad termodinámica menor y la dimerización aumenta su estabilidad conformacional (Mainfroid *et al.*, 1996). Otra posible explicación radica en la interfaz de los monómeros y su estabilización mediante interacciones de tipo puente de hidrógeno (Schliebs *et al.*, 1997), puentes salinos y de residuos con moléculas de agua (Rodríguez-Almazan *et al.*, 2008). Todas estas interacciones de la interfaz se dan principalmente en el asa 3, que se encuentra íntimamente relacionada con sitio activo del monómero contrario.



Existen TIMs que se encuentran en un estado tetramérico, entre ellas *Thermotoga marítima* (Maes *et al.*, 1999), *Pyrococcus woesei* (Walden *et al.*, 2001), *Thermoproteus tenax* (Walden *et al.*, 2004), *Giardia lamblia* (Reyes-Vivas *et al.*, 2007), *Methanocaldococcus jannashii* (Gayathri *et al.*, 2007) y *Pyrococcus furiosus* (Sharma & Guptasarma, 2015). Se ha discutido la probabilidad de que la tetramerización haya surgido como un mecanismo para aumentar la termoestabilidad de las proteínas, ya que se observan que algunos de estos organismos viven en ambientes de altas temperaturas. Sin embargo, también existe la posibilidad de que tengan otra función, ya que se observa que en *G. lamblia*, la TIM cambia el estado oligomérico dependiendo del estado en que se encuentre el parásito.

Es posible calcular la constante de segundo orden para la asociación entre enzimas y sustratos. El límite teórico de catálisis a la que una enzima puede llegar se encuentra en el orden de  $10^8$ - $10^9$   $M^{-1} s^{-1}$  (Alberty & Hammes, 1958). La reacción de isomerización de la TIM es tan efectiva que sólo está limitada por difusión, ya que el valor de  $k_{cat}/K_m$  se encuentra en el orden del límite por difusión, es decir, el paso limitante para la catálisis es la difusión del sustrato al sitio de la enzima. Esta propiedad hizo que Alberty y Knowles plantearan que esta enzima ha alcanzado la “perfección catalítica” (Alberty & Knowles, 1976), argumentando que el objetivo de la evolución sobre las enzimas es que alcancen el límite por difusión (Figura 2A). Esta visión determinista puede no ser el verdadero camino que siguen las enzimas, en especial aquellas que han alcanzado la “perfección catalítica”. Newton, Arcus y Patrick (2015) presentan una visión diferente sobre la evolución de la catálisis de las proteínas. El camino que estos autores proponen para la evolución de una enzima comienza con una eficiencia catalítica muy baja que aumenta rápidamente hasta llegar a ser limitada por difusión. Posteriormente, la eficiencia catalítica se adecua a las necesidades de la célula siempre por encima de un umbral, que por debajo de este afecta la viabilidad de la célula. Hay que tomar en cuenta que el umbral que sea necesario mantener no será siempre el mismo a lo largo del tiempo o entre diferentes células (Figura 2B).



**Figura 2.-** Propuestas de evolución de la actividad de una enzima. A, Modelo más sencillo y determinista donde las enzimas evolucionan de una baja actividad a una alta actividad hasta alcanzar el límite por difusión. B, Modelo hipotético donde se muestra la rápida y temprana evolución de una enzima hasta alcanzar el límite por difusión y poco a poco va disminuyendo sin quedar por abajo del umbral mínimo de actividad (Newton *et al.*, 2015).

Suponiendo que, sí la hipótesis de Albery y Knowles es cierta, las enzimas ancestrales tenderían a una mayor eficiencia catalítica. Por lo tanto, en el pasado existirían pocas enzimas con alta eficiencia catalítica. Gracias a la técnica de ASR, se determinó que la eficiencia de varias enzimas ancestrales como, por ejemplo, la eficiencia de la  $\alpha$ -glicosidasa fue más alta que de enzimas actuales, con estos resultados Newton y colaboradores (2015) propusieron que la actividad de las enzimas se debe mantener sobre un umbral mínimo de actividad más que llegar a la “perfección catalítica”. Es interesante determinar el camino por el cual la TIM llegó a esta eficiencia catalítica limitada por difusión.

## Antecedentes

En nuestro laboratorio se recurrió a la reconstrucción ancestral de secuencias para determinar el camino de la eficiencia catalítica que tomó la TIM durante su evolución, además de otras características que parecían interesantes investigar, cómo encontrarse en los 3 dominios de la vida con la misma estructura terciaria, aunque con una identidad de secuencia hasta del 12% (entre *Plasmodium falciparum* y *Pyrococcus woesei*) y diferentes vías de plegamiento. La reconstrucción se realizó en 2011 por Pardo-Ávila, y la caracterización de estos resultados por Schulte-Sasse (Schulte-Sasse *et al.*, 2019). La ASR se realizó con 44 secuencias de eucariontes, adecuándolas a los árboles establecido por Ciccarelli (2006) y Luzoni (2004) y la reconstrucción por el método de máxima verosimilitud.

Los genes correspondientes a las secuencias de los nodos de ancestros de animales, hongos y el ancestro de ambos (opistoconte) se sintetizaron comercialmente. Las proteínas respectivas fueron sobrepresadas y se purificadas Las proteínas ancestrales caracterizadas de estos nodos poseen estructura secundaria y terciaria característicos del barril TIM, forman dímeros compactos, poseen temperaturas medias de desnaturalización altas y capacidad de realizar la catálisis de isomerización. Sin embargo, su catálisis se lleva a cabo con una eficiencia catalítica menor, incluso en dos órdenes de magnitud, (para TIM63, Tabla 1), con respecto a las enzimas actuales. Las enzimas ancestrales también presentaron una baja estabilidad a la dilución, que sugiere una menor estabilidad del dímero. Además, se observaron otras características en el ancestro de los opistocontes (TIM63) como la reversibilidad en el proceso de dilución, y que, al disociarse, los monómeros aún poseen sus residuos aromáticos aislados en el núcleo hidrofóbico lo que sugiere que el monómero no sufre cambios drásticos en la estructura al disociarse, cualidad muy rara en las TIM de eucariontes actuales.

**Tabla 1.-** Resumen de resultados del trabajo de reconstrucción de secuencias ancestrales de la TIM de Schulte-Sasse y colabores (2019). Se muestran resultados de estructura secundaria y terciaria, estado oligomérico, catálisis, temperatura media de desnaturalización y estabilidad a la dilución.

| TIM                | DC <sup>b</sup> /Rayos-X <sup>c</sup> (%) |       |             |             | $\lambda_{\max}$ | Rs (Å) | Parámetros catalíticos para la isomerización del G3P  |                     |   | Tm (°C)           | K <sub>0.5</sub> <sup>f</sup> (μM) |
|--------------------|---|-------|-------------|-------------|------------------|--------|---|---------------------|---|-------------------|------------------------------------|
|                    | Hélice                                    | Hebra | Vuelta beta | Random coil |                  |        | k <sub>cat</sub> (x 10 <sup>3</sup> s <sup>-1</sup> ) | K <sub>m</sub> (mM) | k <sub>cat</sub> /K <sub>m</sub> (M <sup>-1</sup> s <sup>-1</sup> ) |                   |                                    |
| ScTIM <sup>a</sup> | 38/45                                     | 13/16 | 16/19       | 30/19       | 320              | 30.0   | 6.40 <sup>d</sup>                                     | 1.4 <sup>d</sup>    | 1.14 x 10 <sup>8</sup>  | 59.1 <sup>e</sup> | << 0.001 <sup>g</sup>              |
| TIM63              | 41/48                                     | 12/15 | 16/18       | 29/20       | 330              | 29.5   | 0.96 ± 0.05   | 2.82 ± 0.34         | 8.51 x 10 <sup>6</sup>  | 65.7 ± 0.03       | 0.2                                |
| TIM64              | 42/ND                                     | 11/ND | 15/ND       | 30/ND       | 328              | 30.0   | 2.80 ± 0.09   | 1.98 ± 0.14         | 3.54 x 10 <sup>7</sup>  | 65.7 ± 0.03       | 0.05                               |
| TIM65              | 41/ND                                     | 11/ND | 15/ND       | 30/ND       | 323              | 30.5   | 2.88 ± 0.06   | 0.43 ± 0.03         | 1.67 x 10 <sup>8</sup>  | 54.2 ± 0.10       | 0.05                               |
| TIM86              | 31/ND                                     | 19/ND | 17/ND       | 36/ND       | 329              | 29.9   | 1.83 ± 0.07   | 1.51 ± 0.13         | 3.04 x 10 <sup>7</sup>  | 66.3 ± 0.05       | 0.03                               |
| TIM87              | 35/ND                                     | 15/ND | 16/ND       | 34/ND       | 327              | 29.9   | 2.19 ± 0.08   | 1.59 ± 0.13         | 3.44 x 10 <sup>7</sup>  | 65.8 ± 0.03       | 0.03                               |

ScTIM, TIM de *Saccharomyces cerevisiae*; TIM63, TIM del ancestro de todos los opistocontes; TIM64, TIM del ancestro de animales bilaterales; TIM65, TIM del ancestro de vertebrados con mandíbula; TIM86, TIM del ancestro del orden de los Saccharomycetales; TIM87, TIM del ancestro de la familia Saccharomycetaceae.

<sup>a</sup>PDB ID: 7TIM. <sup>b</sup> Valores obtenidos de la deconvolución del espectro con el programa CDNN. <sup>c</sup> Valores obtenidos con el servidor STRIDE (Heinig & Frishman, 2004). <sup>d</sup> Peimbert *et al.*, 2008. <sup>e</sup> Benítez-Cardoza *et al.*, 2001. <sup>f</sup> Concentración de proteína requerida para la inactivación de la mitad de la población. <sup>g</sup> Nájera *et al.*, 2003. Modificado de Schulte-Sasse *et al.*, 2019.

Estas características sugirieron la posibilidad de que la TIM ancestral tuviera un monómero más estable y robusto que las TIM actuales, y que, en algún momento de su historia pudo funcionar como monómero. Retomando la discusión anterior, encontrar un fenotipo distinto al esperado en una proteína ancestral es un claro ejemplo de la necesidad aumentar la certidumbre del fenotipo de la proteína ancestral debido a la incertidumbre metodológica, ya que no podemos aseverar que estas características hayan pertenecido al verdadero ancestro. Asegurar la robustez de la reconstrucción ancestral, por lo tanto, es parte fundamental del método (Hochberg & Thornton, 2017).

La importancia de este trabajo radica en reconstruir uno de los nodos caracterizados anteriormente (el último ancestro de los opistocontes) con mayor número de secuencias y con una mejor selección de éstas, además de realizar dos metodologías diferentes para la reconstrucción, la máxima verosimilitud y el método bayesiano. También se incluyó la estrategia denominada peor caso plausible para evaluar la robustez de las secuencias ante la incertidumbre de los sitios ambiguos.

## **Justificación**

La TIM63 (ancestro de los opistocontes) de la anterior reconstrucción presentó una baja estabilidad a la dilución y eficiencia catalítica, además de monómeros más estructurados, características inusuales en el grupo de los eucariontes. Para determinar si éstas y otras características son robustas, sin importar la incertidumbre del método, se propone realizar una nueva reconstrucción sobre un mismo nodo (el ancestro de los opistocontes), con una nueva base de datos, un mejor manejo de la redundancia de secuencias, diferentes metodologías de reconstrucción y también usando la estrategia de peor caso plausible.

## **Hipótesis**

La reconstrucción de un solo nodo con distintos métodos de reconstrucción y variando los sitios ambiguos darán lugar a enzimas con características fisicoquímicas, cinéticas y estructurales similares. Aquellas características que no tienen variación significativa, nos proporcionarán información confiable acerca de la evolución de la proteína.

## **Objetivos**

Realizar la reconstrucción de secuencias ancestrales de la triosafosfato isomerasa del ancestro común de los opistocontes mediante el uso del método de máxima verosimilitud e inferencia bayesiana con sus respectivas peores secuencias plausibles y determinar la conservación de características catalíticas, fisicoquímicas y estructurales de las proteínas.

## **Objetivos particulares**

- Realizar la reconstrucción de secuencias ancestrales mediante las metodologías de máxima verosimilitud e inferencia bayesiana con sus respectivos peores casos plausibles.
- Sobrexpresar y purificar las enzimas ancestrales.
- Determinar los parámetros catalíticos, estado oligomérico, inactivación por dilución y temperatura media de desnaturalización
- Comparar las 4 proteínas reconstruidas y determinar qué parámetros son robustos a la reconstrucción de secuencias.
- Comparar los datos obtenidos con la anterior reconstrucción.

## **Metodología**

### **Reconstrucción de secuencias ancestrales**

La metodología computacional para la reconstrucción de secuencias ancestrales tiene 5 pasos:

- La búsqueda de secuencias actuales
- El alineamiento múltiple de secuencias
- La determinación del mejor modelo de sustitución
- La obtención del árbol filogenético

-La estimación de las secuencias ancestrales

Posteriormente se evalúa experimentalmente las proteínas predichas.

### **Selección de secuencias**

Las secuencias de proteínas de la TIM fueron seleccionadas mediante una búsqueda por palabra en la base de datos del *National Center for Biotechnology Information* (NCBI), con la búsqueda "Triosephosphate isomerase". La búsqueda por sí sola da 92295 entradas de proteínas (marzo, 2017), combinando especies de eucariontes, procariontes y arqueas. El número de secuencias de eucariontes fue de 9206 entradas. Posteriormente, se procedió a filtrar las secuencias adecuadas de los organismos ya que la mayoría de las secuencias se encuentran repetidas y existen secuencias incompletas. Además, se descartaron aquellas secuencias que fueran fragmentos y mutantes de la proteína. Se buscó una única secuencia por especie, cuyos tamaños se encontrará alrededor de los 250 aminoácidos. De encontrarse dos secuencias de un mismo organismo, pero distinta secuencia, se seleccionó aquella que tuviera un porcentaje de identidad mayor al 70% con un organismo cercanamente emparentado. Es importante resaltar que los peces con mandíbula poseen una isoenzima neural de la TIM. En este caso, sólo se agregaron aquellas pertenecientes a la isoenzima general y no la neural (Merrit & Quattro, 2001). Se eliminaron aquellas secuencias que carecieran de cualquiera de los aminoácidos catalíticos, que tuvieran inserciones o deleciones muy grandes, que no tuviera algún otro organismo cercanamente relacionado y que tuviera aminoácidos indeterminados (estos se señalizan con una letra X)

Con estos criterios, se obtuvieron de 907 secuencias de la TIM de eucariontes. Sin embargo, se realizó otro proceso de filtración para eliminar secuencias que no cumplían los criterios anteriormente mencionados y que no habían sido detectadas en el primer filtro.

Para apoyar a la selección de secuencias, se recurrió al programa GUIDANCE (v. 2.02; Sela, Katoh & Pupko, 2015). Este programa permite determinar sitios del alineamiento y secuencias poco confiables mediante *bootstrap*. El *bootstrap* genera árboles a partir de alineamientos donde se eliminan ciertos sitios y se calcula para cada par de residuos en el alineamiento base, la fracción de pseudoalineamientos en la que un par de residuos fueron alineados juntos. Se corrió el programa con el algoritmo GUIDANCE2, y con el algoritmo de alineamiento MAFFT. Con base en esto, se eliminaron las secuencias que tuvieran un score menor a 0.8 y que no cumplieran con los criterios de selección mencionados anteriormente (falta de aminoácidos catalíticos, in-del, etc.).

Otro programa utilizado para depurar secuencias fue CD-HIT (v. 4.6; Ying *et al.*, 2010), un programa en línea que permite visualizar secuencias redundantes de un compilado de secuencias, dependiendo del porcentaje de identidad. El programa se ajustó para que agrupara secuencias con un 100% de identidad. Esta estrategia se realizó para eliminar secuencias de organismos a nivel de subespecie que fueran redundantes.

Se obtuvieron 718 secuencias tras correr los dos programas anteriores, sin embargo, se decidió reducir aún más el número de secuencias por varios motivos como: reducir el tiempo de cómputo, aumentar la verosimilitud del árbol al reducir las posibles topologías que pueden existir y principalmente, tener una mejor representación de los grupos sin tantas secuencias redundantes.

Por lo tanto, se eligió un límite de corte más grande (0.8) para el programa de CD-HIT, lo que resultó en un número final de 239 secuencias.

### **Alineamiento múltiple de secuencias**

Para generar el árbol filogenético por los métodos de máxima verosimilitud e inferencia bayesiana, es necesario contar con un alineamiento de secuencias. El alineamiento múltiple de secuencias es una herramienta que permite alinear nucleótidos o aminoácidos de características similares en la misma columna. Las secuencias se alinearon con el programa MAFFT (v. 7; Katoh, Misawa, Kuma, & Miyata, 2002), con el algoritmo FFT-NS-i y los demás valores predeterminados (matriz BLOSUM62, penalización por in-del 1.53). El alineamiento resultante presentaba menor cantidad de in-del's comparado con MUSCLE, algoritmo de alineamiento utilizado en la anterior reconstrucción.

### **Selección de modelo de sustitución**

El modelo de sustitución indica de qué manera los aminoácidos cambian entre ellos, lo que es fundamental para establecer la probabilidad de que un aminoácido cambie por otro al realizar la reconstrucción de secuencias. A partir de la página en línea ATGC, *Montpellier Bioinformatics Platform* (<http://www.atgc-montpellier.fr/phyml/>) se puede encontrar el mejor modelo de sustitución que se ajuste a los datos con el programa *Smart Model Selection* (SMS; Lefort *et al.*, 2017). En lugar de probar todas las combinaciones de modelos y seleccionar el mejor, esta herramienta realiza un estudio heurístico para evitar estos cálculos tan extensos. La selección del mejor modelo de sustitución se hace con base en un parámetro llamado Criterio de Información de Akaike (AIC). Este estimador mide la distancia esperada entre el modelo evaluado y el modelo estimado a partir de las secuencias. El AIC se puede interpretar como la cantidad de información que se pierde cuando se usa cierto modelo, para aproximarse al proceso real de evolución molecular (Salemi, Vandamme & Lemey, 2009).

### **Árbol filogenético**

La construcción del árbol filogenético se puede realizar de diferentes maneras, desde utilizar métodos de distancia como lo son UPGMA y *Neighbor Joining*, hasta métodos basados en caracteres como máxima parsimonia, máxima verosimilitud e inferencia bayesiana (Hall, 2011). La elección del método para realizar el árbol se basa en las preferencias del usuario y en el objetivo del árbol. En este caso, ya que se quiere realizar una reconstrucción ancestral de secuencias, los métodos de distancia ofrecen poca resolución ya que toman en cuenta únicamente los sitios que difieren entre secuencias pareadas, sin embargo, no se toma en cuenta la historia por la que se dieron esas sustituciones. El método de máxima parsimonia tampoco

es útil para la reconstrucción de secuencias, ya que las secuencias no siempre evolucionan de manera parsimoniosa, es decir, no siempre se sigue el camino de mínimos cambios durante la evolución de las proteínas. Por lo tanto, se recurrió a la construcción de árboles mediante los métodos de máxima verosimilitud y de inferencia bayesiana.

Para el árbol de máxima verosimilitud, se realizó el árbol filogenético en la página ATGC, *Montpellier Bioinformatics Platform*, mediante el programa PhyML (v. 3.0), que también implementa el programa anteriormente mencionado, el SMS para selección de modelo. Por lo tanto, el árbol implementa el modelo de sustitución con mayor AIC. Se seleccionó la realización del árbol inicial por el método de *Neighbor Joining*, se indicó que realizara el árbol a partir de 10 árboles raíz y, por último, se optimizaron las ramas mediante la prueba *Approximate Likelihood-Ratio Test* (aLRT), ya que el portal no permitió realizar la prueba por *bootstrap* debido al alto tiempo de cómputo.

Para el árbol por inferencia bayesiana, se recurrió al programa MrBayes (v. 3.2.6; Huelsenbeck & Ronquist, 2001 & 2003), a través del portal CIPRES Science Gateway (V. 3.3; Miller *et al.*, 2010). El portal CIPRES contiene un compilado de programas bioinformáticos y tiene a su disposición una supercomputadora para acortar tiempos de cómputo. Al mismo tiempo, MrBayes realizó tanto la estimación del árbol filogenético como la reconstrucción de estados ancestrales. Los parámetros que indican cómo se realizará el árbol y la reconstrucción, se incluyeron dentro del archivo que contiene el alineamiento en formato NEXUS, llamado bloque de datos de MrBayes (Tabla 2).

**Tabla 2.-** Bloque de datos utilizado para la inferencia del árbol filogenético y la reconstrucción de secuencias ancestrales por inferencia bayesiana.

|  |
|--|
| <b>begin mrbayes;</b>  |
| <b>prset aamodelpr=fixed(lg);</b>  |
| <b>outgroup outEcoli;</b>  |
| <b>constraint opisthokont = 1-184;</b>   |
| <b>prset pinvarpr=fixed(0.102) topologypr = constraints(opisthokont);</b>                      |
| <b>report applyto = (1) ancstates = yes;</b>   |
| <b>mcmc ngen=10000000 printfreq=50000 samplefreq=5000 nchains=4 temp=0.2 checkfreq = 50000</b> |
| <b>diagnfreq = 1000 stopval = 0.01 stoprule = yes;</b>   |
| <b>sump relburnin = yes burninfrac =0.25;</b>  |
| <b>sumt relburnin = yes burninfrac =0.25;</b>  |
| <b>end;</b>  |

Dentro del portal existe una opción para indicar que el archivo contiene el bloque de datos anteriormente mencionado. También se indicó el máximo número de horas que corra el programa sea de 40, un tiempo más que suficiente para que termine por completo. Por lo demás, no se ajustaron otros parámetros dentro del portal CIPRES (Miller *et al.*, 2010).

## Reconstrucción de secuencias ancestrales

La reconstrucción de secuencias ancestrales se realizó por los dos métodos estadísticos anteriormente mencionados, máxima verosimilitud e inferencia bayesiana.

Se trabajó sobre un único ancestro en lugar de la caracterización usual, donde se reconstruyen diferentes ancestros a lo largo de la línea evolutiva. Se trabajó sobre el último ancestro de los opistocontes, el ancestro más reciente de animales y hongos que se calcula vivió hace 1600-900 millones de años (Hedges *et al.*, 2006). Se trabajó con este ancestro debido a las características que presentó en la reconstrucción anterior y a que es un nodo relativamente antiguo, lo que aumenta la incertidumbre sobre la secuencia.

La reconstrucción por el método de máxima verosimilitud se realizó con el programa *codeml* del paquete PAML (v. 4.9; Yang, 1997). Para correr este programa se requieren los siguientes archivos en una sola carpeta: el alineamiento en formato PHYLIP, la matriz de sustitución (en este caso la matriz LG), el árbol que se obtuvo por máxima verosimilitud y una hoja de código llamada *codeml* donde se indican los parámetros para realizar la reconstrucción (Tabla 3).

**Tabla 3.-** Hoja *codeml* utilizada para realizar la reconstrucción de secuencias ancestrales.

|                             |   |
|-----------------------------|---|
| <b>seqfile = prueba.txt</b> | <b>* sequence data filename</b>   |
| <b>treefile = tree.txt</b>  | <b>* tree structure file name</b>   |
| <b>outfile = mlc</b>        | <b>* main result file name</b>  |
| <b>noisy = 9</b>            | <b>* 0,1,2,3,9: how much rubbish on the screen</b>  |
| <b>verbose = 2</b>          | <b>* 0: concise; 1: detailed, 2: too much</b>   |
| <b>runmode = 0</b>          | <b>* 0: user tree; 1: semi-automatic; 2: automatic<br/>* 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise</b>   |
| <b>seqtype = 2</b>          | <b>* 1:codons; 2:AAs; 3:codons--&gt;AAs</b>   |
| <b>ndata = 10</b>           |   |
| <b>clock = 0</b>            | <b>* 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis</b>   |
| <b>aaRatefile = lg.dat</b>  | <b>* only used for aa seqs with model=empirical(_F)<br/>* dayhoff.dat, jones.dat, wag.dat, mtmam.dat, or your own</b>   |
| <b>model = 2</b>            | <b>* models for AAs or codon-translated AAs:<br/>* 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F<br/>* 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)</b> |
| <b>Mgene = 0</b>            | <b>* codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff<br/>* AA: 0:rates, 1:separate</b>   |
| <b>fix_alpha = 0</b>        | <b>* 0: estimate gamma shape parameter; 1: fix it at alpha</b>  |
| <b>alpha = 0.5</b>          | <b>* initial or fixed alpha, 0:infinity (constant rate)</b>   |
| <b>Malpha = 0</b>           | <b>* different alphas for genes</b>   |
| <b>ncatG = 4</b>            | <b>* # of categories in dG of NSSites models</b>  |
| <b>getSE = 0</b>            | <b>* 0: don't want them, 1: want S.E.s of estimates</b>   |
| <b>RateAncestor = 1</b>     | <b>* (0,1,2): rates (alpha&gt;0) or ancestral states (1 or 2)</b>   |
| <b>Small_Diff = .5e-6</b>   |   |
| <b>cleandata = 0</b>        | <b>* remove sites with ambiguity data (1:yes, 0:no)?</b>  |
| <b>method = 0</b>           | <b>* Optimization method 0: simultaneous; 1: one branch a time</b>  |

Se obtuvieron múltiples documentos tras terminar el programa (que requiere de aproximadamente una noche). El archivo relevante para obtener las secuencias ancestrales es el llamado *rst*, que contiene el árbol con el que se trabajó, al que se le adicionaron números en los nodos para numerar los ancestros y relacionarlos con las secuencias reconstruidas. Más abajo del archivo se encuentran todos los nodos enumerados, según el árbol, y para cada sitio se encuentra la probabilidad posterior de reconstrucción de todos los aminoácidos (Figura 3).





encuentren en esa columna. Por lo tanto, fue necesario eliminar sitios en donde es más probable que se encuentre un in-del en lugar de un aminoácido en las secuencias ancestrales.

La eliminación de aminoácidos que corresponden a in-dels de los nodos se realizó de la siguiente manera.

Primero se realizó otra reconstrucción tomando únicamente animales y hongos, así se puede determinar las in-dels que pertenecen a este grupo y los que son externos. Para eliminar los aminoácidos que correspondientes a una in-del, se alineó la secuencia del nodo reconstruido con las demás secuencias. Se comparó la secuencia del nodo con las secuencias de sus descendientes. Si existen aminoácidos que posee cierto grupo, por ejemplo, el grupo de los primates, pero no lo tienen los demás mamíferos (asumiendo que estamos reconstruyendo al último ancestro de los mamíferos), se asume que esa inserción apareció posterior al último ancestro de los mamíferos. Por lo tanto, se eliminó ese sitio en la secuencia ancestral. Si se observa que un aminoácido está presente y ausente en diferentes organismos de un grupo, se asume que esta inserción se encontraba en el ancestro y algunos organismos la perdieron y otros la conservaron. Por lo tanto, se conservó ese sitio en la secuencia ancestral.

## **Procedimiento experimental**

### **Sobrexpresión de las proteínas ancestrales**

Los genes con la secuencia predicha mediante el proceso de reconstrucción de secuencias ancestrales se sintetizaron comercialmente por a la compañía GeneScript®. Los genes se encuentran en un plásmido de sobrexpresión pET-28b (+). Este plásmido contiene el operador de sobrexpresión inducible por lactosa, un marcador de selección de resistencia a kanamicina, un sitio múltiple de clonación, un origen de replicación y una secuencia que codifica una sucesión de seis histidinas al inicio de la secuencia. La secuencia se encuentra flanqueada por dos sitios de restricción, NdeI y XhoI en la región 5' y 3' respectivamente. Cada plásmido liofilizado, se diluyó en 50 µL de agua libre de nucleasas para obtener una concentración aproximada de 80 ng/µL. Se transformaron células competentes *E. coli* BL21 (DE3) GOLD, por choque térmico con 0.6 µL de plásmido y se plaquearon 100 µL de células transformadas. De las células que fueron capaces de integrar el plásmido, se creció un pre-cultivo toda la noche en un volumen de 10 mL de medio LB. Este volumen se agregó a 250 mL de medio fresco y se esperó a una densidad óptica entre 0.6 y 0.8 para inducir con IPTG 1 mM por toda la noche. Las células se cosecharon por centrifugación a 8000 rpm por 15 minutos (Thermo/SLA-3000). Se decantó el medio y las células se resuspendieron en el amortiguador A (35 mM de NaH<sub>2</sub>PO<sub>4</sub>, 300 mM de NaCl y 5 mM de imidazol a pH 8.0) con 0.2 mM PMSF como inhibidor de proteasas. Las células fueron lisadas por sonicación con 30 segundos de pulso y 30 segundos de descanso por 6 minutos con una amplitud del 25% en un sonicador Sonics Materials™ VC 750 Ultrasonic Processor. El lisado celular se centrifugó para separar la fase soluble de la fase insoluble a 15 000 rpm, por 20 minutos (Sorvall/SS-34).

## Purificación por columna de afinidad

El sobrenadante que se obtuvo de la centrifugación se filtró por membranas de 0.45  $\mu\text{m}$  y se inyectó a un equipo de cromatografía ÄKTA FPLC™ a través de una columna de níquel HisTrap™. La proteína recombinante fue posteriormente eluida con un gradiente de imidazol de 0 a 500 mM a un flujo de 4 mL/min. Las fracciones fueron colectadas en un volumen de 4 mL. Aquellas fracciones donde se observó absorbancia a 280 nm se juntaron en una única fracción y se verificó su pureza mediante un gel SDS-PAGE.

## Dicroísmo circular

Las proteínas puras se dializaron contra amortiguador de fosfatos  $\text{NaH}_2\text{PO}_4$  10 mM pH 8.0. Posteriormente, se realizó una dilución a una concentración de proteína de 0.4 mg/mL. Se utilizó el espectropolarímetro Chirascan™. Se tomó el espectro de dicroísmo circular en una celda de cuarzo de 0.1 cm de paso de luz, 0.5 nm de ancho de banda, de 260 a 185 nm. Se calculó la elipticidad molar a partir de la siguiente ecuación:

$$\text{Elipticidad molar } (\theta) \text{ (deg cm}^2 \text{ dmol}^{-1}) = \frac{\theta \text{ (en miligrados)} \times \text{peso promedio por residuo} \times 100}{[\text{Proteína (mg/mL)}] \times \text{paso de luz de la celda (cm)} \times 1000} \quad (4)$$

$$\text{peso promedio por residuo} = \frac{\text{peso molecular de la proteína (Da)}}{\# \text{ residuos de la proteína}}$$

Posteriormente, se realizó la deconvolución de los datos con el software CDNN de *Applied Photophysics* (Bohm *et al.*, 1992).

## Fluorescencia

Se utilizó la proteína en el mismo amortiguador de fosfatos utilizado en el experimento de DC. Se diluyeron las proteínas a una concentración de 10  $\mu\text{g/mL}$ . Se utilizó el espectrofluorómetro ISS PC1, en celdas de cuarzo de 0.5 x 0.5 cm. Las proteínas se excitaron a una longitud de onda de 295 nm y se midió la emisión de 310 a 400 nm.

## Calorimetría Diferencial de Barrido (DSC)

Los experimentos de calorimetría se realizaron con las proteínas ancestrales a una concentración de

1 mg/mL en amortiguador de fosfatos. Los experimentos se realizaron en el equipo VP-Capillary (MicroCal, GE Healthcare), en el laboratorio de Termofísica de la Facultad de Química. Se realizó una corrida con calentamiento de 10 a 100°C a una velocidad de 90 °C por hora, con un segundo calentamiento, para determinar la reversibilidad de las proteínas. Se corrió la línea base, con el amortiguador de fosfatos. Los datos se analizaron con el software Origin 7 y se ajustaron a un modelo de dos estados irreversible:

$$C_p^{APP} = \frac{\Delta H E_A}{RT_m^2} \exp(x) \times \exp[-\exp(x)] \quad (5)$$

$$x = \frac{E_A}{RT_m^2} (T - T_m)$$

Donde  $C_p^{APP}$  es la capacidad calorífica aparente,  $T_m$  es la temperatura media de desnaturalización,  $T$  es la temperatura de cada punto de la endoterma,  $\Delta H$  es la entalpía de desplegamiento,  $E_A$  es la energía de activación y  $R$  la constante de los gases ideales.

### Actividad enzimática

La proteína pura se dializó contra amortiguador de actividad (Trietanolamina 100 mM, EDTA 10 mM, DTT 10 mM pH 7.4). La actividad de isomerización de la enzima se siguió a través de una reacción acoplada con la enzima glicerol-3-fosfato deshidrogenasa ( $\alpha$ -GDH), la cual convierte la dihidroxiacetona fosfato y al NADH en glicerol-3-fosfato y  $\text{NAD}^+$ . La actividad de la TIM, por lo tanto, se siguió midiendo la desaparición de NADH con un espectrofotómetro Beckman DU 7500 a 340nm, en celdas de 1 cm de paso de luz a 25 °C. Se trabajó en un volumen de 700  $\mu\text{L}$  de amortiguador de actividad, utilizando NADH 0.2 mM, GDH 1.75  $\mu\text{g}/\text{mL}$ , las enzimas ancestrales a 100 ng/mL para ML y a 1 ng/mL para B, AM y AB y variando la cantidad de gliceraldehído-3-fosfato de 0.1 a 4 mM. Los datos de velocidades iniciales se ajustaron al modelo cinético de Michaelis-Menten y se obtuvieron los parámetros cinéticos de cada una de las enzimas.

### Estado oligomérico y estabilidad a la dilución

#### Cromatografía de exclusión molecular

Para determinar el estado oligomérico de las proteínas ancestrales nativas se realizó una cromatografía de exclusión molecular. Se utilizó la columna de exclusión Superdex 75 10/300 GL en el equipo ÄKTA FPLC™. Se utilizaron los estándares de pesos moleculares de GE Healthcare (Gel filtration calibration kit LMW), que se diluyeron en amortiguador Trietanolamina 10 mM, EDTA 1 mM, DTT 1 mM y NaCl 100 mM a pH 7.4. Se utilizó este mismo amortiguador para equilibrar la columna. Se aplicó un flujo de 0.75 mL/min a la columna. Las proteínas ancestrales se inyectaron a la columna a una concentración de 1mg/mL y a 10  $\mu\text{g}/\text{mL}$ .

#### Inactivación por dilución

Se realizaron diferentes diluciones de las proteínas ancestrales en un rango de 2 mg/mL a 7 ng/mL, las cuales se incubaron a 25 °C por 24 horas. Posterior al tiempo de incubación, se midió la actividad de las proteínas ancestrales. Se utilizaron las mismas concentraciones de  $\alpha$ -GDH, NADH y concentración de enzima final que se utilizaron en la medición de parámetros

catalíticos, pero se mantuvo constante la concentración de sustrato a 2 mM. Las actividades se normalizaron con un promedio de las actividades a concentraciones de proteína más altas. Los datos obtenidos se ajustaron a la ecuación de disociación (Romero-Romero, 2016)

$$f_N = \frac{K_{diss} + 4P_c \pm \sqrt{K_{diss}^2 + 8K_{diss}P_c}}{4P_c} \quad (6)$$

Donde  $f_N$  es la actividad catalítica normalizada,  $K_{diss}$  es la constante de disociación del dímero y  $P_c$  es la concentración de proteína.

Para los cálculos en los que se involucran el número de moles, los cálculos se realizaron tomando en cuenta el peso molecular del monómero.

## Resultados

### Resultados bioinformáticos

#### Selección de secuencias

Posterior a la depuración de secuencias, de acuerdo con los criterios señalados en la sección de metodología, el número final de secuencias fue de 239, con 81 secuencias de animales, 103 de hongos, 15 de plantas y 40 de protistas. Con respecto a la reconstrucción anterior, se aumentaron entre dos y seis veces el número de secuencias de cada grupo. Como grupos externos, se eligieron tres secuencias de TIM de bacterias, *Escherichia coli*, *Salmonella enterica* y *Klebsiella pneumoniae*.

#### Alineamiento Múltiple de secuencias

En el alineamiento de las 239 secuencias y los tres grupos externos se observa que, con la adición de in-del's, la longitud de la secuencia aumenta a 279 aminoácidos. Los sitios catalíticos (K14, H107 y E189 en el alineamiento; Figura 4) se encuentran conservados en todas las secuencias, así como los sitios aledaños a éstos. Adicionalmente, se observan sitios altamente conservados además de los aminoácidos catalíticos. Entre los aminoácidos 76 a 92 hay sitios altamente conservados pertenecientes al asa 3, que tiene una gran participación en la interacción entre monómeros. Del sitio 192 al 200 del alineamiento se encuentra el asa 6, que, tras un cambio conformacional, "atrapa" al sustrato impidiendo la entrada de agua al sitio activo. Finalmente, se observa alta conservación en los sitios que van del 232 al 236 y del 249 al 261, los extremos C-terminal de las hebras  $\beta 7$  y  $\beta 8$ , así como las aspas  $\beta\alpha 7$  y  $\beta\alpha 8$ . Estos sitios participan en la formación del sitio activo mediante interacciones de tipo puente de hidrógeno de la cadena principal con el fosfato del sustrato. También promueven la creación de un ambiente hidrofóbico que impide la entrada de agua al sustrato.

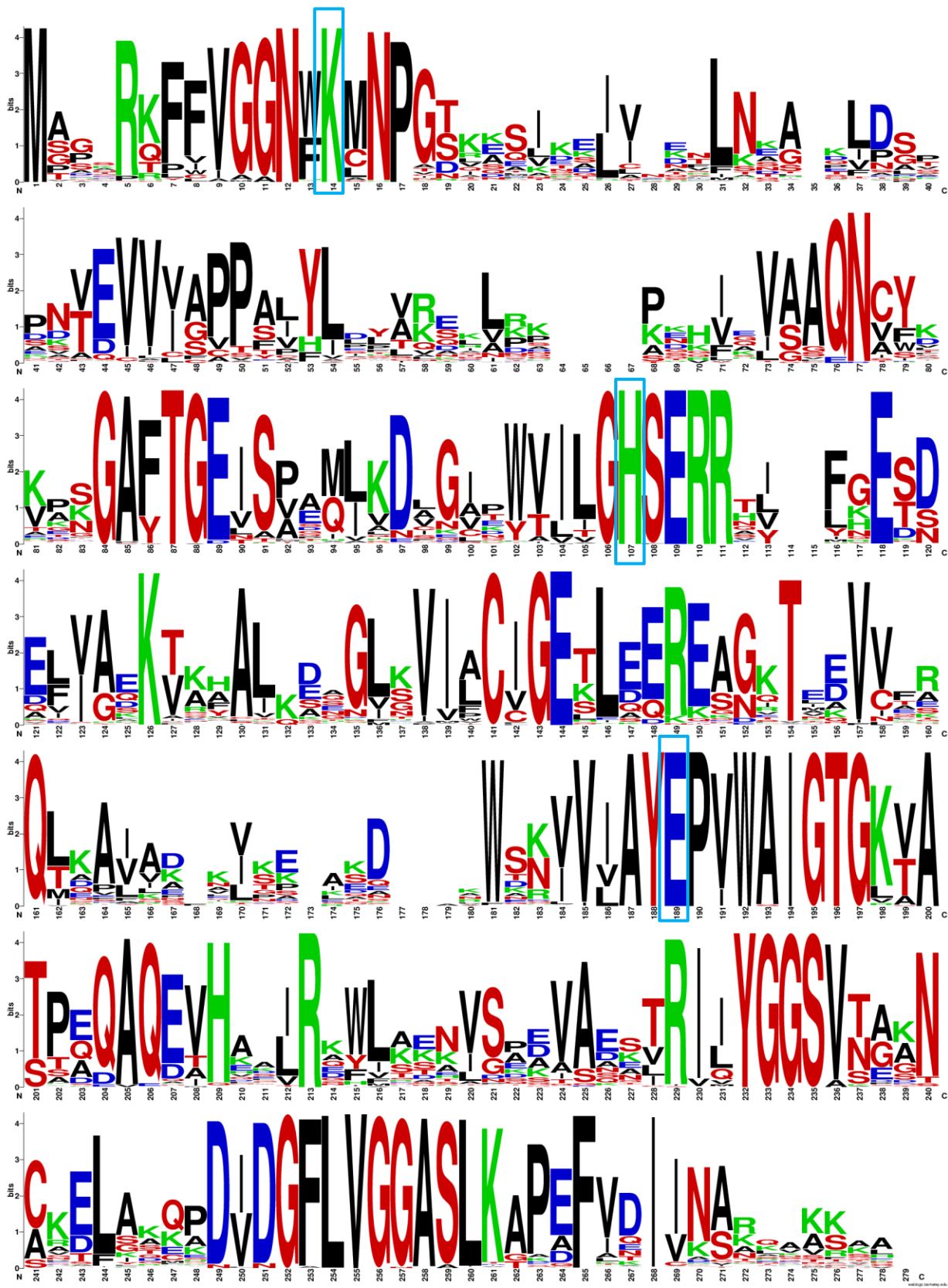


Figura 4.- Logo del alineamiento múltiple de secuencias para las 239 secuencias con los 3 grupos externos. En verde se encuentran los aminoácidos básicos, en azul los ácidos, en rojo los polares y en negro los no polares. Encerrados en un cuadro azul se encuentran los aminoácidos catalíticos <https://weblogo.berkeley.edu/logo.cgi>

## Selección del modelo de sustitución

El programa SMS arrojó los modelos que mejor se ajusten a los datos basado en la verosimilitud de éste y en el AIC. El mejor modelo seleccionado fue el modelo LG (Le & Gascuel, 2008) +G (0.847) +I (0.072) +F, escogido por los valores más altos de verosimilitud de -52887 y AIC de 106778. Esta matriz fue creada con el software XRATE a partir de 3,912 alineamientos con Pfam. Reúne aproximadamente 50,000 secuencias y aproximadamente 6.5 millones de residuos. Es una matriz empírica la cual describe la manera observada en que cambian los aminoácidos, en lugar de modelos teóricos.

**Tabla 4.-** Resultado del programa SMS de selección del mejor modelo de sustitución. El programa arroja los mejores modelos de sustitución probando diferentes combinaciones de modelos con decoraciones evaluados por el AIC.

| Model    | Decoración | K   | Llk    | AIC    | BIC    |
|----------|------------|-----|--------|--------|--------|
| LG       | +G+I+F     | 502 | -52887 | 106778 | 108601 |
| LG       | +G+F       | 501 | -52996 | 106995 | 108814 |
| RtREV    | +G+I+F     | 502 | -53074 | 107153 | 108976 |
| LG       | +G+I       | 483 | -53143 | 107253 | 109007 |
| WAG      | +G+I+F     | 502 | -53213 | 107431 | 109254 |
| LG       | +G         | 482 | -53246 | 107456 | 109206 |
| WAG      | +G+I       | 483 | -53288 | 107542 | 109296 |
| VT       | +G+I+F     | 502 | -53433 | 107871 | 109694 |
| Blosum62 | +G+I+F     | 502 | -53565 | 108134 | 109957 |
| VT       | +G+I       | 483 | -53633 | 108233 | 109987 |
| CpREV    | +G+I+F     | 502 | -53676 | 108356 | 110179 |
| DCMut    | +G+I+F     | 502 | -53853 | 108711 | 110534 |
| Dayhoff  | +G+I+F     | 502 | -53854 | 108712 | 110535 |
| JTT      | +G+I+F     | 502 | -53984 | 108972 | 110795 |
| MtZoa    | +G+I+F     | 502 | -54036 | 109076 | 110899 |
| MtREV    | +G+I+F     | 502 | -54345 | 109694 | 111517 |
| MtArt    | +G+I+F     | 502 | -54569 | 110142 | 111964 |
| Flu      | +G+I+F     | 502 | -55553 | 112111 | 113934 |
| HIVb     | +G+I+F     | 502 | -55875 | 112754 | 114577 |
| MtMam    | +G+I+F     | 502 | -56137 | 113279 | 115102 |
| AB       | +G+I+F     | 502 | -56222 | 113448 | 115271 |
| HIVw     | +G+I+F     | 502 | -58557 | 118119 | 119942 |

## Árbol filogenético

Se obtuvieron dos árboles filogenéticos, uno por máxima verosimilitud y otro por inferencia bayesiana. Hay que hacer varias menciones respecto a la creación de árboles filogenéticos. Primeramente, este no es el árbol que representa la filogenia de las especies, por lo que es normal que se encuentren secuencias de organismos que están lejanamente emparentados, pero

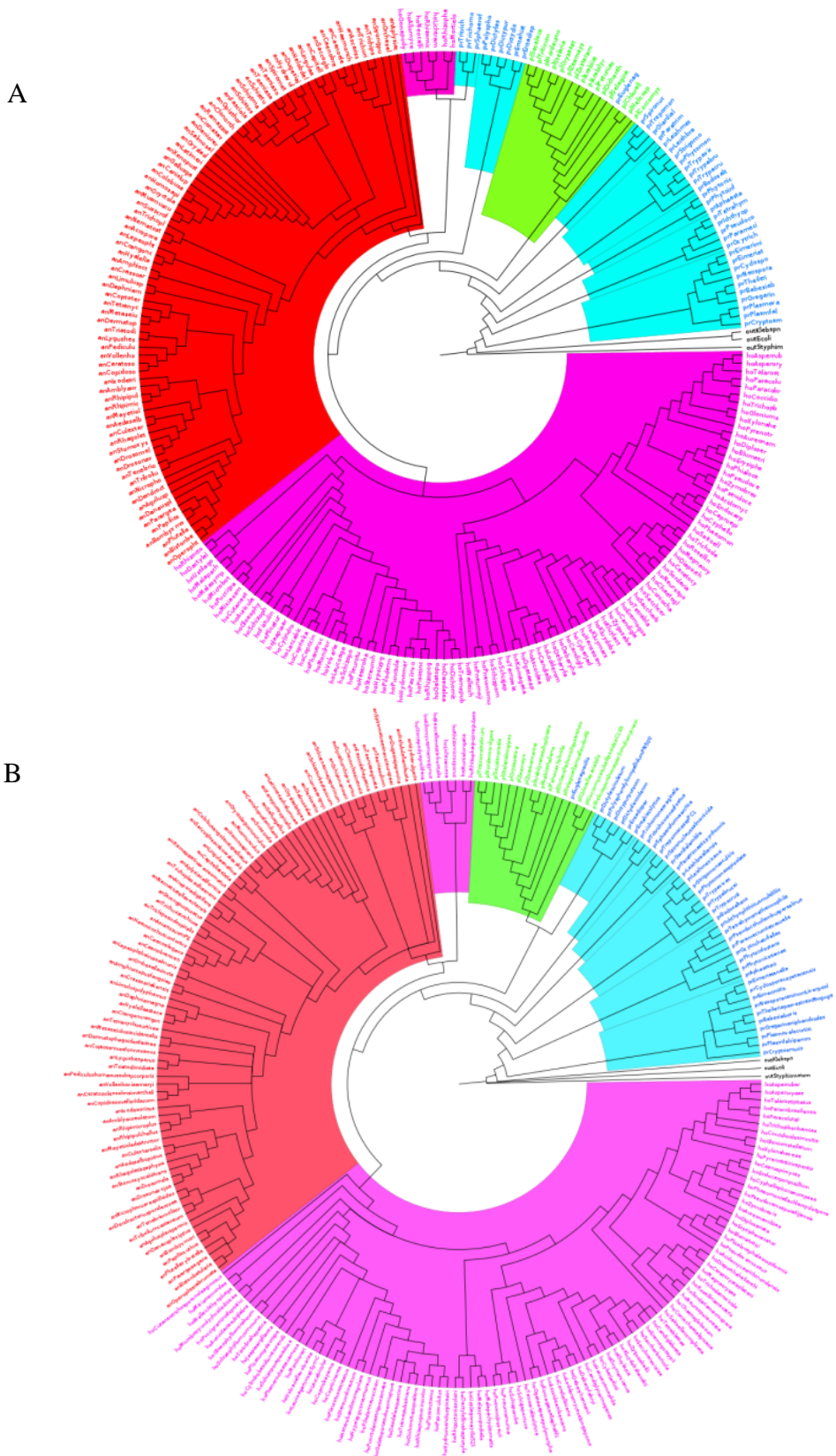
que comparten una secuencia muy parecida, ya sea por mero azar o por homoplasia. Segundo, este árbol no es “él” árbol filogenético que representa la relación de las secuencias de la proteína, es el mejor que se obtuvo con los parámetros y secuencias utilizados. Finalmente, el árbol que se muestra en la figura 5B correspondiente al árbol inferido por método bayesiano, es el resultado de un consenso de los árboles más probables. En este árbol, se observaron politomias en este árbol, que no se habían observado con el método de máxima verosimilitud ni con mayor número de secuencias, por lo que posiblemente estas politomias sean resultado del consenso del árbol.

Los árboles inferidos se encuentran muy bien resueltos a lo largo de las diferentes filogenias, salvo unos casos especiales. El nodo que se desea reconstruir, el último ancestro de los opistocontes, abarca al clado de los animales y al de los hongos. Podemos observar que se define bien la separación entre estos dos.

Observamos que un grupo de hongos se agrupa dentro del grupo de los animales en lugar de agruparse con el resto de los hongos. Los organismos de este grupo pertenecen al filo *Microsporidia*, microorganismos intracelulares que se encuentran dentro del reino fungi, pero muy diferentes a la forma de vida usual de los demás hongos.

También observamos a la especie de *Euglena gracilis* agrupada con las plantas. Ésta especie es considerada un alga, aunque filogenéticamente se encuentra más emparentado con *Trypanosoma* y *Leishmania* (Oneill *et al.*, 2015).

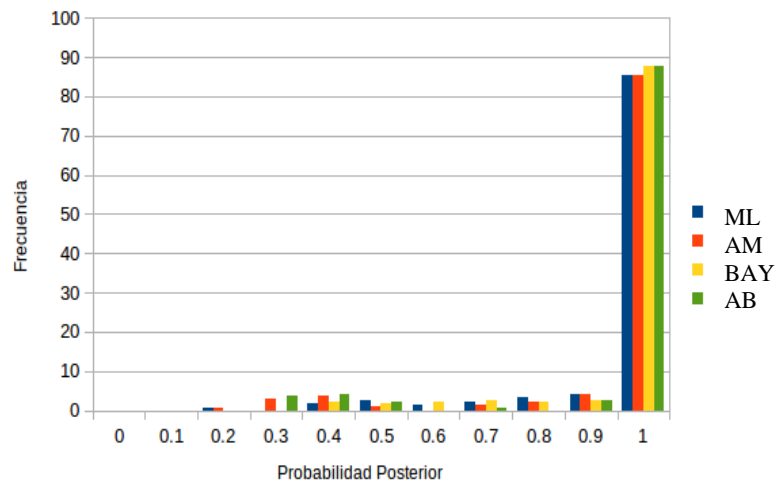




**Figura 5.-** Árboles filogenéticos representados en forma de cladograma obtenidos por el método de máxima verosimilitud (A) e inferencia bayesiana (B). En color se representan diferentes clados. Rojo, animales; rosa, hongos; verde, plantas y azul; protistas. En negro se observan los grupos externos.

## Reconstrucción Ancestral de Secuencias

Posterior al manejo de in-dels, se obtuvieron 4 secuencias de igual tamaño (250 aa.), ya que este proceso se realizó de igual manera para todas las reconstrucciones. Más del 80% de los sitios se reconstruyeron con una probabilidad posterior mayor a 0.9 (Figura 6). Como resultado, las probabilidades posteriores de las secuencias, que se obtiene promediando las probabilidades de todos los sitios, son altas (Tabla 3). Se realizó una búsqueda por BLAST de estas secuencias, todos los resultados obtenidos corresponden a TIMs. La secuencia con mayor puntaje e identidad es la secuencia putativa del hongo del filo *Zygomycota*, *Rhizopus microsporus*, tanto para la secuencia por máxima verosimilitud como por inferencia bayesiana con al menos un 70% de identidad. También, tiene alta puntuación e identidad con el hongo *Rhizopus delemar*; el hongo *Acomycota*, *Coniochaeta ligniaria*; el hongo *Zygomycota*, *Mortierella elongata* y el molusco *Conus litteratus*. Con respecto a la secuencia que se obtuvo en la reconstrucción anterior (TIM63), se obtuvo más de un 80% de identidad. Cabe resaltar que es la secuencia con la que mayor porcentaje de identidad con las cuatro proteínas ancestrales de éste trabajo, sin embargo, se observa que el manejo de in-dels se realizó de manera diferente.



**Figura 6.-** Probabilidad posterior de los sitios de las secuencias reconstruidas. Se graficó la frecuencia de sitios que poseen su respectiva probabilidad posterior. Más del 80% de sitios tienen una probabilidad posterior entre 0.9 y 1.

**Tabla 5 .-** Estadísticos de las secuencias reconstruidas por los métodos de máxima verosimilitud y bayesiano, además de las peores secuencias plausibles. El error esperado se refiere al número de residuos que probablemente son incorrectos debido a la baja probabilidad de reconstrucción de los sitios. Ésta se obtiene al sumar la diferencia entre la unidad y la probabilidad de cada sitio. Los cambios con respecto a la secuencia original se refieren a los sitios ambiguos y que se cambiaron con el segundo mejor aminoácido.

|   | <b>Probabilidad posterior</b> | <b>Error esperado</b> | <b>Cambios con respecto a la secuencia original</b> |
|---|-------------------------------|-----------------------|---|
| <b>Máxima verosimilitud</b>                     | 0.9418                        | 14.53                 | -   |
| <b>Pero Caso Plausible Máxima verosimilitud</b> | 0.9272                        | 18.18                 | 18 cambios  |
| <b>Inferencia Bayesiana</b>                     | 0.9469                        | 13.26                 | -   |
| <b>Peor Caso Plausible Bayesiano</b>            | 0.9247                        | 18.81                 | 22 cambios  |

```

ScTIM      MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPENVEVVICPPATYLDYSVSLVKKP-QV 59
RmTIM      MARKFFVGGNWKMNQSVAVKELVELLNKIDVPANTEVVVAPPALYVDRVNQAVK--KEI 58
TIM63      MSRKFFVGGNWKMNGLTESIKALVETLNSAQLDPNTEVVVAPPAYLPPFARSKLKKPKKEI 60
ML         MARKFFVGGNWKMNGLTESIKSLVETLNNAELDDNTEVVVAPPAYLPLVRQSLR--KDI 58
AM         MARKFFVGGNWKMNQSVESIKSLVETLNNAELDPNVEVVIAPPAYLPLVREALR--KDI 58
BAY        MARKFFVGGNWKMNQSKASIKSLVETLNASKLDNKTEVVVAPPAYLPPVQRQALR--KDI 58
AB         MARKFFVGGNWKMNQTESIKSLVETLNANNLDNKVEVVIAPPAYLPPVQRQQLR--KDI 58
          *:.*.*****:.*:**: .:* :** ** .: .:***:.*** *: . : :

ScTIM      TVGAQNAYLKASGAFTGENSVQIKDVGAKWVILGHSERRSYFHEDDKFIADKTKFALGQ 119
RmTIM      QVAAQNNTYTKASGAYTGEISPQMLKDLGNWVILGHSERREYFKESDEFVGGKVRFALDA 118
TIM63      QVAAQNNTYTKASGAYTGEISAEMLKDLGVPWVILGHSERRTIFGESDELIAEKVKYALDQ 120
ML         QVAAQNNTYTKASGAYTGEISAEMLKDLGIPWVILGHSERRQIFGESDEFVAEKTKYALDN 118
AM         QVAAQNNTYTKASGAYTGEISAEMLKDLGIPWVILGHSERRHIFGESDELVAEKTKYALDS 118
BAY        QVAAQNNTYTKASGAYTGEISPEMLKDLGIPWVILGHSERRHIFGESDEFVAEKTKYALDQ 118
AB         QVAAQNNTYTKASGAYTGEISPEMLKDLGIPWVILGHSERRNIFGESDELVAEKTVHALEA 118
          *.*** * * .**:* ** * : :**:* ***** * *.*:::.*:.*

ScTIM      GVGVILCIGETLEEKKAGKTLDVVERQLN-AVLEEV--KDWTNVVVAEYEPVWAI GTGLAA 176
RmTIM      GVSVIACIGETLEEREANVTNDVVARQM-KAIANEI--NDWTNVVVAEYEPVWAI GTGKVA 175
TIM63      GLKVIACIGETLEEREAGKTMEVVARQILKAIADKI--KDWSNVVIAEYEPVWAI GTGKVA 178
ML         GLKVILCIGETLEEREAGKTMDVVT RQL-KAIADKISESDWSKVVIAYEPVWAI GTGKVA 177
AM         GLSVILCIGETLEEREAGKTMDVVARQL-KAIADKITENDWSKVVIAYEPVWAI GTGKVA 177
BAY        GLKVIACIGETLEEREAGKTMEVCARQL-KAIAKKISESDWSNVVIAEYEPVWAI GTGKVA 177
AB         GLKVIACIGETLEEREAGKTTEVVARQL-KAIAKKISESDWSKVVIAYEPVWAI GTGKVA 177
          *: ** ****.***:*.* * : * ** : * : .: .**:*.*:***** *

ScTIM      TPEDAQDIHASIRKFLASKLGDKAASELRILYGGSSANGSNAVTFKDKADVDFLVGGASL 236
RmTIM      TPEQAQEVHGFRLRQWLAENVSQEAEEKTRILYGGSSVNGGNC KTLASKPDI DGFVGGASL 235
TIM63      TPEQAQEVHAAIRKWLKENVSPEVAESTRIIYGGSSVNAANCKELAKQPDI DGFVGGASL 238
ML         TPEQAQEVHAEIRKWLKENVSPEVAESTRIIYGGSSVNGGNC KTLAKQPDI DGFVGGASL 237
AM         TPEQAQEVHAEIRKWLKENVSPEVAESTRIIYGGSSVNGGNC KTLAKQPDI DGFVGGASL 237
BAY        TPEQAQEVHAAIRKWLKENVSPEVAESTRIIYGGSSVNGGNC AELAKQPDI DGFVGGASL 237
AB         TPEQAQEVHAAIRKWLKENVSPEVAESTRIIYGGSSVNGSNCAELAKQPDI DGFVGGASL 237
          ***:***:*. :*** * .: .: *.. **:* **.*..* . : : *:*:*****

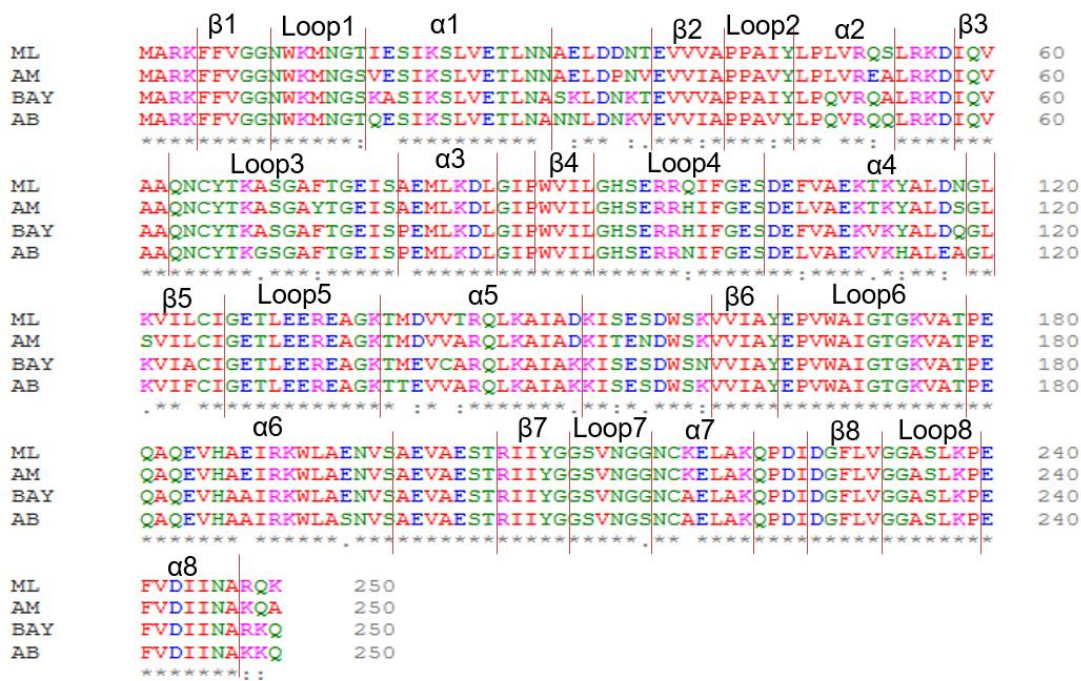
```

|          |        |        |        |        |        |        |        |
|----------|--------|--------|--------|--------|--------|--------|--------|
| 1: ScTIM | 100.00 | 60.41  | 58.87  | 60.82  | 60.82  | 60.00  | 59.18  |
| 2: RmTIM | 60.41  | 100.00 | 71.66  | 72.58  | 73.79  | 74.19  | 70.16  |
| 3: TIM63 | 58.87  | 71.66  | 100.00 | 87.04  | 85.43  | 86.23  | 83.40  |
| 4: ML    | 60.82  | 72.58  | 87.04  | 100.00 | 92.80  | 90.40  | 87.60  |
| 5: AM    | 60.82  | 73.79  | 85.43  | 92.80  | 100.00 | 88.00  | 87.60  |
| 6: BAY   | 60.00  | 74.19  | 86.23  | 90.40  | 88.00  | 100.00 | 91.20  |
| 7: AB    | 59.18  | 70.16  | 83.40  | 87.60  | 87.60  | 91.20  | 100.00 |

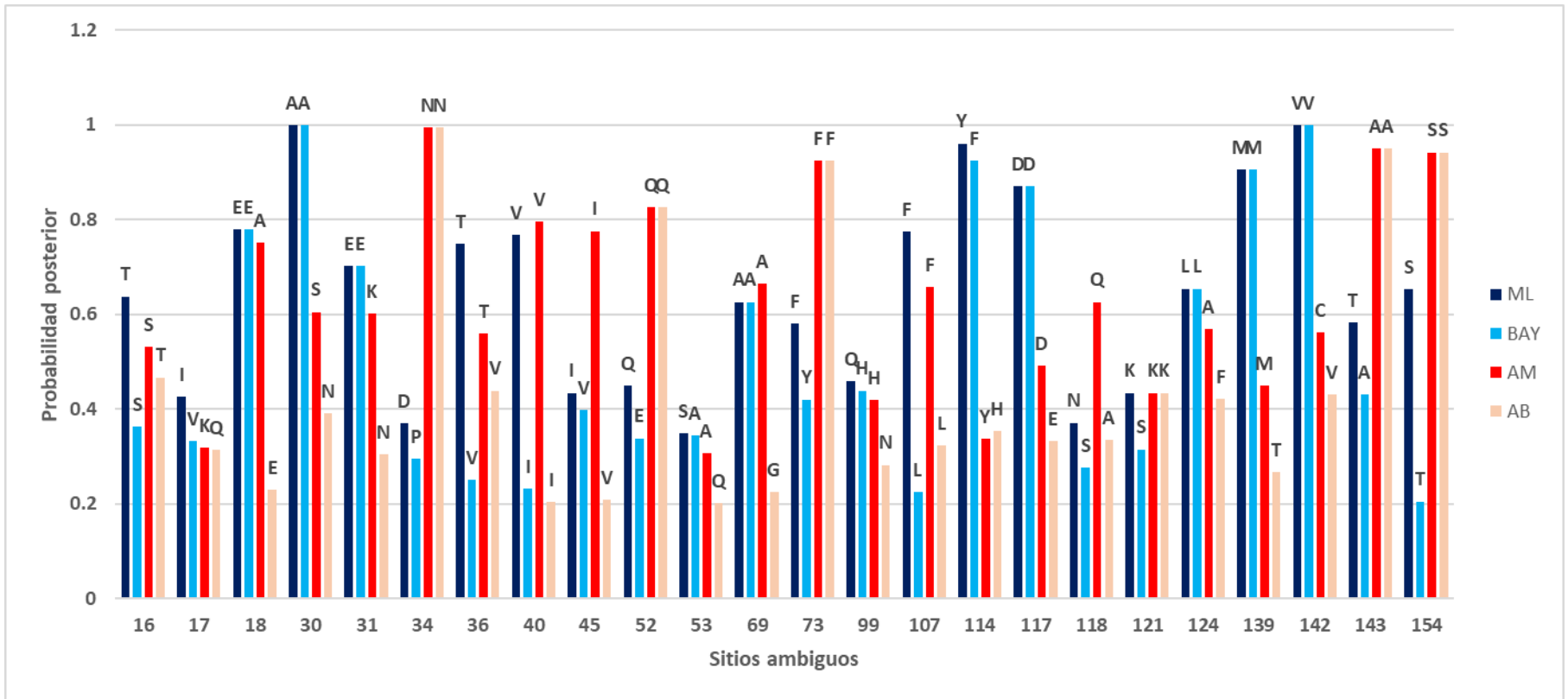
**Figura 7.-** Alineamiento de secuencias con los porcentajes de identidad de las cuatro secuencias obtenidas en esta reconstrucción con la secuencia de la TIM de *Saccharomyces cerevisiae* (ScTIM), la TIM de *Rhizopus microsporus* (RmTIM, la secuencia de mayor identidad al realizar el BLAST) y la secuencia del nodo del último ancestro de los opistocontes obtenida de la anterior reconstrucción (TIM63). Se utilizó el programa Clustal Omega para este alineamiento (Sievers, 2014); Código de colores: Rojo, pequeño hidrofóbico; Azul, ácido; Magenta, básico; Verde, hidroxilo, sulfhidrido, amina y glicina.

Entre las cuatro proteínas se observa alta identidad de secuencia. El menor porcentaje de identidad corresponde a la que hay entre la obtenida por máxima verosimilitud y la proteína obtenida por la estrategia de peor caso plausible por inferencia bayesiana con 87.60 %. Entre las 4 proteínas se observan 15 cambios no conservados. De los sitios ambiguamente reconstruidos se comparten 10 sitios entre las proteínas obtenidas por máxima verosimilitud e inferencia bayesiana, 8 aminoácidos más se encuentran exclusivamente en la peor secuencia plausible de máxima verosimilitud y 12 exclusivamente en la peor secuencia plausible de inferencia bayesiana.

Entre la proteína obtenida por máxima verosimilitud y su peor caso plausible hay 18 cambios. De estos cambios 10 son no equivalentes en cuanto a las propiedades fisicoquímica de los residuos. Por cambio de propiedades fisicoquímicas, entiéndase aquellos donde se cambian un aminoácido polar por uno no polar, por uno con carga (positiva o negativa) o aromático, y las combinaciones entre éstos. Entre la proteína obtenida por inferencia bayesiana y su peor caso plausible hay 22 cambios de los cuales 15 son cambios no equivalentes en cuanto a la fisicoquímica del aminoácido.



**Figura 8.-** Alineamiento de secuencias de las cuatro secuencias obtenidas de la reconstrucción, mostrando las estructuras secundarias del barril TIM. Código de colores: Rojo, pequeño hidrofóbico; Azul, ácido; Magenta, básico; Verde, hidroxilo, sulfhidrilo, amina y glicina. Alineamiento obtenido por Clustal Omega (Sievers, 2014)

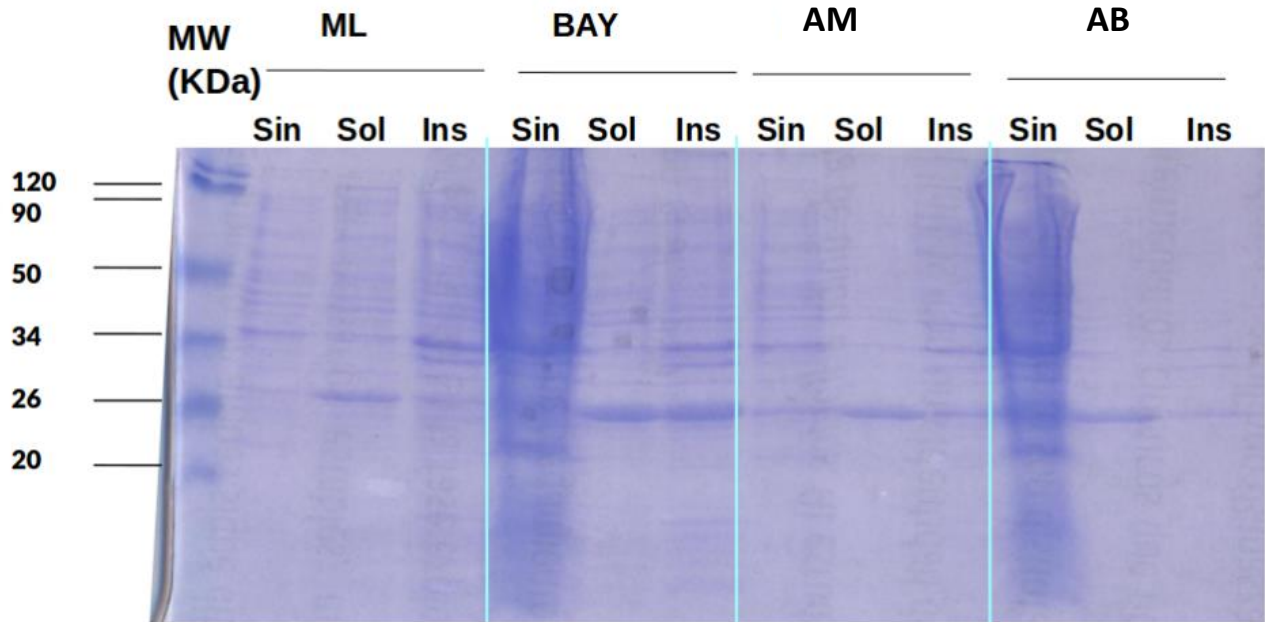


**Figura 9.-** Aminoácidos ambiguamente reconstruidos de las secuencias ancestrales. Se grafica la probabilidad posterior contra cada uno de los sitios ambiguos para las secuencias. Sobre la barra se muestra la identidad del aminoácido que se colocó en la secuencia correspondiente. El detalle de las probabilidades posteriores de cada uno de los aminoácidos se encuentra en el anexo 7.

## Resultados experimentales

### Sobrexpresión de las proteínas ancestrales

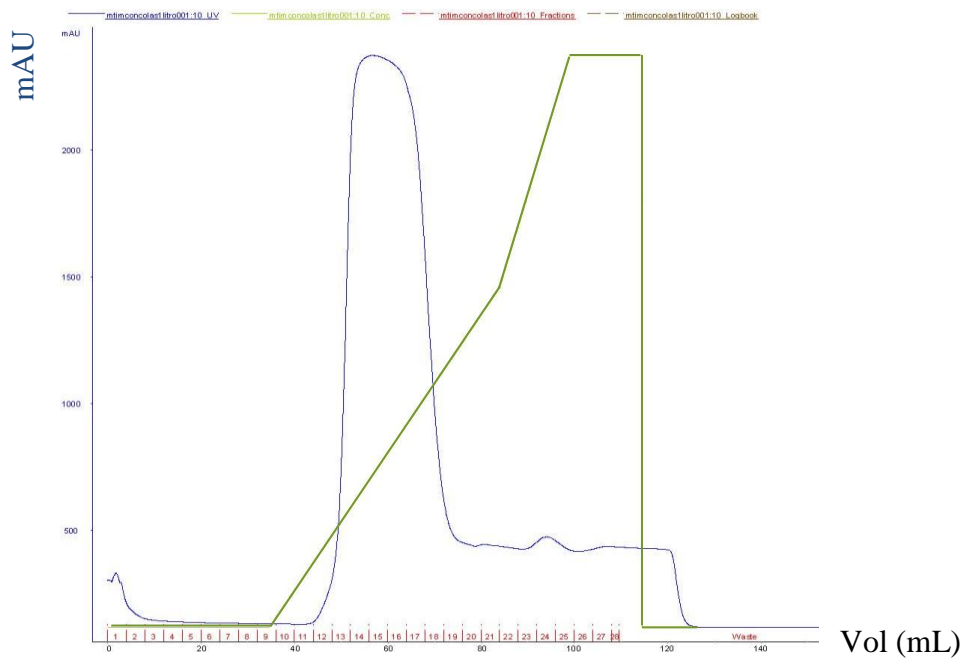
Las proteínas ancestrales tienen una masa molecular de 29 kDa (Anexo 5), se puede observar en la figura 10 que la banda correspondiente aumenta en intensidad posterior a la inducción. Las proteínas ancestrales recombinantes se expresan a través de la inducción por IPTG en la fase soluble, por lo que no fue necesario un procedimiento extra para recuperarla de la fase insoluble.



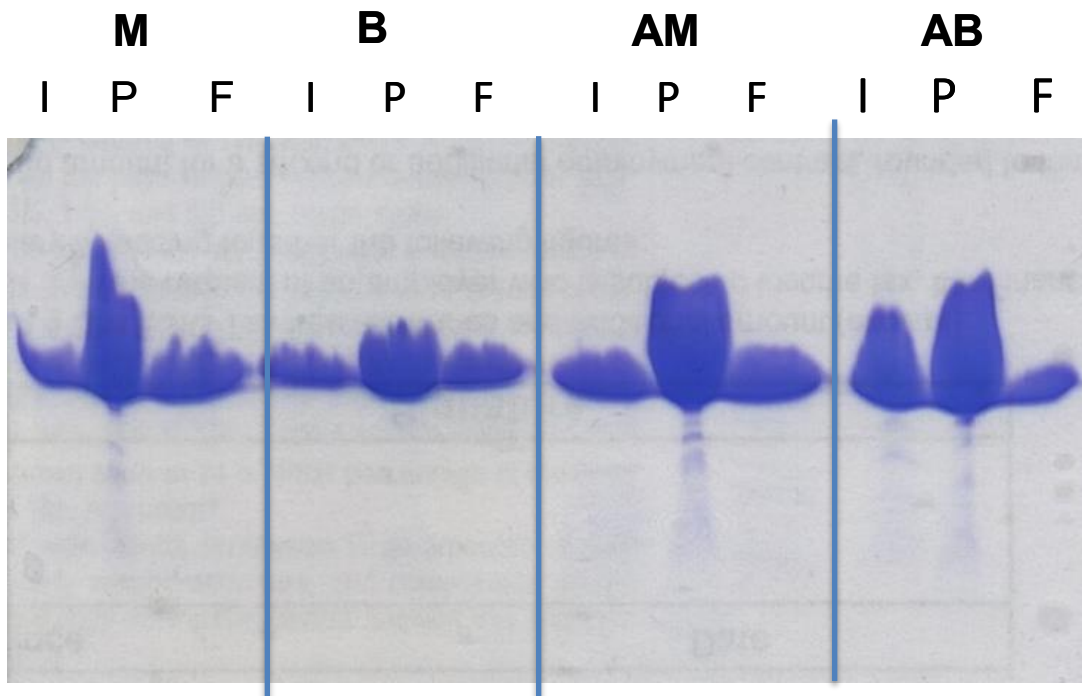
**Figura 10.-** Gel de poliacrilamida desnaturalizante al 12.5% de la inducción de las proteínas ancestrales. Se muestra la sobre expresión de cada una de las proteínas ancestrales, cuya masa molecular es de aproximadamente 29 kDa. Para cada proteína se muestra la fracción sin inductor, la fracción soluble y la fracción insoluble. En el primer carril se muestran un estándar de pesos moleculares (Thermo Scientific Prestained Protein Molecular Weight Marker).

### Purificación de las proteínas ancestrales

Las proteínas ancestrales se purificaron con ayuda de la etiqueta de histidinas mediante una columna de afinidad con níquel. Se observa un pico de alta absorbancia a 280 nm en la figura 11 al eluir con gradiente de imidazol a 60 mL, aproximadamente. En el gel de poliacrilamida se observa una única banda correspondiente a las proteínas ancestrales (Figura 11).



**Figura 11.-** Cromatograma de la purificación de las proteínas ancestrales. En verde se muestra el aumento de la concentración del amortiguador B que contiene una concentración de 500 mM de imidazol, en azul la absorbancia a 280 nm y en rojo las fracciones colectadas.



**Figura 12.-** Gel de poliacrilamida desnaturalizante al 12.5% de la purificación de las proteínas ancestrales. Se muestra la purificación de cada una de las proteínas ancestrales, cuya masa molecular es de aproximadamente 29 kDa. Para cada proteína se muestra la fracción inicial (I), la fracción principal (P) y la fracción final (F) del pico observado en el cromatograma.

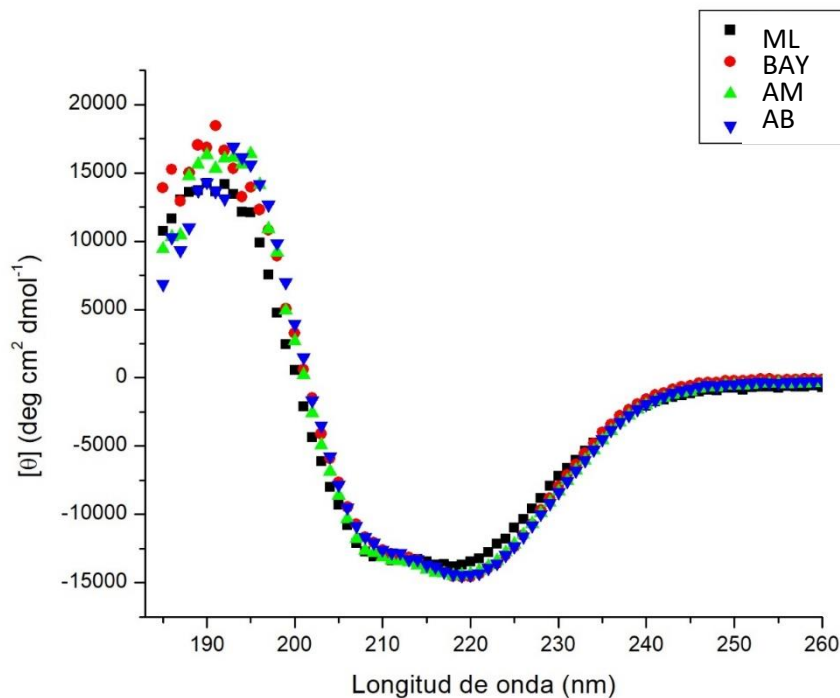


## Propiedades espectroscópicas

### Dicroísmo circular

Las cuatro proteínas presentan un espectro muy similar de dicroísmo circular, con los mismos mínimos a 208 y 220 nm correspondientes a la estructura secundaria de hebra beta y hélice alfa, respectivamente.

Posteriormente, los datos se deconvolucionaron para obtener el porcentaje de estructura secundaria. Tanto el espectro de DC (Figura 13) como el porcentaje de estructura secundaria obtenido mediante la deconvolución de los espectros corresponden a lo esperado para un barril TIM.



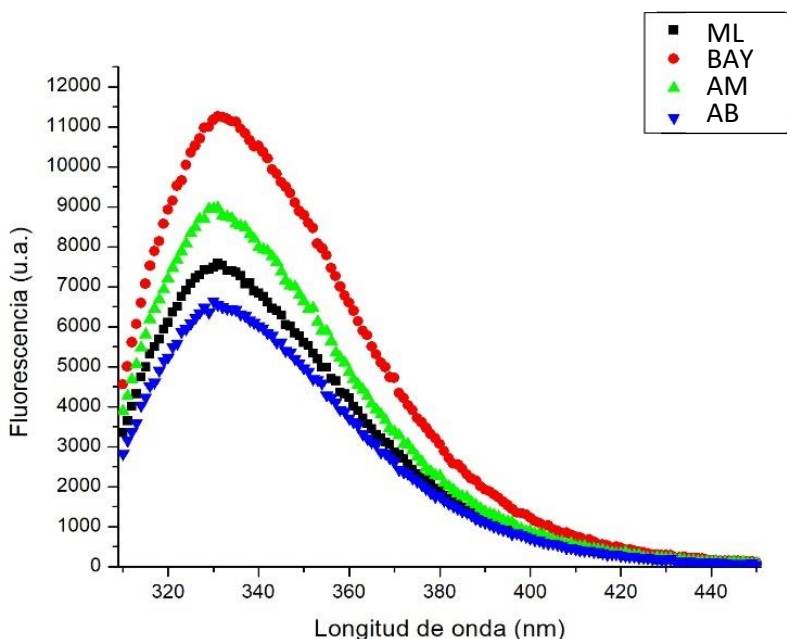
**Figura 13.-** Espectro de dicroísmo circular de las 4 proteínas ancestrales. Los datos se graficaron en unidades de elipticidad molar.

**Tabla 6.-** Deconvolución a partir del espectro de DC con las proporciones de estructura secundaria para cada una de las proteínas ancestrales. Los valores de ScTIM se obtienen de la deconvolución del espectro de DC (Schulte-Sasse, 2015)

| Proteína \ % | Hélice | Hebra | Giro beta | Random coil |
|--------------|--------|-------|-----------|-------------|
| ML           | 34.5   | 16.7  | 17.1      | 27.7        |
| BAY          | 36.2   | 14.5  | 16.4      | 29.2        |
| AM           | 36.9   | 14.2  | 16.4      | 28.2        |
| AB           | 36     | 14.6  | 16.4      | 29.7        |
| ScTIM        | 38     | 13    | 16        | 30          |

## Fluorescencia

El espectro de fluorescencia de las 4 proteínas presenta la misma forma y un máximo de emisión muy similar (Figura 14) con solo una diferencia máxima de 2 nm. La longitud de onda a la que se encuentra este máximo de emisión (332 nm) también indica que los aminoácidos aromáticos se encuentran escondidos al interior de la proteína (Tabla 6), por lo que podemos suponer que las proteínas se encuentran plegadas. Se observa una diferencia en las intensidades de fluorescencia para las 4 proteínas (Figuras 14) por lo que puede estar ocurriendo un fenómeno de apagamiento entre los aminoácidos vecinos (Lakowicz, 2006).



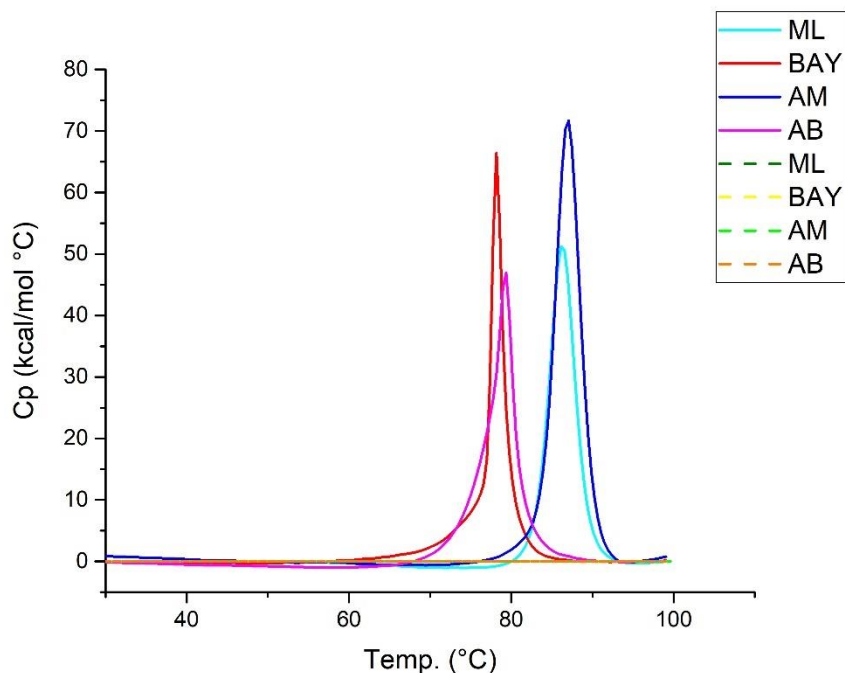
**Figura 14.-** Espectro de fluorescencia de las 4 proteínas ancestrales al excitarse con una longitud de onda de 295 nm.

**Tabla 7.-** Número de residuos aromáticos de las proteínas ancestrales y los resultados obtenidos a partir de la fluorescencia; la longitud de onda de emisión máxima y el centro espectral de masa.

|            | Aminoácidos aromáticos |   |   | $\lambda_{\max}$ (nm) | Centro espectral de masa |
|------------|------------------------|---|---|-----------------------|--------------------------|
|            | W                      | Y | F |                       |                          |
| <b>ML</b>  | 5                      | 5 | 7 | 332                   | 345.8                    |
| <b>BAY</b> | 5                      | 5 | 7 | 332                   | 346.7                    |
| <b>AM</b>  | 5                      | 6 | 5 | 330                   | 345.7                    |
| <b>AB</b>  | 5                      | 4 | 7 | 331                   | 346.5                    |

### Calorimetría diferencia de barrido

Realizando el barrido a 90 °C/hora, se observa una sola transición para cada una de las proteínas. Se observa que la transición de las proteínas ancestrales obtenidas por inferencia bayesiana se da a una temperatura menor que las proteínas inferidas por máxima verosimilitud (Figura 15). Así mismo, las proteínas obtenidas por inferencia bayesiana presentan menor  $\Delta H$  que aquellas obtenidas por máxima verosimilitud, las cuales poseen el doble de  $\Delta H$  (Tabla 7).



**Figura 15.-** Calorimetría diferencial de barrido de las cuatro proteínas ancestrales. Las líneas continuas muestran la endoterma obtenida al calentar por primera vez la muestra mientras que las líneas discontinuas muestran el segundo calentamiento de las proteínas tras dejar enfriar la muestra.

**Tabla 8.-** Temperaturas medias de desnaturalización y cambios en la entalpía de desnaturalización de las proteínas ancestrales. Los errores de la  $T_m$  es una desviación estándar de tres experimentos independientes.

|     | $T_m$ (°C)     | $\Delta H$ (kJ/mol) |
|-----|----------------|---------------------|
| M   | $85.3 \pm 1.1$ | 938.7               |
| BAY | $77.2 \pm 1.1$ | 552.9               |
| AM  | $86.4 \pm 1.1$ | 1216.0              |
| AB  | $77.2 \pm 1.7$ | 564.35              |

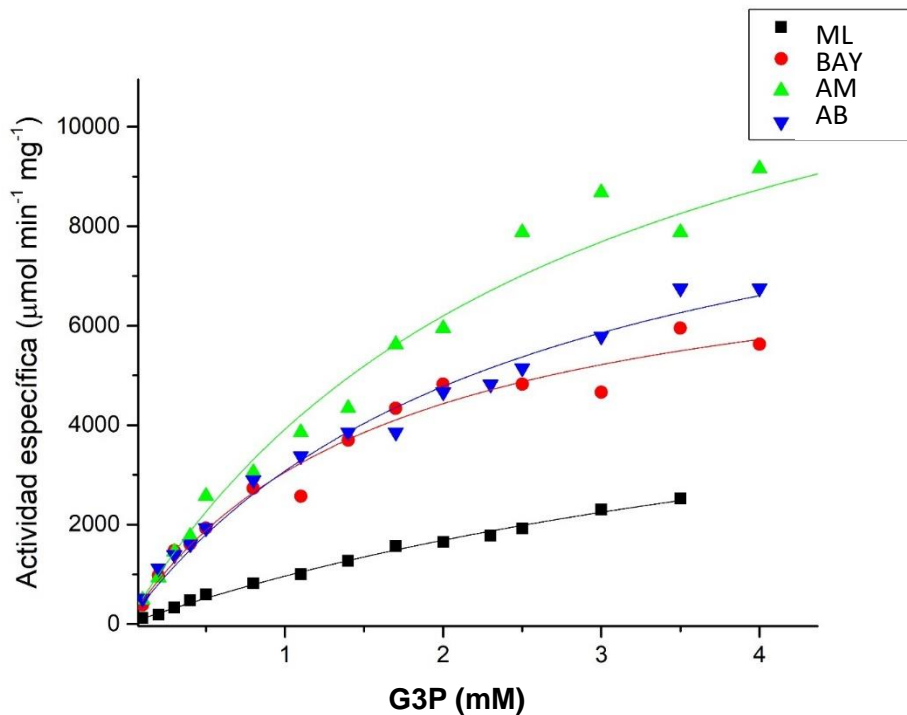
Con el fin de determinar la reversibilidad del desplegamiento, todas las muestras se sometieron a un segundo ciclo de calentamiento, en el cual no se observó una endoterma significativa, lo cual indica que al enfriar la muestra después del primer ciclo, la proteína no es capaz de

replegarse. Por lo tanto, el proceso es irreversible. Se ajustaron los datos a un modelo de dos estados irreversible para su análisis.

Las temperaturas medias de desnaturalización resultaron ser altas, en promedio 17 grados más alta que la TIM63 y la TIM del mesófilo *Saccharomyces cerevisiae* (65.8°C, Peimbert *et al.*, 2008) (Tabla 7). Sin embargo, no son tan altas como la Tm de la TIM de la bacteria termófila *Thermotoga maritima*, que tiene una Tm de 102 °C (Alvarez *et al.*, 1999).

### Actividad enzimática

Las proteínas ancestrales son capaces de realizar la actividad catalítica de isomerización propia de la TIM. En los cuatro casos, fue posible ajustar los datos al modelo de Michaelis-Menten. Aunque se observan variaciones ligeras en cuanto a la velocidad máxima, los parámetros catalíticos de constante catalítica y constante de afinidad, la eficiencia catalítica de las cuatro enzimas se encuentra en el mismo orden de magnitud (Tabla 9).



**Figura 16.-** Actividad enzimática de las cuatro enzimas ancestrales medida contra la concentración de proteína. La línea continua es el ajuste al modelo de Michaelis-Menten.

**Tabla 9.-** Constantes catalíticas de las proteínas ancestrales. Se realizó la corrección sobre la eficiencia catalítica tomando en cuenta que solo 4% del sustrato se encuentra disponible para su catálisis.

|            | $k_{cat}$ ( $\times 10^3 \text{ s}^{-1}$ ) | $K_m$ (mM)      | $k_{cat}/K_m$ ( $\text{M}^{-1} \text{ s}^{-1}$ ) |
|------------|--|-----------------|--|
| <b>ML</b>  | $3.32 \pm 0.34$                            | $5.99 \pm 0.86$ | $1.39 \pm 0.34 \times 10^7$                      |
| <b>BAY</b> | $3.95 \pm 0.31$                            | $1.64 \pm 0.29$ | $6.01 \pm 1.53 \times 10^7$                      |
| <b>AM</b>  | $7.25 \pm 0.71$                            | $2.79 \pm 0.53$ | $6.51 \pm 1.87 \times 10^7$                      |
| <b>AB</b>  | $5.23 \pm 0.41$                            | $2.47 \pm 0.37$ | $5.29 \pm 1.20 \times 10^7$                      |

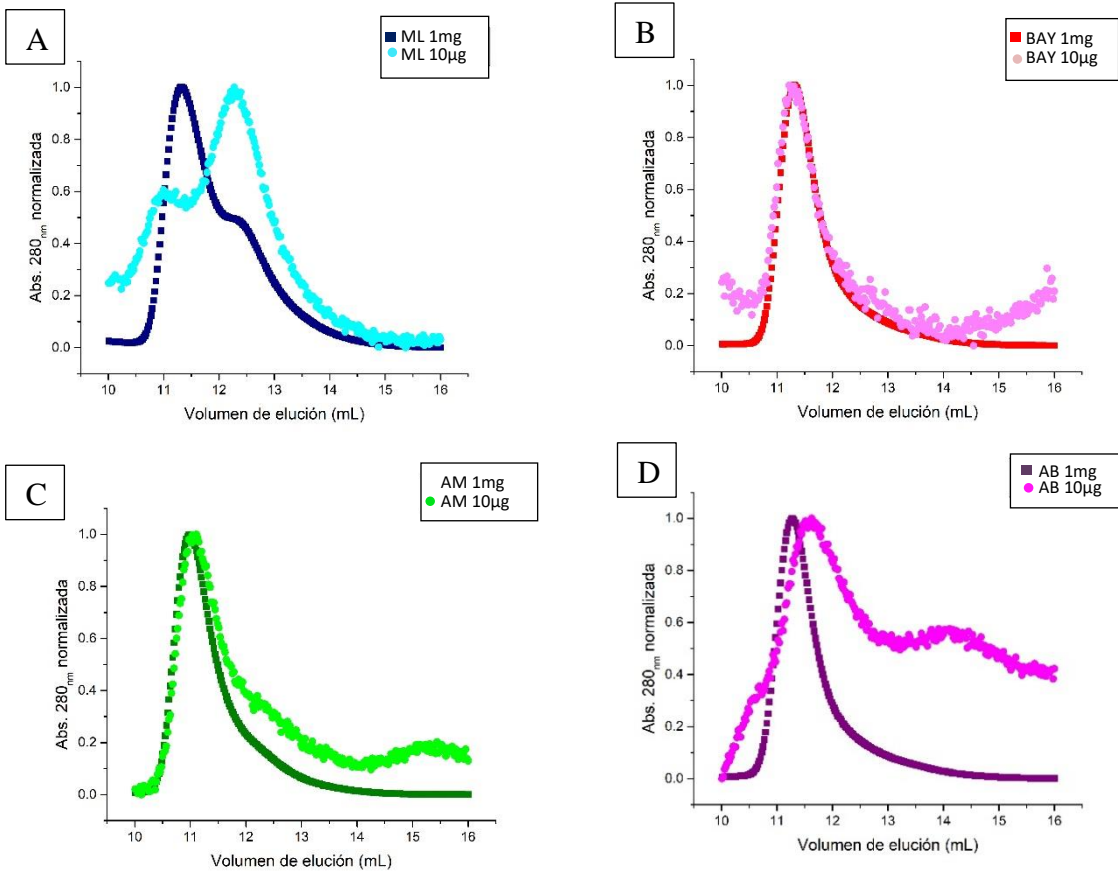
### Radio hidrodinámico

A partir del análisis de las proteínas en la columna de exclusión molecular analítica, se logró determinar el radio hidrodinámico de las cuatro proteínas. Las proteínas poseen un radio hidrodinámico comparable al de una TIM nativa dimérica ( $29.6 \pm 0.25 \text{ \AA}$ ; Morgan *et al.*, 2000).

**Tabla 10.-** Radios hidrodinámicos de las cuatro proteínas ancestrales a dos concentraciones diferentes, 1 mg/mL y 10  $\mu\text{g/mL}$ .

|            | Radio hidrodinámico ( $\text{\AA}$ ) |                     |
|------------|--------------------------------------|---------------------|
|            | 1 mg/mL                              | 10 $\mu\text{g/mL}$ |
| <b>ML</b>  | 31.6                                 | 26.8                |
| <b>BAY</b> | 31.7                                 | 31.9                |
| <b>AM</b>  | 33.5                                 | 33.1                |
| <b>AB</b>  | 32.9                                 | 30.2                |

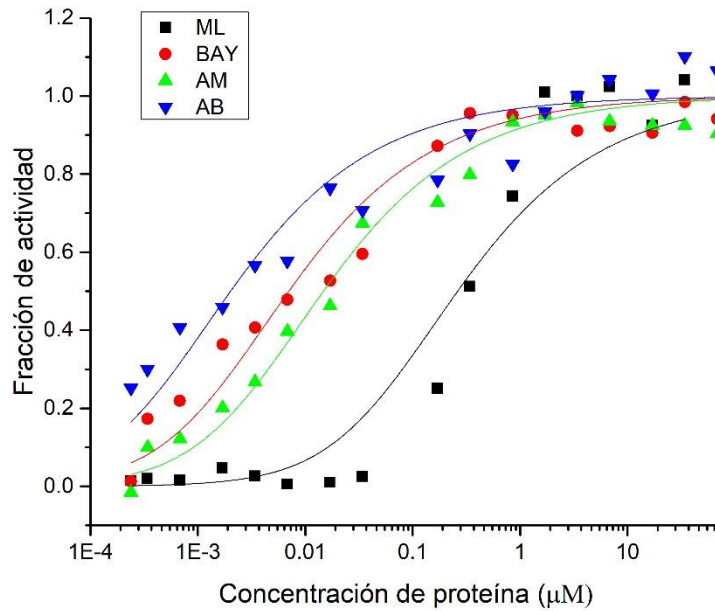
A altas concentraciones de proteína (1 mg/mL) se observa un solo pico en el perfil de elución de las proteínas (Figura 17), excepto para la obtenida por máxima verisimilitud donde se empieza a observar la presencia de un segundo pico (Figura 17A) que se hace más presente a bajas concentraciones (10  $\mu\text{g/mL}$ ).



**Figura 17.-** Perfiles de elución de las cuatro proteínas ancestrales con la absorbancia normalizada sobre el pico de mayor absorbancia. A, ML; B, BAY; C, AM; D, AB. Los colores oscuros representan concentraciones mayores mientras que los colores claros concentraciones bajas de proteínas cargadas a la columna.

### Estabilidad a la dilución

Al medir la actividad enzimática, incubando previamente las enzimas a diferentes concentraciones de proteína, se observa que las cuatro proteínas ancestrales sufren de inestabilidad a la dilución. Se observa también que existen grandes diferencias entre ellas, siendo este parámetro uno de los más variables entre éstas (Tabla 11). También se observa que el ajuste de los puntos con el modelo no es del todo adecuado como se puede observar en las TIMs bacterianas (Romero-Romero *et al.*, 2015). Por esta razón también se determinó la concentración a la que se obtiene la mitad de la actividad ( $K_{0.5}$ ). Se observan que los datos de  $K_{0.5}$  son ligeramente mayores a la  $K_{diss}$  (Tabla 11).



**Figura 18.-** Estabilidad a la dilución de las proteínas ancestrales. Se normalizó la actividad tomando como referencia el promedio de los valores de actividades obtenidos a las concentraciones más altas de proteína. La línea continua es el ajuste a la ecuación de disociación (Ecuación 6).

**Tabla 11.-** Constantes de disociación y  $K_{0.5}$  de las cuatro proteínas ancestrales obtenido del ajuste de los datos al modelo de disociación. Los errores asociados corresponden a los errores al ajustar el modelo.

|            | $K_{diss}^{APP}$ ( $\mu M$ ) | $K_{0.5}$ ( $\mu M$ ) |
|------------|------------------------------|-----------------------|
| <b>ML</b>  | $0.2610 \pm 0.0873$          | 0.2976                |
| <b>BAY</b> | $0.0070 \pm 0.0013$          | 0.0086                |
| <b>AM</b>  | $0.0147 \pm 0.0021$          | 0.0197                |
| <b>AB</b>  | $0.0020 \pm 0.0004$          | 0.0026                |

**Tabla 12.-** Compilación de las características determinadas para las cuatro proteínas ancestrales, con el promedio de estas además de la comparación con el mismo nodo (TIM63) de la reconstrucción anterior. Los errores de las cinéticas y de la constante de disociación corresponden a los errores asignados por el ajuste al modelo. El promedio y la desviación estándar corresponden a las cuatro proteínas ancestrales de esta reconstrucción

|                  | Hélice | Hebra | Vuelta beta | Random coil | $\lambda_{\max}$ (nm) | Rs (Å) | $k_{\text{cat}}$ ( $\times 10^3 \text{ s}^{-1}$ ) | $K_m$ (mM)      | $k_{\text{cat}}/K_m$ ( $\text{M}^{-1} \text{ s}^{-1}$ ) | Tm (°C)         | $K_{\text{diss}}^{\text{APP}}$ ( $\mu\text{M}$ ) | $\Delta G_{\text{diss}}^{\text{APP}}$ (kcal/mol) |
|------------------|--------|-------|-------------|-------------|-----------------------|--------|---|-----------------|---|-----------------|--|--|
| <b>TIM63</b>     | 41     | 12    | 16          | 29          | 330                   | 29.5   | $0.96 \pm 0.05$                                   | $2.82 \pm 0.34$ | $8.51 \times 10^6$                                      | $65.7 \pm 0.03$ | 0.2  | 9.1  |
| <b>ML</b>        | 34.5   | 16.7  | 17.1        | 27.7        | 332                   | 31.6   | $3.32 \pm 0.34$                                   | $5.99 \pm 0.86$ | $1.39 \pm 0.34 \times 10^7$                             | 86.6            | $0.2610 \pm 0.0873$                              | $9.0 \pm 3.0$                                    |
| <b>BAY</b>       | 36.2   | 14.5  | 16.4        | 29.2        | 332                   | 31.7   | $3.95 \pm 0.31$                                   | $1.64 \pm 0.29$ | $6.01 \pm 1.53 \times 10^7$                             | 78.4            | $0.0070 \pm 0.0013$                              | $11.1 \pm 2.1$                                   |
| <b>AM</b>        | 36.9   | 14.2  | 16.4        | 28.2        | 330                   | 33.5   | $7.25 \pm 0.71$                                   | $2.79 \pm 0.53$ | $6.51 \pm 1.87 \times 10^7$                             | 87.4            | $0.0147 \pm 0.0021$                              | $10.7 \pm 1.5$                                   |
| <b>AB</b>        | 36     | 14.6  | 16.4        | 29.7        | 331                   | 32.9   | $5.23 \pm 0.41$                                   | $2.47 \pm 0.37$ | $5.29 \pm 1.20 \times 10^7$                             | 79.2            | $0.0020 \pm 0.0004$                              | $11.9 \pm 2.4$                                   |
| <b>Prom.</b>     | 35.9   | 15    | 16.58       | 28.7        | 331.2                 | 32.2   | 4.94  | 3.22            | $4.80 \times 10^7$                                      | 82.9            | 0.0368   | 10.7   |
| <b>STD. DEV.</b> | 1.010  | 1.146 | 0.35        | 0.913       | 0.957                 | 0.88   | 1.74  | 1.72            | $2.30 \times 10^7$                                      | 4.8             | 0.060  | 1.2  |



## Discusión

Previo a la discusión de las propiedades de las cuatro proteínas ancestrales, hay que tener en cuenta un detalle muy importante acerca de los resultados obtenidos en la parte bioinformática. El alineamiento, el árbol filogenético, el modelo de sustitución y las secuencias reconstruidas anteriormente no son los resultados “reales”. Son aproximaciones con base en modelos evolutivos aceptados en la actualidad. De esta premisa surge la necesidad de tratar de realizar un análisis que implique un muestreo razonable de las posibilidades. Esto se realizó aquí probando diferentes metodologías y buscando reducir la incertidumbre que introducen los sitios reconstruidos ambiguamente mediante la metodología del peor caso plausible.

Los árboles reconstruidos por los métodos de máxima verosimilitud e inferencia bayesiana son muy parecidos y comparables con los árboles filogenéticos de las especies. Se observan excepciones donde la TIM de ciertas especies no se agrupan dentro del clado que debería pertenecer, ya sea porque la historia evolutiva de la proteína no resulta ser la misma que la de la especie o la aparición de homoplasia (Benner *et al.*, 2007)

Con respecto al modelo de sustitución, resultó ser el mismo modelo al que se recurrió en la reconstrucción anterior (LG). Se observa que la primera, la segunda y la cuarta opción resultan ser el mismo modelo (Tabla 2), lo que aumenta la confianza con respecto a que se escogió un modelo adecuado para la reconstrucción de secuencias. Al ser un modelo de sustitución empírico, no es de extrañar que sea el modelo con mayor verosimilitud para los datos que se presentan, ya que se observan los cambios de aminoácidos tal y como son y no con base en suposiciones teóricas.

Las probabilidades posteriores de las secuencias resultaron ser más altas que el estimado de una buena reconstrucción (0.8), y más alta que las probabilidades posteriores de las anteriores reconstrucciones (>60% de sitios con probabilidades entre 0.9 y 1).

Randall y colaboradores (Randall *et al.*, 2016), estudiaron la capacidad de la técnica de reconstrucción ancestral para dar secuencias y fenotipos correctos. Evaluaron la técnica mediante una reconstrucción de la proteína rojo fluorescente sobre una sucesión de mutaciones que se realizaron experimentalmente y, por lo tanto, conocen con seguridad la filogenia y la secuencia de los ancestros. En ese artículo comparan las secuencias reales con las secuencias predichas por la reconstrucción de secuencias ancestrales y observan un error máximo de 10 sitios inferidos incorrectamente en el nodo más antiguo, de una proteína que contiene 225 residuos. Por lo tanto, la cantidad de errores esperados en las secuencias ancestrales de la TIM no es mayor a la cantidad de errores inherentes a la técnica.

En este mismo artículo (Randall *et al.*, 2016) determinaron qué tanto afectan ciertas variaciones a la reconstrucción de secuencias ancestrales como la cantidad de secuencias, la adecuación de las secuencias a un árbol filogenético realizado a partir de varios marcadores (denominado *species-tree-aware*), la utilización de distintos métodos estadísticos de inferencia y la tasa de variaciones. Determinaron que el número de secuencias, el uso de un árbol de especies o el propio de la

proteína, el uso de diferentes modelos estadísticos, o el uso o no de las tasas de variaciones, no afectan significativamente el fenotipo de las proteínas fluorescentes ancestrales.

La importancia de este trabajo con respecto al anterior (Schulte-Sasse *et al.*, 2019) no radica en rehacer la reconstrucción y probar distintas variaciones en cómo se realiza la reconstrucción de secuencias y mejorar la reconstrucción, ya que está demostrado que éstas no afectan significativamente el fenotipo de las proteínas ancestrales. Este trabajo se centra en aumentar la robustez ante la incertidumbre de la reconstrucción de secuencias sobre el fenotipo de la TIM que resultó con características poco comunes en las TIMs actuales y determinar cuáles características son variables y cuales se mantienen, ya sea por la cooperación de factores variables o por su inherente robustez.

Con esto en mente, se realizó el análisis de las características de las proteínas ancestrales de la TIM sobre el nodo del ancestro de los opistocontes.

Las propiedades espectroscópicas indican que las cuatro proteínas ancestrales poseen estructura secundaria (DC) y terciaria (fluorescencia) comparables a la de las proteínas actuales, es decir, los espectros y los porcentajes de estructura secundaria son muy similares con los de un barril TIM actual (Tabla 5).

Al excitar la muestra a 295 nm, se están excitando principalmente los triptófanos. Si tomamos en cuenta que las cuatro proteínas poseen el mismo número de triptófanos, la fluorescencia máxima de los espectros debería ser la misma. Esto no ocurre así ya que los espectros difieren bastante en cuanto a intensidad máxima. Esta diferencia puede deberse a que los aminoácidos aromáticos no se encuentran en el mismo ambiente. Se ha observado que la fluorescencia del triptófano se ve disminuida, cuando hay una histidina protonada cerca (Lowenthal *et al.*, 1991) o aumentada, por transferencia energética de la tirosina (Boteva *et al.*, 1996).

Además de poseer estructura secundaria y terciaria, las proteínas ancestrales son capaces de dimerizarse, comparando su radio hidrodinámico con el de proteínas actuales dimericas ( $29.6 \pm 0.25 \text{ \AA}$ ; Morgan *et al.*, 2000). Podemos observar, para el caso de la proteína reconstruida por el método de máxima verosimilitud, dos picos cuando la proteína se encuentra diluida, en tanto que cuando está concentrada se observa un pico definido y un pico de menor intensidad y más ancho (Figura 17, panel A). El segundo pico de la proteína M, presenta una diferencia de  $3.4 \text{ \AA}$  con el monómero de ScTIM ( $23.5 \text{ \AA}$ ) y si se compara con el monómero de la TIM de *Entamoeba histolytica* obtenido al calentar la proteína ( $26.7 \text{ \AA}$ ; Tellez *et al.*, 2008) la diferencia es únicamente de  $0.1 \text{ \AA}$ . Por lo que lo más probable es que este pico sea el componente monomérico.

Tras confirmar que las enzimas reconstruidas poseen la estructura de un barril TIM y estado oligomérico dimerico, se determinó la actividad catalítica, ya que el tener el mismo plegamiento no implica que sean capaces de realizar la catálisis correspondiente a la isomerización. Se observó que las cuatro enzimas son capaces de realizar la reacción de isomerización. Sin embargo, si comparamos con el rango propuesto por Alberty & Hammes (1958), las cuatro enzimas se encuentran un orden de magnitud por debajo del límite por difusión. No hay que ser totalmente estrictos con estos límites, ya que a lo largo de medio siglo se han caracterizado una gran cantidad

de TIMs con una eficiencia catalítica muy variable que va desde el orden de  $10^5$  a  $10^9$   $M^{-1} s^{-1}$ . En la tabla 12 podemos observar las eficiencias catalíticas de la TIM de diversos opistocontes, desde *Tenebrio molitor* hasta *Saccharomyces cerevisiae*. De estas siete especies, dos presentan la eficiencia catalítica en el mismo orden de magnitud que las enzimas reconstruidas, mientras que las otras cinco presentan un orden de magnitud mayor ( $10^8$   $M^{-1} s^{-1}$ ). Así mismo observamos que los valores de  $K_m$  y  $k_{cat}$  no varían entre este grupo, a excepción de la TIM de conejo y de la levadura.

**Tabla 13.-** Parámetros catalíticos de TIMs de diferentes organismos. Los organismos se encuentran agrupados en dominios.

| Organismo                       | $K_m$ (mM)      | $k_{cat}$ ( $min^{-1}$ )      | Eficiencia catalítica ( $M^{-1} s^{-1}$ ) | Referencias   |
|---------------------------------|-----------------|-------------------------------|---|---|
| <b>Bacteria</b>                 |                 |                               |   |   |
| <i>Escherichia coli</i>         | $1.03 \pm 0.10$ | $(5.4 \pm 0.1) \times 10^5$   | $(2.18 \pm 0.24) \times 10^8$             | Alvarez <i>et al.</i> , 1998 <sup>a</sup>   |
| <i>Moritella marina</i>         | $1.90 \pm 0.20$ | $(4.2 \pm 0.1) \times 10^5$   | $(9.21 \pm 0.42) \times 10^7$             | Alvarez <i>et al.</i> , 1998 <sup>a</sup>   |
| <i>Streptomyces coelicolor</i>  | $0.74 \pm 0.02$ | $(3.40 \pm 0.03) \times 10^5$ | $(1.93 \pm 0.06) \times 10^8$             | Romero-Romero <i>et al.</i> , 2016 <sup>b</sup>                                     |
| <i>Deinococcus radiodurans</i>  | $0.69 \pm 0.03$ | $(3.40 \pm 0.05) \times 10^5$ | $(2.06 \pm 0.18) \times 10^8$             | Romero-Romero <i>et al.</i> , 2016 <sup>b</sup>                                     |
| <i>Nostoc punctiforme</i>       | $0.51 \pm 0.04$ | $(1.04 \pm 0.02) \times 10^5$ | $(8.63 \pm 0.29) \times 10^7$             | Romero-Romero <i>et al.</i> , 2016 <sup>b</sup>                                     |
| <i>Gemmata obscuriglobus</i>    | $0.34 \pm 0.02$ | $(7.83 \pm 0.16) \times 10^4$ | $(9.51 \pm 0.14) \times 10^7$             | Romero-Romero <i>et al.</i> , 2016 <sup>b</sup>                                     |
| <i>Clostridium perfringens</i>  | $0.72 \pm 0.04$ | $(3.80 \pm 0.07) \times 10^5$ | $(2.24 \pm 0.10) \times 10^8$             | Romero-Romero <i>et al.</i> , 2016 <sup>b</sup>                                     |
| <b>Eucariota</b>                |                 |                               |   |   |
| <i>Trichomonas vaginalis</i>    | $0.21 \pm 0.02$ | $8.2 \times 10^4$             | $1.63 \times 10^8$                        | Figuroa-Angulo <i>et al.</i> , 2012 <sup>c</sup>                                    |
| <i>Taenia solium</i>            | $0.37 \pm 0.05$ | $7.5 \times 10^4$             | $8.49 \times 10^7$                        | Jiménez <i>et al.</i> , 2003; Jiménez <i>et al.</i> , 2000 <sup>c</sup>             |
| <i>Plasmodium falciparum</i>    | 0.58            | $1.1 \times 10^5$             | $7.54 \times 10^7$                        | Joubert <i>et al.</i> , 2001 <sup>d</sup>   |
| <i>Trypanosoma brucei</i>       | $0.25 \pm 0.05$ | $3.7 \times 10^5$             | $6.17 \times 10^8$                        | Hernández-Alcántara <i>et al.</i> , 2002; Lambeir <i>et al.</i> , 1987 <sup>a</sup> |
| <i>Tenebrio molitor</i>         | 0.84            | $2.6 \times 10^5$             | $1.28 \times 10^8$                        | Knobeloch <i>et al.</i> , 2010 <sup>e</sup>   |
| <i>Gallus gallus</i>            | 0.47            | $2.6 \times 10^5$             | $2.27 \times 10^8$                        | Plaut y Knowles, 1972; Putman <i>et al.</i> , 1972 <sup>a</sup>                     |
| <i>Homo sapiens</i>             | 0.49            | $2.7 \times 10^5$             | $2.30 \times 10^8$                        | Mainfroid, <i>et al.</i> , 1996; Mande <i>et al.</i> , 1994 <sup>a</sup>            |
| <i>Oryctolagus cuniculus</i>    | $0.32 \pm 0.02$ | $5.2 \times 10^4$             | $6.77 \times 10^7$                        | Knobeloch <i>et al.</i> , 2010; Krietsch <i>et al.</i> , 1970 <sup>a</sup>          |
| <i>Sus scrofa</i>               | 0.51            | $1.2 \times 10^5$             | $9.33 \times 10^7$                        | Jiménez <i>et al.</i> , 2003 <sup>c</sup>   |
| <i>Saccharomyces cerevisiae</i> | $1.27 \pm 0.06$ | $1.0 \times 10^6$             | $3.28 \times 10^8$                        | Lambeir <i>et al.</i> , 2000; Krietsch <i>et al.</i> , 1970 <sup>a</sup>            |

| Organismo                        | $K_m$ (mM)      | $k_{cat}$ ( $\text{min}^{-1}$ ) | Eficiencia catalítica ( $\text{M}^{-1} \text{s}^{-1}$ ) | Referencias                                      |
|----------------------------------|-----------------|---------------------------------|---|--|
| <i>Chlamydomonas reinhardtii</i> | $3.04 \pm 0.1$  | $(2.02 \pm 0.02) \times 10^5$   | $1.12 \times 10^5$                                      | Zaffagnini <i>et al.</i> , 2014 <sup>c</sup>     |
| <i>Arabidopsis thaliana</i>      | $0.48 \pm 0.05$ | $(1.54 \pm 0.02) \times 10^5$   | $5.34 \times 10^6$                                      | López-Castillo <i>et al.</i> , 2016 <sup>c</sup> |
| <b>Arquea</b>                    |                 |                                 |   |  |
| <i>Metanococcus jannaschii</i>   | $0.19 \pm 0.01$ | $(5.36 \pm 0.09) \times 10^4$   | $(1.4 \pm 0.9) \times 10^8$                             | Muñiz -Trejo, 2017 <sup>b</sup>                  |
| <i>Methanococcus aeolicus</i>    | $0.6 \pm 0.04$  | $(4.36 \pm 0.12) \times 10^3$   | $(2.3 \pm 0.33) \times 10^6$                            | Muñiz -Trejo, 2017 <sup>b</sup>                  |
| <i>Methanococcus maripaludis</i> | $0.8 \pm 0.03$  | $(1.62 \pm 0.27) \times 10^3$   | $(9,9 \pm 0.57) \times 10^5$                            | Muñiz -Trejo, 2017 <sup>b</sup>                  |

<sup>a</sup> Valores obtenidos a 25 °C, pH 7.6; <sup>b</sup>Valores obtenidos a 25 °C, pH 8.0; <sup>c</sup> Valores obtenidos a pH 7.4, temperatura no especificada; <sup>d</sup>Valores obtenidos a 25 °C, pH 7.9; <sup>e</sup> Valores obtenidos a 25 °C, pH 7.4

Sin duda, la característica en la que existe más variabilidad entre las 4 proteínas ancestrales es la energía de interacción entre los monómeros. La variabilidad la podemos observar con los experimentos de DSC y de estabilidad a la dilución.

Los experimentos de estabilidad térmica seguidos por DSC muestran que la energía necesaria, para disociar y desplegar a las proteínas es diferente. Se puede dividir en dos grupos dependiendo de su  $T_m$ , las obtenidas por inferencia bayesiana con una  $T_m$  de 78.4 y 79.2 °C y las obtenidas por máxima verosimilitud con una  $T_m$  de 86.6 y 87.4 °C. Hay que notar que ambos valores de  $T_m$  son elevados cuando se comparan con la  $T_m$  de TIMs de algunos organismos actuales.

Así mismo, podemos observar una diferencia notable de 664 kJ/mol, en el  $\Delta H$  entre la proteína alternativa de máxima verosimilitud y la proteína alternativa de inferencia bayesiana. Si comparamos el  $\Delta H$  de las secuencias ancestrales obtenidas con el método de máxima verosimilitud con la TIM de *Entamoeba histolytica*, observamos que no hay diferencia significativa (EhTIM, 1111 kJ/mol; Tellez *et al.*, 2008), sin embargo, observamos una diferencia del doble de  $\Delta H$  para las proteínas obtenidas con inferencia bayesiana (Tabla 7).

Es importante resaltar la discrepancia de casi 10 °C entre las  $T_m$ s de los dos métodos de reconstrucción; resultados semejantes ya se han observado antes. En una comparación de distintos métodos de reconstrucción para comprobar su precisión mediante simulaciones computacionales, se llegó a la conclusión de que los métodos de máxima parsimonia y máxima verosimilitud sobrestiman la termoestabilidad mientras que la inferencia bayesiana no lo hace. Esta propiedad se observa, ya que los aminoácidos que tienden a disminuir la estabilidad de la proteína ancestral se reconstruyen con una probabilidad posterior menor en el método de máxima verosimilitud y, por lo tanto, no son escogidos para la secuencia más probable. (Williams *et al.*, 2006). Por el contrario, cuando se observa la actividad de la proteína ancestral, el método que ofrece una mejor precisión es el de máxima verosimilitud sobre la inferencia bayesiana.

La TIM63, inferida con el método de máxima verosimilitud, en la reconstrucción anterior (Schulte-Sasse *et al.*, 2019) tiene una  $T_m$  idéntica a la  $T_m$  de ScTIM (65.8 °C, Peimbert *et al.*, 2008). Por lo que no se observa el aumento de la  $T_m$  típico en las reconstrucciones ancestrales. Es posible que esta diferencia se deba al sesgo de las secuencias utilizadas, pero sin más información relacionada es difícil determinar la razón. Una investigación más profunda en este ámbito podría ayudar a evitar la sobrestimación de la  $T_m$  por el método de máxima verosimilitud.

Hart y colaboradores (Hart *et al.*, 2014) realizaron una investigación sobre proteínas termoestables y al obtener las secuencias alternativas, observan diferencias en un intervalo 5 °C entre 10 secuencias diferentes. Este rango es ligeramente mayor al obtenido en este trabajo para las secuencias alternativas de cada uno de los métodos, de apenas 0.8 °C. En este trabajo observamos sobrestimación de la termoestabilidad en secuencias inferidas por máxima verosimilitud, por lo que podríamos sugerir que la temperatura media de desnaturalización más probable de la TIM del último ancestro de los opistocontes es la estimada por los métodos bayesianos; por lo que esta se encontraría entre 78.4 y 79.2 °C.

La alta temperatura de desnaturalización de diferentes reconstrucciones ancestrales ha sido tema de discusión en diversos artículos y se ha relacionado con una adaptación a las altas temperaturas de los océanos en tiempos ancestrales (Gaucher *et al.*, 2008). Sin embargo, Trudeau y colaboradores (2016) observaron que la paraxonasa sérica del ancestro de los mamíferos posee una  $T_m$  30 °C mayor a la proteína actual de humano, siendo que la temperatura de la tierra era la misma que ahora cuando surgió este ancestro (Trudeau *et al.*, 2016).

Trudeau y colaboradores (2016) proponen dos teorías que explican la alta termoestabilidad. Proponen que puede ser consecuencia de un efecto consenso como el observado por la sobrestimación de la  $T_m$  por la inserción de aminoácidos no disruptivos en la secuencia como mencionan Williams y colaboradores (2006). Si bien, no se puede descartar el efecto de consenso para el aumento de la estabilidad de una proteína, una comparación hecha por Riso y colaboradores (2013) con  $\beta$ -lactamasas ancestrales y la secuencia consenso indican que, a pesar de que la secuencia consenso aumenta su termoestabilidad, no es tan grande como la termoestabilidad obtenida para la secuencia ancestral.

Otra razón que Trudeau y colaboradores (2016) proponen para explicar la alta termoestabilidad de las proteínas ancestrales es que ésta es un reflejo del aumento general de la estabilidad de la proteína a factores que afecten su función. Entre estos factores proponen a los errores de transcripción y traducción, incapacidad de las proteínas de replegarse por sí mismas y factores físicos y químicos. Esto lo observan por la resistencia a la quelación por calcio y a la radiación UV de la paraxonasa sérica ancestral. Si bien, podemos observar el efecto que tiene integrar aminoácidos con efecto consenso en las secuencias de máxima verosimilitud, no podemos asegurar que las secuencias obtenidas por inferencia bayesiana presentan un aumento de temperatura por este mismo efecto o como resultado de un aumento general en la estabilidad de la proteína y que también se refleja en una  $T_m$  alta. Solamente estudios parecidos a los realizados por Trudeau y colaboradores (2016) podrían indicar por qué las TIMs estudiadas en este trabajo poseen una mayor estabilidad y, aun así, no podemos asegurar que esto sea un artificio del método.

En los experimentos de estabilidad a la dilución, al seguir la actividad enzimática a diferentes concentraciones de proteína, observamos que las TIMs ancestrales se disocian al diluirse y los monómeros son inactivos. La constante de disociación varía en tres órdenes de magnitud entre las cuatro proteínas ancestrales, siendo la proteína obtenida por máxima verosimilitud la más sensible a este fenómeno. También observamos que el ajuste al modelo de disociación de dos estados no es adecuado. Este modelo implica que el proceso de disociación/asociación es reversible. Para algunas TIMs bacterianas se ha encontrado experimentalmente que la disociación es reversible y que los datos de la curva de dilución se ajustan adecuadamente al modelo (Romero-Romero *et al.* 2015). En contraste, la disociación de la TIM de maíz es irreversible y los datos de la curva de dilución no se ajustan al modelo de dos estados. Además, la inactivación ocurre en un intervalo de concentraciones mucho menor al predicho por el modelo (Romero-Romero *et al.* 2018). Ya que en nuestro caso no hemos comprobado experimentalmente si la disociación de las TIMs ancestrales es reversible, además de calcular la  $K_{diss}$  mediante la ecuación de dos estados reversibles, determinamos la disociación como la concentración donde la mitad de la actividad se pierde ( $K_{0.5}$ ), se observan valores ligeramente aumentados pero similares (Tabla 11).

Las diferencias entre las constantes de disociación de las TIMs ancestrales parecieran ser altas, pero al comparar con las constantes de otras TIM actuales observamos que también existe variación (Tabla 13). Las constantes de disociación de las proteínas ancestrales son entre dos y cuatro órdenes de magnitud menor que las TIM de conejo, levadura y humano, descendientes del ancestro de los opistocontes. En particular la TIM obtenida por máxima verosimilitud que tiene una  $K_{diss}$  del orden de  $10^{-7}$ , dentro del grupo eucarionte es la TIM que se disocia más fácilmente, es decir, a concentraciones altas, mientras que es común ver que en el grupo de las bacterias también tienden a disociarse a concentraciones altas de proteína.

**Tabla 14.-** Constantes de disociación de varias TIM actuales de distintos clados.

| <b>Organismo</b>                          | <b><math>K_{diss}</math></b>                 | <b>Referencias</b>                      |
|---|--|---|
| <b>Eucariota</b>                          |  |   |
| <i>Homo sapiens</i>                       | $3.21 (\pm 2.3) \times 10^{-11}$ M -ND       | Mainfroid <i>et al.</i> , 1996          |
| <i>Oryctolagus cuniculus</i> <sup>a</sup> | 2 ng/mL ( $7.51 \times 10^{-11}$ M)-<br>ND   | Zabori <i>et al.</i> , 1980             |
| <i>Entamoeba histolytica</i> <sup>a</sup> | 0.3 ng/mL ( $1.07 \times 10^{-11}$ M)-<br>ND | Landa <i>et al.</i> , 1997              |
| <i>Trypanosoma brucei</i>                 | $3.1 \times 10^{-8}$ M-ND                    | Reyes-Vivas <i>et al.</i> , 2002        |
| <i>Trypanosoma cruzi</i>                  | $4.11 \times 10^{-8}$ -ND                    | Zomosa-Signoret <i>et al.</i> ,<br>2003 |
| <i>Giardia lamblia</i>                    | $5.06 \pm 12 \times 10^{-8}$ M -Rev          | López-Velazquez <i>et al.</i> ,<br>2004 |
| <i>Zea mays</i>                           | $6.09 \pm 2.38 \times 10^{-8}$ M -Irrev      | Romero-Romero <i>et al.</i> ,<br>2018   |

| <b>Organismo</b>                | <b>K<sub>diss</sub></b>               | <b>Referencias</b>                 |
|---------------------------------|---------------------------------------|------------------------------------|
| ML                              | 2.3 × 10 <sup>-7</sup> M - ND         | Este trabajo                       |
| BAY                             | 7.0 × 10 <sup>-9</sup> M - ND         | Este trabajo                       |
| AML                             | 1.47 × 10 <sup>-8</sup> M - ND        | Este trabajo                       |
| ABAY                            | 2.0 × 10 <sup>-9</sup> M - ND         | Este trabajo                       |
| <b>Bacteria</b>                 |                                       |                                    |
| <i>Saccharomyces cerevisiae</i> | 1.01 × 10 <sup>-11</sup> M            | Nájera <i>et al.</i> , 2003        |
| <i>Nostoc punctiforme</i>       | 5.6 ± 0.7 × 10 <sup>-8</sup> M -Irrev | Romero-Romero <i>et al.</i> , 2016 |
| <i>Gemmata obscuriglobus</i>    | 1.1 ± 0.2 × 10 <sup>-7</sup> M -Irrev | Romero-Romero <i>et al.</i> , 2016 |
| <i>Deinococcus radiodurans</i>  | 1.4 ± 0.5 × 10 <sup>-7</sup> M -Rev   | Romero-Romero <i>et al.</i> , 2016 |
| <i>Streptomyces coelicolor</i>  | 1.8 ± 0.5 × 10 <sup>-7</sup> M -Rev   | Romero-Romero <i>et al.</i> , 2016 |
| <i>Clostridium perfringens</i>  | 2.9 ± 0.3 × 10 <sup>-8</sup> M -Rev   | Romero-Romero <i>et al.</i> , 2016 |

<sup>a</sup> Las constantes de las TIMs no se obtuvieron por un ajuste al modelo. Se obtuvieron las constantes a partir de la concentración en ng/mL donde la proteína pierde actividad. Se coloca la reversibilidad ante la dilución ya sea reversible (Rev), irreversible (Irrev) o no determinado (ND).

Podemos observar que la TIM inferida por máxima verosimilitud posee una constante de disociación más parecidas a la observada en la TIMs de protistas que a la observada en la TIMs de opistocontes.

Ya que MLTIM se disocia con relativa facilidad, se analizó la secuencia de esta proteína para determinar a qué se debe este comportamiento. Hay 7 sitios en los que MLTIM presenta residuos que no se observan en las otras tres reconstrucciones. De éstos, uno se encuentra en la región del carboxilo terminal, por lo tanto, no se analizó. Los demás son isoleucina en el sitio 17, aspartato en el 34, serina en el 53, glutamina en el 99, asparagina en el 118 y treonina en el 143 (Tabla 14).

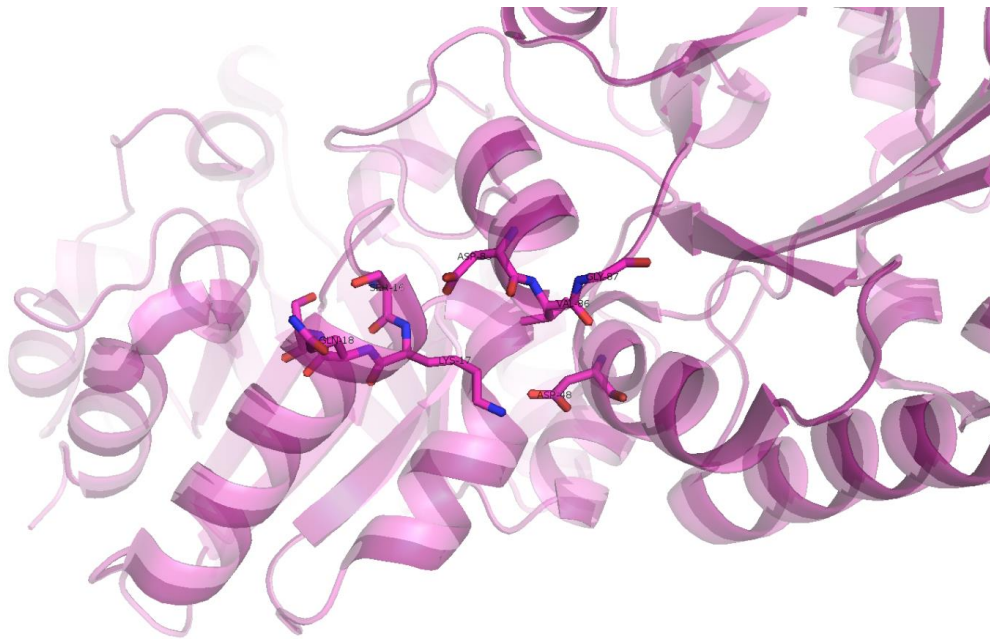
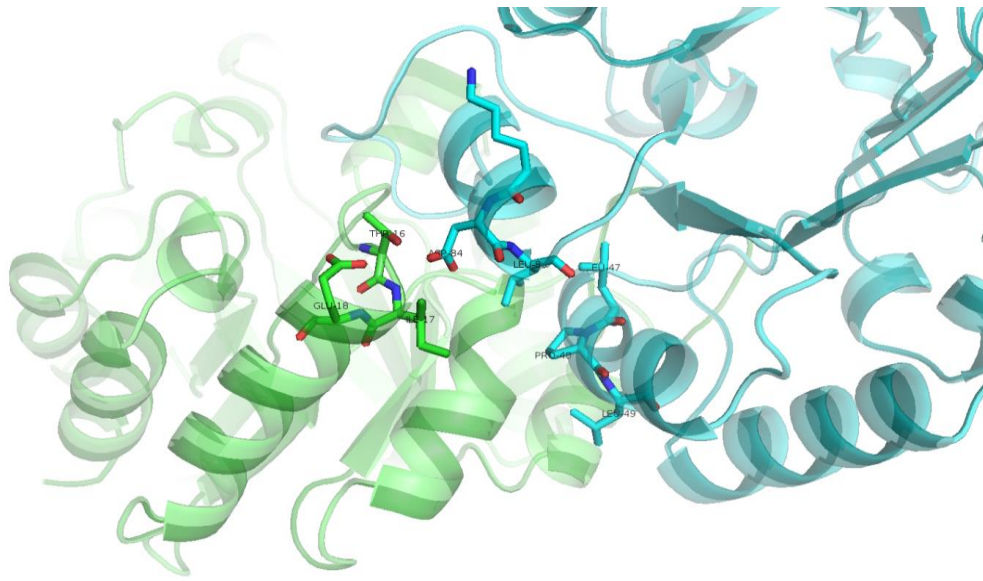
**Tabla 15.-** Identidad y localización de los sitios a analizar de la proteína de máxima verosimilitud.

|              | Sitio (Ubicación en estructura secundaria) |                        |                   |                        |                         |                    |
|--------------|--|------------------------|-------------------|------------------------|-------------------------|--------------------|
|              | 17 ( $\alpha 1$ )                          | 34 ( $\alpha\beta 1$ ) | 53 ( $\alpha 2$ ) | 99 ( $\beta\alpha 4$ ) | 118 ( $\alpha\beta 4$ ) | 143 ( $\alpha 5$ ) |
| <b>ML</b>    | I  | D                      | S                 | Q                      | N                       | T                  |
| <b>BAY</b>   | V  | P                      | A                 | H                      | S                       | A                  |
| <b>AM</b>    | K  | N                      | A                 | H                      | Q                       | A                  |
| <b>AB</b>    | Q  | N                      | Q                 | N                      | A                       | A                  |
| <b>ScTIM</b> | K  | E                      | L                 | S (100)                | Q (119)                 | E (144)            |

Se muestran los sitios de las otras secuencias para su comparación.  $\alpha$ , hélice alfa;  $\alpha\beta$ , asa inferior del barril TIM;  $\beta\alpha$ , asa superior del barril TIM.

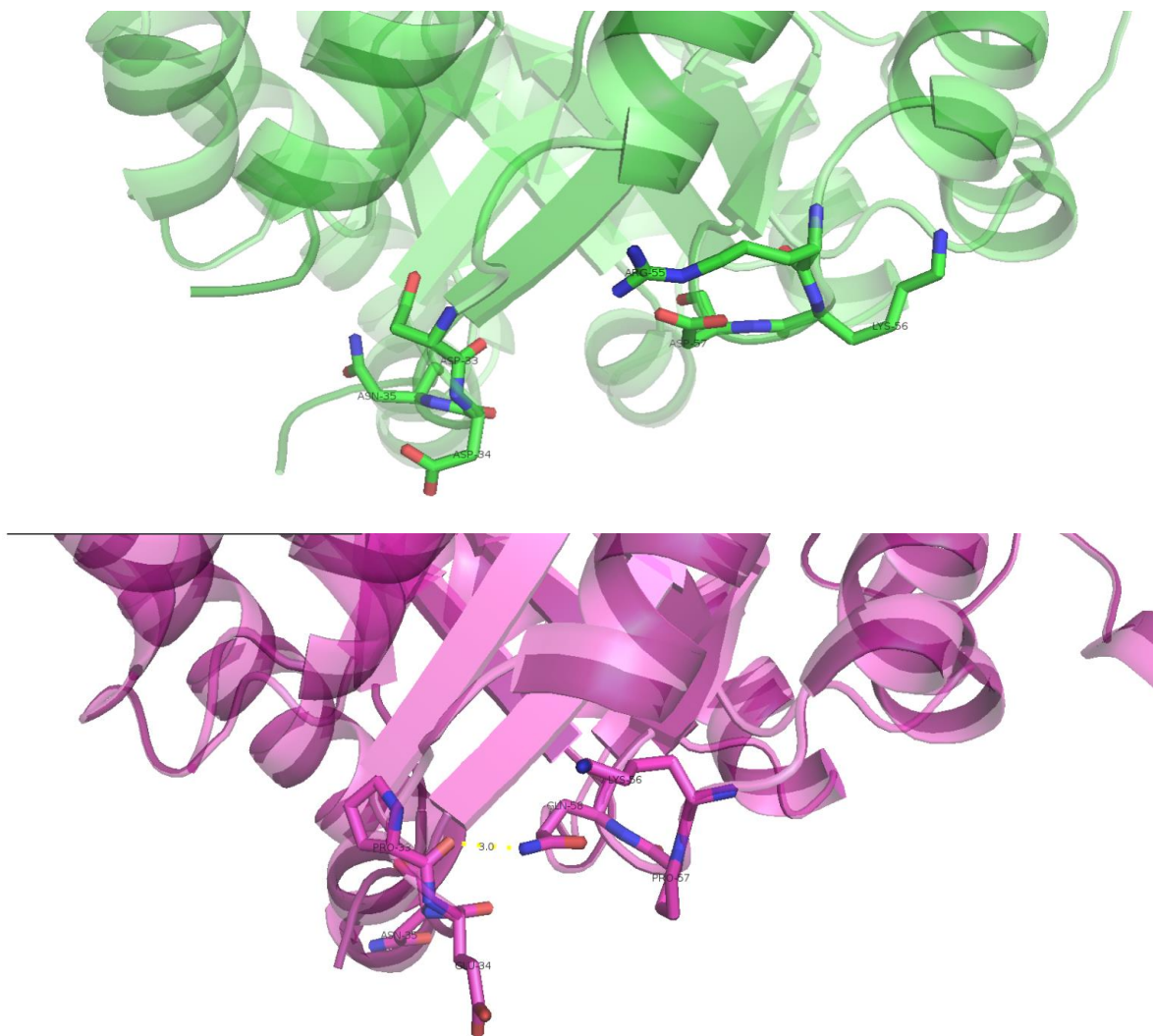
Se comparó la estructura cristalográfica de la TIM de *Saccharomyces cerevisiae* (PDB: 1I45) con un modelo predicho por ITasser para la proteína inferida por máxima verosimilitud. Se escogió ScTIM al ser una TIM resistente a la disociación por dilución.





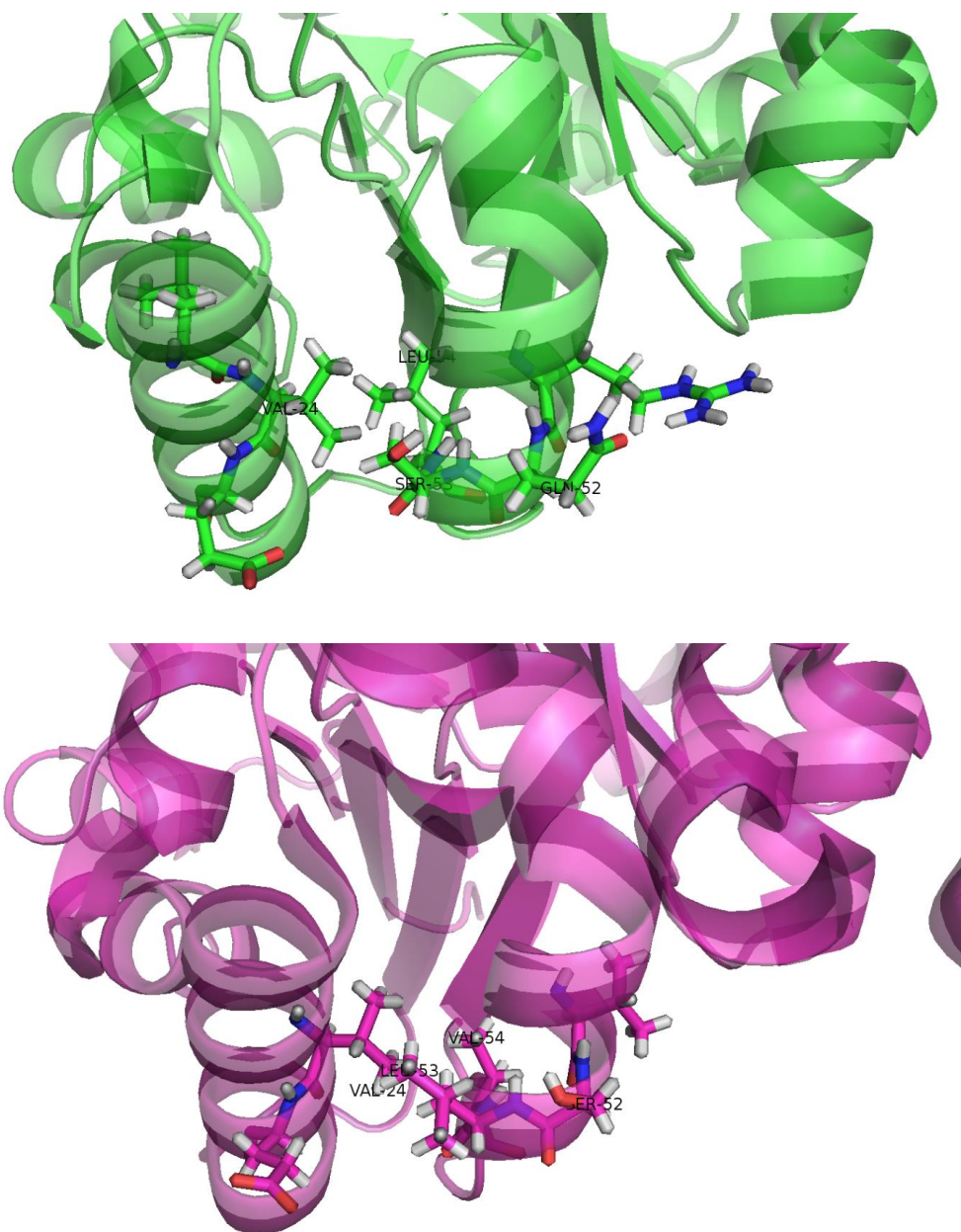
**Figura 19.-** Localización del sitio Ile 17 en la TIM inferida por ML (verde y azul el otro monómero) y Lys 17 en ScTIM (morada).

En la Figura 19, en ScTIM , podemos observar a la lisina 17 interaccionando con el aspartato 48 de la hélice 2 de la otra subunidad. Este puente salino es común en varias especies de animales y ciertas especies de hongos. También se observa la interacción del aspartato 84 en la hélice 3 con el amino de la cadena principal. Para la proteína ancestral, se observa una isoleucina 17 en lugar de la lisina. Por lo tanto, no se establece el puente de salino como se observa en la ScTIM. Sí se observa la interacción del aspartato 84 con el amino de la cadena principal. A pesar de que la secuencia obtenida por inferencia bayesiana (BAY) posee esta lisina, ninguna de las cuatro proteínas ancestrales posee el aspartato en la hélice 2 para generar este puente salino. Todas las secuencias poseen una prolina.



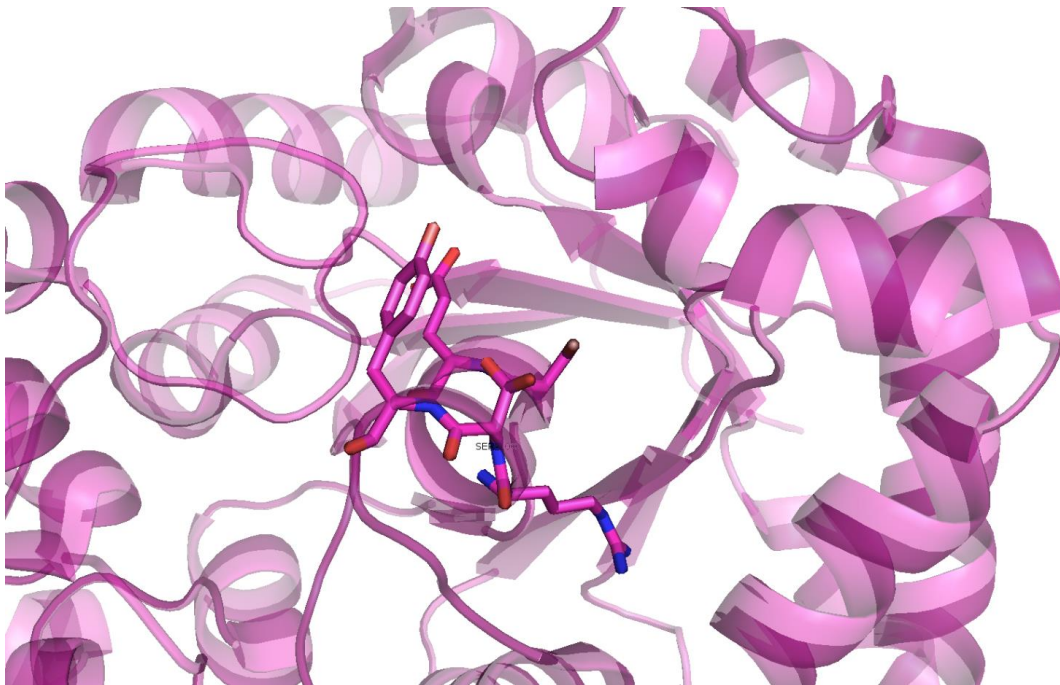
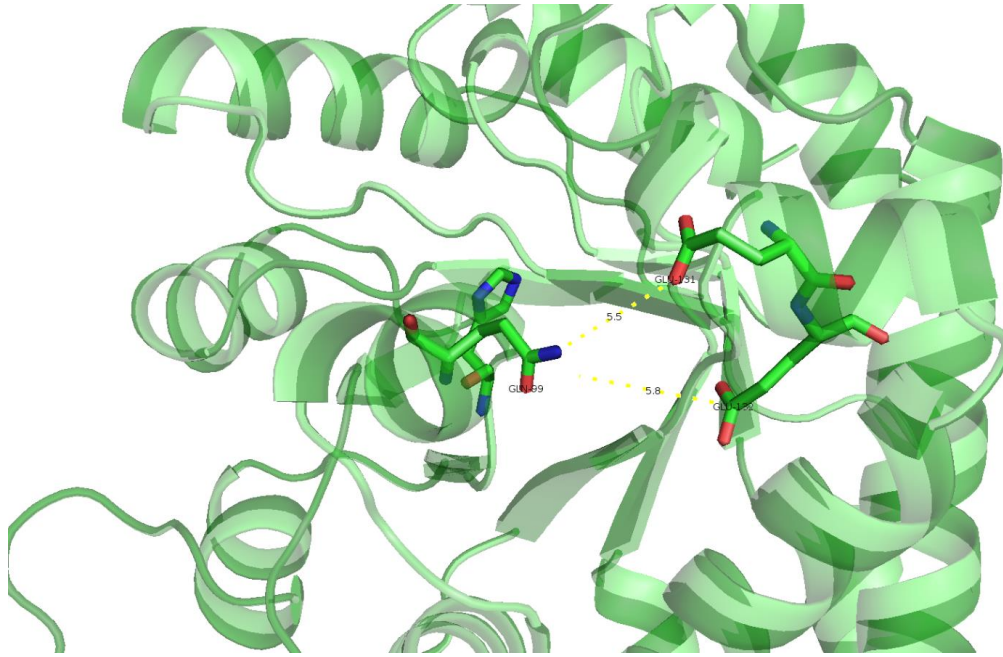
**Figura 20.-** Localización del sitio Asp 34 en la TIM inferida por ML (verde) y Glu 34 ScTIM (morada).

Para el aspartato en el sitio 34, observamos que se encuentra en el asa inferior del barril (Figura 20). En ScTIM encontramos un glutamato en ese sitio. Lo relevante es una posible interacción que se puede dar entre la glutamina 58 y el carboxilo de la cadena principal. Sin embargo, a pesar de que MLTIM tiene una arginina (en posición 55) que podría realizar una interacción parecida, no se encuentra lo suficientemente cerca. Esto debido a que en las proteínas ancestrales se introdujo una deleción en el asa donde se encuentra esa arginina. Aun así, a todas las proteínas se les introdujo esta deleción por lo que es improbable que exista una interacción parecida a la de ScTIM para alguna de las 4 (en el asa inferior entre la hélice 2 y la hebra 3).



**Figura 21.-** Localización del sitio Ser53 en la TIM inferida por ML (verde) y Leu 53 en ScTIM (morada).

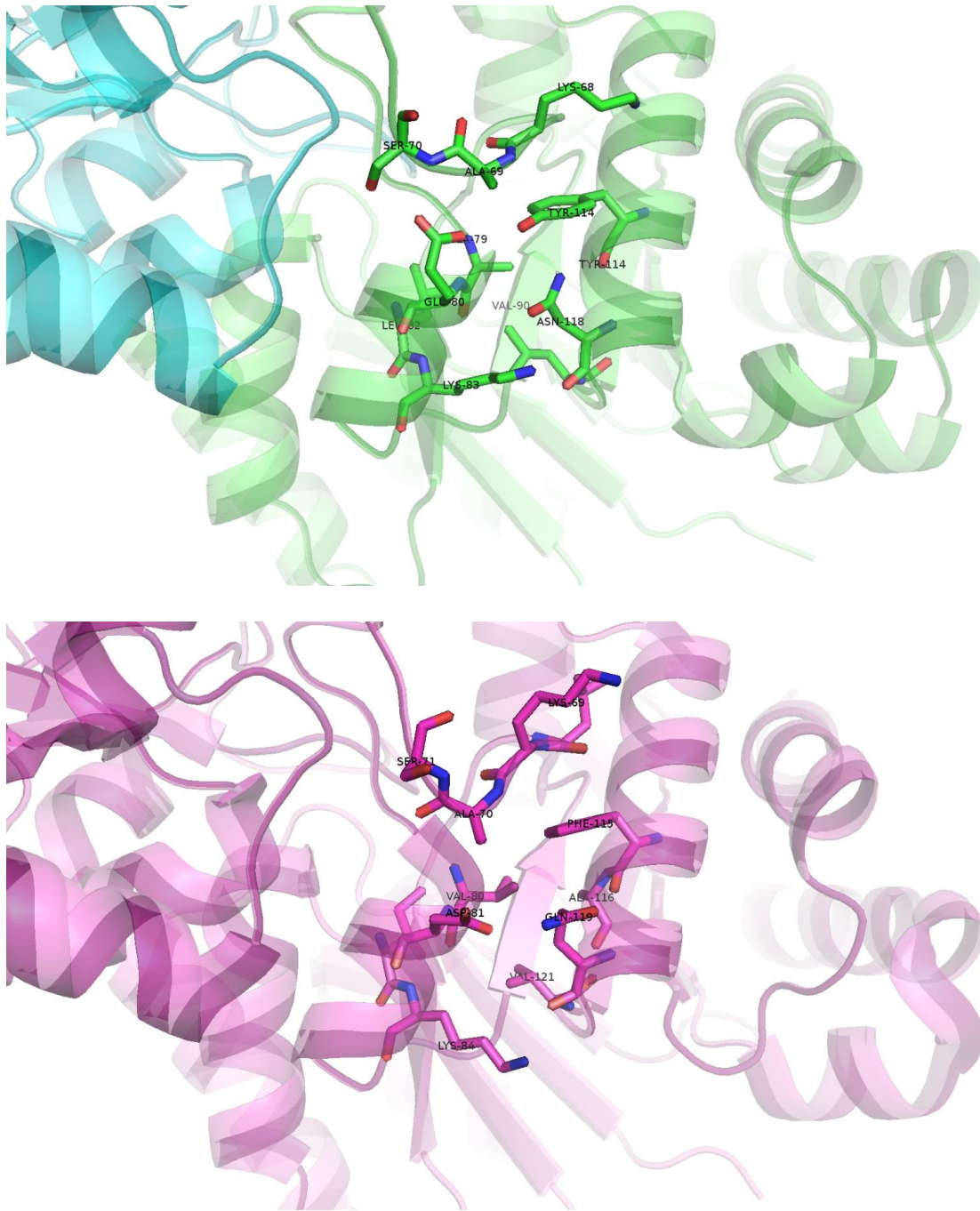
La serina en posición 53 de MLTIM se observa similar a la leucina 53 de ScTIM (Figura 21). Se observan interacciones tipo dispersión entre las dos hélices (hélice 1 y 2) en ScTIM entre esa leucina 53 y la valina 24. En MLTIM, se observa que la serina no se encuentra interaccionando con algún aminoácido, sin embargo, se observa que la interacción se mantiene entre la leucina 54 y la valina 24.



**Figura 22.-** Localización del sitio Gln 99 en la TIM inferida por ML (verde) y Ser 100 en ScTIM (morada).

La glutamina 99, presente en el asa 4, forma una pequeña hélice alfa. En ScTIM hay una serina (en posición 100) que no se encuentra interaccionando con algún residuo en específico (Figura 22). Cabe señalar que éste aminoácido se encuentra precedido por la región de la histidina catalítica que se encuentra altamente conservada, sin embargo, a partir de este residuo disminuye la conservación (Figura 4). En MLTIM, la glutamina se encuentra muy lejos para interaccionar con

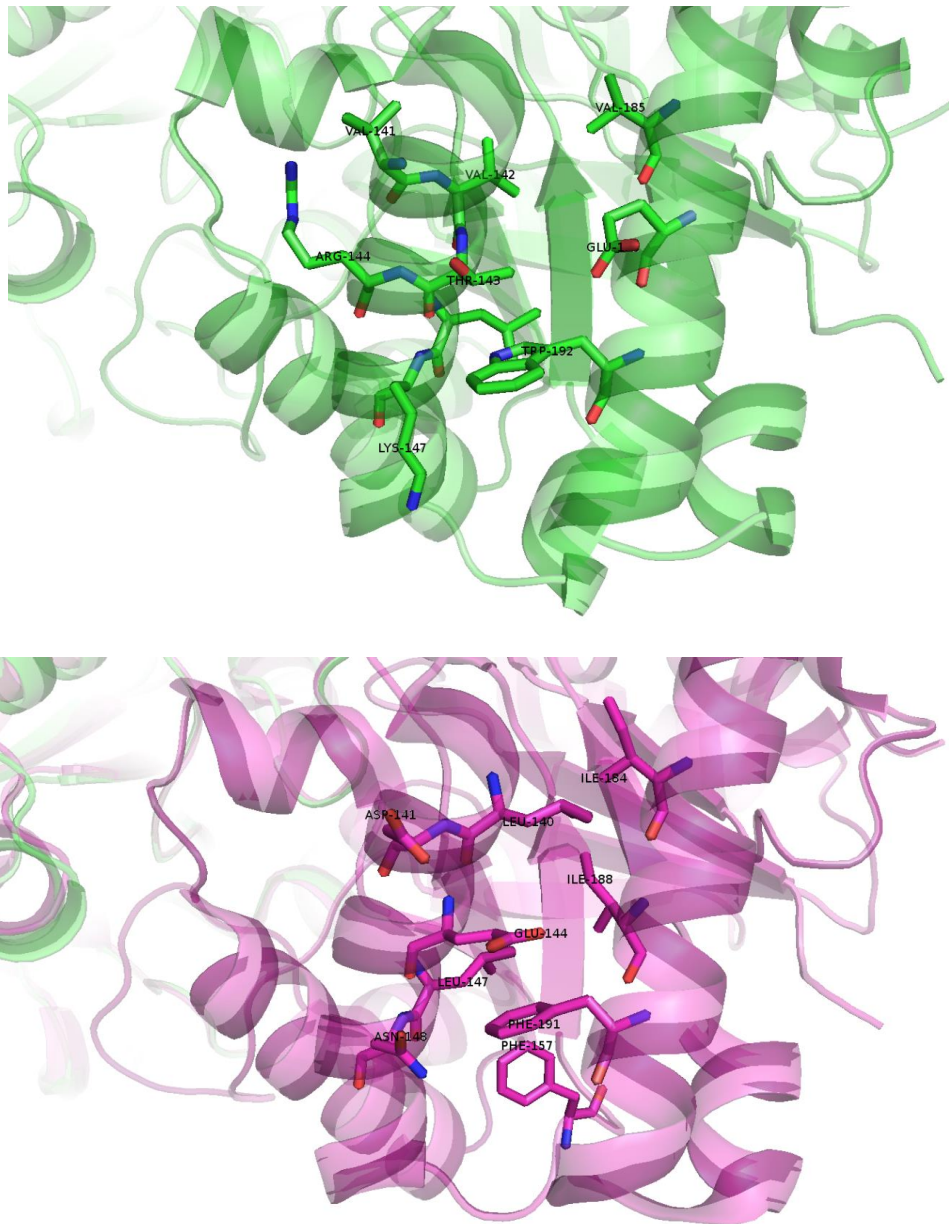
cualquier residuo mediante puentes de hidrógeno, por lo que, probablemente, este residuo funcione únicamente para la formación de la hélice alfa.



**Figura 23.-** Localización del sitio Asn 118 en la TIM inferida por ML (verde y en azul el otro monómero) y la Gln 119 en ScTIM (morada).

Se observa en la figura 23 la localización de la asparagina 118 en MLTIM en la hélice 4, cuya cadena lateral se encuentra interaccionando con el carbonilo del enlace peptídico de la tirosina 114 y también se encuentra muy cerca del hidroxilo de esa misma tirosina. Es una red de enlaces diversos en esa zona de la proteína. También, se observa una lisina muy cerca de la misma

asparagina y del carbonilo de la cadena principal. En lugar de esa asparagina, en ScTIM se observa una glutamina interaccionando con un aspartato en la hélice 3, además de eso, no se observan otras interacciones importantes.



**Figura 24.-** Localización del sitio Thr 143 en la TIM inferida por ML (verde) y Glu144 en la ScTIM (morada).

La treonina 143 ubicada en la hélice 5 no interacciona con algún residuo cercano. De igual manera, el glutamato 144 de ScTIM se encuentra con su cadena lateral hacia el exterior de la proteína. Se observan en ambas proteínas que los residuos de la hélice 5 y 6 existen interacciones hidrofóbicas.

De entre los residuos analizados, únicamente la asparagina 118 parece generar más interacciones que en ScTIM. La serina 53, y treonina 143, se encuentran con interacciones parecidas con ScTIM.

Los demás residuos no se encuentran interaccionando con algún residuo cercano, aun así, las demás proteínas ancestrales (BAY, AM y AB) tampoco son capaces de realizar estas interacciones. Curiosamente, todos estos aminoácidos analizados se encuentran en la interfaz entre hélices alfa, a excepción de la isoleucina 17 que de ser lisina podría formar un puente salino. Por lo que, de los residuos analizados, no hay ninguno que aparentemente este causando la disrupción de entre los monómeros. Hay que tener en cuenta que las suposiciones que se realizaron anteriormente son basadas en una estructura inferida por homología. Se deberían confirmar estas suposiciones con base en la estructura obtenida a partir de la difracción de cristales de proteína.

Se ha observado una disminución de la estabilidad del dímero al mutar residuos que se encuentran en la interfaz entre monómeros, que se refleja en una disminución de cuatro órdenes de magnitud en la constante de disociación (Mainfroid *et al.*, 1996). En el caso de las proteínas ancestrales, ninguno de los cambios únicos observados en la proteína inferida por máxima verosimilitud, se encuentra en la interfaz de monómeros. Aun así, observamos un fenotipo tan distinto en la estabilidad a la dilución que pareciera que se modifican las interacciones en la interfaz. Es posible que los cambios que se analizaron estén irrumpiendo la estabilidad del monómero y al disociarse éste se agregue, impidiendo que exista un equilibrio dímero-monómero.

Otra posible explicación a la variabilidad en la estabilidad a la dilución para las proteínas ancestrales es que esta variabilidad es intrínseca a las TIMs de diversas especies reportadas (Tabla 13). Se ha encontrado que la concentración intracelular de la enzima TIM se encuentra en un rango de  $1 \times 10^{-5}$  M y  $1 \times 10^{-6}$  M (Albe *et al.*, 1990). Estos valores son mucho más altos, al menos dos órdenes de magnitud, que la constante de disociación de las enzimas más sensibles a la dilución, las TIMs bacterianas. Por lo que en cualquier caso la afinidad entre monómeros es lo suficientemente alta para que la enzima se encuentre en su forma dimérica *in vivo*.

Por otro lado, si determinamos el  $\Delta G_{\text{diss}}^{\text{APP}}$  de las cuatro enzimas ancestrales, observamos que varía por apenas 3 kcal/mol entre la proteína obtenida por máxima verosimilitud y la alternativa obtenida por inferencia bayesiana (Tabla 11). Puede que esta pequeña diferencia sea la razón por la que no se encuentran cambios estructurales significativos en los sitios de la interfaz, aunque si estén presentes. Análisis comparativos en la estructura entre las cuatro proteínas ancestrales podrían determinar si existen cambios en la interfaz, aunque no se presenten cambios drásticos en los residuos.

Al comparar los resultados de este trabajo con la TIM63, observamos que poseen las mismas señales de estructura secundaria y terciaria y son capaces de formar dímeros. Las diferencias son notables en los demás parámetros. Observamos que la TIM63 tiene una eficiencia catalítica un orden de magnitud menor a la de cualquiera de las TIM de esta reconstrucción, que al igual que las proteínas de esta reconstrucción, ya no es considerada limitada por difusión. También, observamos un aumento en las temperaturas medias de desnaturalización de las proteínas de esta reconstrucción de hasta 15°C (comparadas con el promedio). Observamos diferencias más notables en la estabilidad a la dilución, ya que el valor de la constante de disociación de la TIM63 es muy diferente con tres de las cuatro proteínas ancestrales (hasta dos órdenes de magnitud).

Sin embargo, si se compara con la proteína inferida por máxima verosimilitud las constantes son muy parecidas.

Hay que notar que tanto TIM63 como MLTIM, se infirieron bajo el mismo método estadístico para la determinación de la secuencia ancestral. De alguna manera, la incorporación de los aminoácidos que el método considera más verosímiles está afectando la estabilidad a la dilución de este nodo, ya que los nodos más recientes de la anterior reconstrucción son menos sensibles a la dilución, al menos por un orden de magnitud. Tampoco se ve afectada la estabilidad a la dilución por la incorporación de sitios ambiguos.

Finalmente, se muestra un análisis con respecto a la variabilidad en los valores de  $T_m$  y las constantes de disociación desde el punto de vista de la reconstrucción. Hart y colaboradores (Hart *et al.*, 2014) realizaron un trabajo donde evaluaron la  $T_m$  de los ancestros de la ribonucleasa H1 tanto en ancestros de organismos termófilos como mesófilos, así como los mecanismos por los que se modifica la estabilidad de la proteína, es decir, cambiando los valores de  $\Delta C_p$ ,  $\Delta H$  y  $T_s$ . En el trabajo se observa un cambio gradual en la  $T_m$ , aumentando en la rama termófila o disminuyendo en la rama mesófila, a partir del ancestro común de ambos. Sin embargo, los valores que modifican la estabilidad sufren cambios drásticos en los ancestros sin patrón alguno. Con estas evidencias, los autores sugieren que las proteínas disponen de diferentes estrategias para modificar su estabilidad. Es importante señalar que realizaron una caracterización de la incertidumbre de los valores de las proteínas ancestrales. Una caracterización por simulación se realizó sobre un pseudo conjunto de datos generado por *bootstrap* sobre los ancestros, que muestran un gran intervalo para los valores que modifican la estabilidad ( $\Delta C_p$ ,  $\Delta H$  y  $T_s$ ), no así para el rango de valores de  $T_m$  (5 °C), al caracterizadas experimentalmente secuencias ancestrales alternativas.

Con base en el artículo anterior y los distintos mecanismos de la ribonucleasa para su estabilización, nos damos cuenta de que hay características que no varían mucho sin importar el ancestro que se reconstruye o la incertidumbre de los aminoácidos ambiguos ( $T_m$ ) mientras que hay valores que se modifican drásticamente a lo largo de las ramas o no son tan robustos en el mismo ancestro ( $\Delta C_p$ ,  $\Delta H$  y  $T_s$ ). Es decir, que mantener un parámetro como el  $\Delta G$  y la  $T_m$  puede ser más relevante para la proteína que la manera en el que se consigue dicho parámetro.

Con los estudios de Hart y colaboradores (2014) se podría suponer que la constante de disociación no es robusta porque su modificación en conjunción con otro parámetro podría estar regulando otra característica de mayor importancia para la célula como se observaba en la estabilidad de las ribonucleasas H1.

La ecuación de desplegamiento de tres estados para un dímero con intermediario monomérico se presenta a continuación:

$$\Delta G_D = \Delta G_{Diss} + 2\Delta G_{Unf} \quad (7)$$

Donde  $\Delta G_D$  es el cambio en la energía libre de desplegamiento de la proteína,  $\Delta G_{Diss}$  es el cambio en la energía libre de disociación de los monómeros y  $\Delta G_{Unf}$  se refiere al cambio en la energía



libre del desplegamiento únicamente de monómero. Estudios posteriores se deben realizar para determinar si, a pesar de que  $\Delta G_{\text{Diss}}$  este cambiando,  $\Delta G_{\text{Unf}}$  se mantenga constante como observaron Mainfroid y colaboradores (1996). O si el parámetro que se esté manteniendo sea el cambio en la energía libre total.

Desde el punto de vista estructural es posible que la variabilidad de las propiedades en las cuatro proteínas ancestrales sobre un mismo nodo nos esté hablando sobre la tolerancia de estas propiedades a las mutaciones. Puede que no estemos observando la capacidad de la técnica de reconstrucción ancestral para proponer características más o menos robustas, sino un reflejo de la robustez de la proteína a las mutaciones que afecten su plegamiento y por lo tanto su función.

Por ejemplo, los sitios asociados a la catálisis como los aminoácidos catalíticos y aminoácidos aledaños se encuentran altamente conservados. Mientras que las regiones asociadas a la oligomerización no lo están tanto (Figura 8). Las constantes catalíticas de catálisis de las cuatro proteínas ancestrales son semejantes, mientras que existe más variación en cuanto a los valores de la constante de disociación. Es más probable que existan cambios en esta zona que afecten la estabilidad entre los monómeros debido a los aminoácidos ambiguos o a utilizar diferentes metodologías y que estos cambios afecten, únicamente por la incertidumbre asociada al alineamiento.

Se determinaron aquí propiedades que, sin importar la metodología de reconstrucción y los errores asociados a estos, son robustos (lo observamos en este trabajo con los porcentajes de estructura secundaria y terciaria, los parámetros catalíticos y el estado oligomérico). Observamos también parámetros que cambian dependiendo de la metodología utilizada para la inferencia (temperatura de desnaturalización, y  $\Delta H$ ). Finalmente, aquellos que son muy variables, tanto al utilizar diferentes metodologías de reconstrucción como al incorporar sitios ambiguos (intensidad de fluorescencia y la constante de disociación).

La determinación de los parámetros que se mantuvieron sin importar la incertidumbre de la reconstrucción puede proporcionar información más confiable sobre las reconstrucciones ancestrales y aumentar la confianza sobre las características de proteínas ancestrales y las inferencias que se realizan sobre ellas en el campo de la evolución molecular. Pero, aquellas características variables también nos pueden proporcionar información sobre las variaciones que existen entre las diferentes secuencias utilizadas y en el motivo por el que existe tanta diversidad.

## Conclusiones

- Las propiedades espectroscópicas, estado oligomérico y parámetros catalíticos, son semejantes entre las cuatro reconstrucciones, indicando que estas propiedades dentro de las proteínas ancestrales son robustas, sin importar el método de reconstrucción o la incorporación de elecciones subóptimas en sitios ambiguos.
- La  $T_m$  y  $\Delta H$  son diferentes dependiendo de la metodología de reconstrucción con la que se obtuvo la proteína ancestral. Seguramente esta diferencia se deba a que el método de máxima verosimilitud descarta aminoácidos perjudiciales a la estabilidad estructural por ser menos frecuentes.

- La constante de estabilidad a la dilución es el parámetro de más variación entre las cuatro proteínas ancestrales. Existen múltiples razones que pueden estar afectando este valor, desde las variaciones propias en la familia de la TIM, hasta las metodologías de reconstrucción o parámetros relacionados a la constante.

## Perspectivas

Una caracterización más profunda de los nodos ancestrales se podría llevar a cabo, ya que aún hay muchas características por determinar y verificar la robustez de estos parámetros de las cuatro reconstrucciones. Entre los experimentos que se pueden realizar está la desnaturalización por agente caotrópico, ya sea urea o cloruro de guanidinio, seguida por DC o por fluorescencia. El seguimiento de la desnaturalización por este método permitirá determinar el  $\Delta G$  del plegamiento de la proteína y el  $\Delta G$  de desnaturalización de los monómeros, además de determinar si las cuatro proteínas poseen el mismo número de intermediarios de despegamiento o si poseen la misma ruta de plegamiento.

Para el caso de la proteína ancestral obtenida por el método de máxima verosimilitud, que sufre más por la dilución, se podría realizar una técnica modificada de calorimetría de titulación isotérmica (ITC) para determinar el calor asociado a la disociación de los monómeros (Schulte-Sasse *et al.*, 2019). Aunque las otras 3 proteínas ancestrales del mismo nodo son más estables a la dilución, se podría intentar realizar este mismo experimento.

Así mismo, se puede usar la técnica de ITC clásica, utilizando en análogo de ligando PGH y determinar la termodinámica asociada a la unión del sustrato, ya que, a pesar de no observar variaciones significativas en la constante de unión, aun así, se encuentra una ligera variación. Determinar las energías asociadas, nos podría indicar con más detalle la manera en que se une el ligando al sitio activo y si es afectada por los cambios en aminoácidos en las diferentes reconstrucciones.

Para los experimentos de calorimetría, únicamente se determinó la irreversibilidad de la desnaturalización las proteínas ancestrales, sin embargo, se puede recurrir a realizar más experimentos para determinar los parámetros aparentes. Por ejemplo, se pueden realizar experimentos de calorimetría a diferentes velocidades de barrido y determinar si el despegamiento se encuentra bajo control cinético.

También, se lograron colocar pruebas de cristalización de las cuatro proteínas ancestrales que resultaron satisfactorias. Sin embargo, para algunos cristales aún no se ha realizado la difracción o se encuentran en proceso del remplazo molecular para la obtención del modelo. La obtención de la estructura tridimensional real podría servir para determinar si de verdad las suposiciones que se hacen con base en el modelo son ciertas y se observan interacciones o no que podría explicar las diferencias tan grandes en las constantes de disociación. También prestando atención a las moléculas de agua, que se ha observado influye mucho en la estabilización de los monómeros.

## Referencias

- Akanuma, S., Nakajima, Y., Yokobori, S., Kimura, M., Nemoto, N., Mase, T., & Yamagishi, A. (2013). Experimental evidence for the thermophilicity of ancestral life. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), 11067-11072.
- Albe, K. R., Butler, M. H., & Wright, B. E. (1990). Cellular concentrations of enzymes and their substrates. *Journal of Theoretical Biology*, 143, 163-195
- Alberty, R. A., & Hammes, G. G. (1958). Application of the theory of diffusion-controlled reactions to enzyme kinetics. *The Journal of Physical Chemistry*, 62(2), 154-159.
- Albery, W. J., & Knowles, J. R. (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*, 15(25), 5631-5640.
- Alvarez, M., Wouters, J., Maes, D., Mainfroid, V., Rentier-Delrue, F., Wyns, L., & Martial, J. A. (1999). Lys13 plays a crucial role in the functional adaptation of the thermophilic triose-phosphate isomerase from *Bacillus stearotherophilus* to high temperatures. *The Journal of Biological Chemistry*, 274(27), 19181-7.
- Alvarez, M., Zeelen, J. P., Mainfroid, V., Rentier-Delrue, F., Martial, J. A., Wyns, L., Wierenga, R.K., & Maes, D. (1998). Triose-phosphate Isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. *Journal of Biological Chemistry*, 273(4), 2199-2206.
- Beaucamp, N., Hofmann, A., Kellerer, B., & Jaenicke, R. (2008). Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: Design and characterization of the separate triosephosphate isomerase. *Protein Science*, 6(10), 2159-
- Benítez-Cardoza, C. G., Rojo-Domínguez, A., & Hernández-Arana, A. (2001). Temperature-induced denaturation and renaturation of triosephosphate isomerase from *Saccharomyces cerevisiae*: Evidence of Dimerization Coupled to Refolding of the Thermally Unfolded Protein. *Biochemistry*, 40(30), 9049-9058.
- Benner, S. A., Sassi, S. O., & Gaucher, E. A. (2007). Molecular paleoscience: Systems biology from the past. *Advances in Enzymology. Related Areas of Molecular Biology*, 75, 1-132.
- Boteva, R., Zlateva, T., Dorovska-Taran, V., Visser, A. J., Tsanev, R., & Salvato, B. (1996). Dissociation Equilibrium of Human Recombinant Interferon  $\gamma$ . *Biochemistry*, 35(47), 14825-14830.
- Bohm, G., Muhr, R., & Jaenicke, R. (1992). CDNN: quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Engineering*, 5(3), 191-195.
- Chu, C., Lai, Y., Huang, H., & Sun, Y. (2008). Kinetic and structural properties of triosephosphate isomerase from *Helicobacter pylori*. *Proteins: Structure, Function, and Bioinformatics*, 71(1), 396-406.
- Ciccarelli, F. D. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311(5765), 1283-1287.
- Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J., & Thornton, J. W. (2016). Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Molecular biology and evolution*. 34(2), 247-261

- Farber, G. K., & Petsko, G. A. (1990) The evolution of  $\alpha/\beta$  barrel enzymes. *Trends in Biochemistry Sciences*, 15, 228–234
- Field, S. F., & Matz, M. V. (2009). Retracing Evolution of red fluorescence in GFP-like proteins from Faviina corals. *Molecular Biology and Evolution*, 27(2), 225-233.
- Figuerola-Angulo, E. E., Estrella-Hernández, P., Salgado-Lugo, H., Ochoa-Leyva, A., Puyou, A. G., Campos, S. S., Montero-Moran, G., Ortega-López, J., Saab-Rincón, G., Arroyo, R., Benítez-Cardoza, C. G., & Brieba, L. G. (2012). Cellular and biochemical characterization of two closely related triosephosphate isomerases from *Trichomonas vaginalis*. *Parasitology*, 139(13), 1729-1738.
- Gaucher, E. A., Govindarajan, S., & Ganesh, O. K. (2008). Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature*, 451(7179), 704-707.
- Gayathri, P., Banerjee, M., Vijayalakshmi, A., Azeez, S., Balaram, H., Balaram, P., & Murthy, M. R. N. (2007). Structure of triosephosphate isomerase (TIM) from *Methanocaldococcus jannaschii*. *Acta Crystallographica*, D63, 206-220.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321.
- Gumulya, Y., & Gillam, E. M. (2016). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The retro approach to protein engineering. *Biochemical Journal*, 474(1), 1-19.
- Hall, B. G. (2011). *Phylogenetic trees made easy: A how-to manual for molecular biologists* (Fourth ed.). Sunderland: Sinauer Associates.
- Hart, K. M., Harms, M. J., Schmidt, B. H., Elya, C., Thornton, J. W., & Marqusee, S. (2014). Thermodynamic system drift in protein evolution. *PLoS Biology*, 12(11).
- Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics*, 22(23), 2971-2972.
- Heinig, M., & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic acids research*, 32(Web Server issue), W500-2.
- Hernández-Alcántara, G., Garza-Ramos, G., Hernández, G. M., Gómez-Puyou, A., & Pérez-Montfort, R. (2002). Catalysis and stability of triosephosphate isomerase from *Trypanosoma brucei* with different residues at position 14 of the dimer interface. Characterization of a catalytically competent monomeric enzyme. *Biochemistry*, 41(13), 4230-4238.
- Hillis, D. M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology*, 47, 3–8
- Hochberg, G. K. A., & Thornton, J. W. (2017). Reconstructing ancient proteins to understand the causes of structure and function. *Annual Review of Biophysics*, 46(1), 247–269.
- Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680-682

- Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 17, 754-755.
- Huelsenbeck, J. P., & Bollback, J. P. (2001). Empirical and hierarchical bayesian estimation of ancestral states. *Systematic Biology*, 50(3), 351-366.
- Jiménez, L., Fernández-Velasco, D. A., Willms, K., & Landa, A. (2003). A comparative study of biochemical and immunological properties of triosephosphate isomerase from *Taenia solium* and *Sus scrofa*. *Journal of Parasitology*, 89(2), 209-214.
- Jiménez, L., Vibanco-Pérez, N., Navarro, L., & Landa, A. (2000). Cloning, expression and characterization of a recombinant triosephosphate isomerase from *Taenia solium*. *International Journal for Parasitology*, 30(9), 1007-1012.
- Joubert, F., Neitz, A., & Louw, A. (2001). Structure-based inhibitor screening: A family of sulfonated dye inhibitors for malaria parasite triosephosphate isomerase. *Proteins: Structure, Function, and Genetics*, 45(2), 136-143.
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059-3066.
- Kim, J. (1998). Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Systematic Biology*, 47, 43-60.
- Knobeloch, D., Schmidt, A., Scheerer, P., Krauss, N., Wessner, H., Scholz, C., Kütner, G., von Rintelen, T., Wessel, A., & Höhne, W. (2010). A coleopteran triosephosphate isomerase: X-ray structure and phylogenetic impact of insect sequences. *Insect Molecular Biology*, 19(1), 35-48.
- Krietsch, W. K., Pentchev, P. G., Klingenburg, H., Hofstatter, T., & Bucher, T. (1970). The isolation and crystallization of yeast and rabbit liver triose phosphate isomerase and a comparative characterization with the rabbit muscle enzyme. *European Journal of Biochemistry*, 14(2), 289-300.
- Lakowicz, J. R. (1999). *Principles of fluorescence spectroscopy* (Third Edition ed.). New York: Kluwer Academic, 540-545, 556-667.
- Lambeir, A., Backmann, J., Ruiz-Sanz, J., Filimonov, V., Nielsen, J. E., Kursula, I., Norledge, B. V., & Wierenga, R. K. (2000). The ionization of a buried glutamic acid is thermodynamically linked to the stability of *Leishmania mexicana* triose phosphate isomerase. *European Journal of Biochemistry*, 267(9), 2516-2524.
- Lambeir, A., Opperdoes, F. R., & Wierenga, R. K. (1987). Kinetic properties of triose-phosphate isomerase from *Trypanosoma brucei*. A comparison with the rabbit muscle and yeast enzymes. *European Journal of Biochemistry*, 168(1), 69-74.
- Landa, A., Rojo-Dominguez, A., Jiménez, L., & Fernández-Velasco, D. A. (1997). Sequencing, expression and properties of triosephosphate isomerase from *Entamoeba histolytica*. *European Journal of Biochemistry*, 247(1), 348-355.
- Le, S., & Gascuel, O. (2008). An improved general amino-acid replacement matrix. *Molecular Biology and Evolution*, 25, 1307-1320

- Lefort, V., Longueville, J. E., Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, 34(9), 2422-2424.
- Loewenthal, R., Sancho, J., & Fersht, A. R. (1991). Fluorescence spectrum of barnase: contributions of three tryptophan residues and a histidine-related pH dependence. *Biochemistry*, 30 (27), 6775-6779
- López-Castillo, M., Jimenez-Sandoval, P., Lara-González, S., Briebe, L., & Baruch, N. (2016). Crystal structure of *Arabidopsis thaliana* cytosolic triose phosphate isomerase.
- López-Velázquez, G., Molina-Ortiz, D., Cabrera, N., Hernández-Alcántara, G., Peon-Peralta, J., Yépez-Mulia, L., & Reyes-Vivas, H. (2004). An unusual triosephosphate isomerase from the early divergent eukaryote *Giardia lamblia*. *Proteins: Structure, Function, and Bioinformatics*, 55(4), 824-834.
- Lutzoni, F., Kauff, F., Cox, C. J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T. Y., Baloch, E., Grube, M., Reeb, V., Hofstetter, V., Schoch, C., Arnold, A. E., Miadlikowska, J., Spatafora, J., Johnson, D., Hambleton, S., Crockett, M., Shoemaker, R., Sung, G., Lücking, R., Lumbsch, T., O'Donnell, K., Binder, M., Diederich, P., Ertz, D., Gueidan, C., Hansen, K., Harris, R. C., Hosaka, K., Lim, Y., Matheny, B., Nishida, H., Pfister, D., Rogers, J., Rossman, A., Schmitt, I., Sipman, H., Stone, J., Sugiyama, J., Yahr, R., & Vilgalys, R. (2004), Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits . *American Journal of Botany*, 91(10), 1446-1480.
- Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Thi, M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R., & Wierenga, R. K. (1999). The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. *Proteins*, 7(3), 441-53
- Mainfroid, V., Terpstra, P., Beauregard, M., Frère, J., Mande, S. C., Hol, W. G., Martial, J. A., & Goraj, K. (1996). Three hTIM mutants that provide new insights on why TIM is a dimer. *Journal of Molecular Biology*, 257(2), 441-456.
- Mande, S., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A., & Hol, W. G. (1995). Crystal structure of recombinant human triosephosphate isomerase at 2.8 Angstroms resolution. Triosephosphate isomerase related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Science*, 3(5), 810-821.
- Mathur, D., Malik, G., & Garg, L. C. (2006). Biochemical and functional characterization of triosephosphate isomerase from *Mycobacterium tuberculosis* H37Rv. *FEMS Microbiology Letters*, 263(2), 229-235.
- Merritt, T. J., & Quattro, J. M. (2001). Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics*, 159(2), 689-697.
- Miller, M.A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 1-8.
- Morgan, C. J., Wilkins, D. K., Smith, L. J., Kawata, Y., & Dobson, C. M. (2000). A compact monomeric intermediate identified by NMR in the denaturation of dimeric triose phosphate isomerase. *Journal of Molecular Biology*, 300(1), 11-16.

Muñiz-Trejo Ricardo. (2017). Producción y caracterización fisicoquímica comparativa de la triosafosfato isomerasa de tres arqueas del orden *Methanococcales*. Tesis de Licenciatura. Universidad Nacional Autónoma de México. Ciudad de México.

Nájera, H., Costas, M., & Fernández-Velasco, D. A. (2003) Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme. *Biochemical Journal*, 370, 785–92.

Newton, M. S., Arcus, V. L., & Patrick, W. M. (2015). Rapid bursts and slow declines: On the possible evolutionary trajectories of enzymes. *Journal of The Royal Society Interface*, 12(107), 20150036

Oneill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., Zimba, P. V., Henrissat, B., & Field, R. A. (2015). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems*, 11(10), 2808-2820.

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science*, 317(5844), 1544-1548.

Pauling, L., Zuckerkandl, E., Henriksen, T., & Löfstad, R. (1963). Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life. *Acta Chemica Scandinavica*, 17 suppl., 9-16.

Peimbert, M., Domínguez-Ramírez, L., & Fernández-Velasco, D. A. (2008). Hydrophobic repacking of the dimer interface of triosephosphate isomerase by in silico design and directed evolution. *Biochemistry*, 47(20), 5556–5564.

Plaut, B., & Knowles, J. R. (1972). pH-dependence of the triose phosphate isomerase reaction. *Biochemical Journal*, 129(2), 311-320.

Pollock, D. D., Zwickl, D. J., McGuire, J. A. & Hillis, D. M. (2002). Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology*, 51, 664–671.

Putman, S. J., Coulson, A. F., Farley, I. R., Riddleston, B., & Knowles, J. R. (1972). Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochemical Journal*, 129(2), 301-310.

Reyes-Vivas, H., Diaz, A., Peon, J., Mendoza-Hernández, G., Hernández-Alcántara, G., De La Mora-De La Mora, I., Enfriquez-Flores, S., Dominguez-Ramírez, L., & López-Velázquez, G. (2007). Disulfide bridges in the mesophilic triosephosphate isomerase from *Giardia lamblia* are related to oligomerization and activity. *Journal of Molecular Biology*, 365(3), 752-763.

Reyes-Vivas, H., Martínez-Martínez, E., Mendoza-Hernández, G., López-Velázquez, G., Pérez-Montfort, R., Gómez-Puyou, M. T., & Gómez-Puyou, A. (2002). Susceptibility to proteolysis of triosephosphate isomerase from two pathogenic parasites: Characterization of an enzyme with an intact and a nicked monomer. *Proteins: Structure, Function, and Bioinformatics*, 48(3), 580-590.

Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A. & Sanchez-Ruiz, J.M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian  $\beta$ -lactamases. *Journal of the American Chemical Society*, 135, 2899–2902

Rodríguez-Almazan, C., Arreola, R., Rodríguez-Larrea, D., Aguirre-Lopez, B., Tuena de Gómez-Puyou, M., Pérez-Montfort, R., Costas, M., Gómez-Puyou, A., & Torres-Larios, A. (2008). Structural basis of

- human triosephosphate isomerase deficiency: mutation E104D is related to alterations of a conserved water network at the dimer interface. *Journal of Biological Chemistry*, 283(34), 23254-23263.
- Romero-Romero, S., Becerril-Sesín, L. A., Costas, M., Rodríguez-Romero, A., & Fernández-Velasco, D. A. (2018). Structure and conformational stability of the triosephosphate isomerase from *Zea mays*. Comparison with the chemical unfolding pathways of other eukaryotic TIMs. *Archives of Biochemistry and Biophysics*, 658, 66-76.
- Romero-Romero, S., Costas, M., Rodríguez-Romero, A., & Fernández-Velasco, D. A. (2016). Reversibility and two state behaviour in the thermal unfolding of oligomeric TIM barrel proteins. *Physical Chemistry Chemical Physics*, 18(15), 10647-10647.
- Ronquist, F. & J. P. Huelsenbeck. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572-1574.
- Rosenberg, M. S., & Kumar, S. (2001). Incomplete taxon sampling is not a problem for phylogenetic inference. *Proceedings of the National Academy of Sciences*, 98, 10751-10756.
- Salemi, M., Vandamme, A. M., & Lemey, P. (2009). *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. Cambridge, UK: Cambridge University Press.
- Schliebs, W., Thanki, N., Jaenicke, R., & Wierenga, R. K. (1997). A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability. *Biochemistry*, 36(32), 9655-9662.
- Schulte-Sasse Jiménez, M. (2015). Evolución de la estabilidad en la enzima triosafosfato isomerasa de eucariontes. (Tesis de maestría). UNAM, Facultad de Medicina. México.
- Schulte-Sasse, M., Pardo-Ávila, F., Pulido-Mayoral, N. O., Vázquez-Lobo, A., Costas, M., García-Hernández, E., Rodríguez-Romero, A. & Fernández-Velasco, D. A. (2019). Structural, thermodynamic and catalytic characterization of an ancestral triosephosphate isomerase reveal early evolutionary coupling between monomer association and function. *The FEBS Journal*, 1-19. Manuscrito aceptado.
- Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43 (Web Server issue), W7-W14.
- Sharma, P., & Guptasarma, P. (2015). 'Super-perfect' enzymes: Structural stabilities and activities of recombinant triose phosphate isomerases from *Pyrococcus furiosus* and *Thermococcus onnurineus* produced in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 460(3), 753-758.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539-539.
- Thornton, J. W. (2004). Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nature Reviews Genetics*, 5(5), 366-375.
- Trudeau, D. L., Kaltenbach, M., & Tawfik, D. S. (2016). On the potential origins of the high stability of reconstructed ancestral proteins. *Molecular Biology and Evolution*, 33(10), 2633-2641.



- Ugalde, J. A. (2004). Evolution of coral pigments recreated. *Science*, 305(5689), 1433-1433.
- Walden, H., Bell, G. S., Russell, R. J., Siebers, B., Hensel, R., & Taylor, G. L. (2001). Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. *Journal of Molecular Biology*, 306, 745-757.
- Williams, P. D., Pollock, D. D., Blackburne, B. P., & Goldstein, R. A. (2006). Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Computational Biology*, 2, 5573-5584.
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13(5), 555-556.
- Yang, Z., Kumar, S., & Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, 141(4), 1641-1650.
- Zabori, S., Rudolph, R., & Jaenicke, R. (1980). Folding and association of triose phosphate isomerase from rabbit muscle. *Zeitschrift für Naturforschung C*, 35(11-12), 999-1004.
- Zaffagnini, M., Michelet, L., Sciabolini, C., Di Giacinto, N., Morisse, S., Marchand, C. H., Trost, P., Fermani, S., & Lemaire, S. D. (2014). High-resolution crystal structure and redox properties of chloroplastic triosephosphate isomerase from *Chlamydomonas reinhardtii*. *Molecular Plant* 7, 101-120.
- Zhang, J. & Nei, M. (1997). Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *Journal of Molecular Evolution*, 44 (Suppl. 1), S139-S146
- Zomosa-Signoret, V., Hernández-Alcántara, G., Reyes-Vivas, H., Martínez-Martínez, E., Garza-Ramos, G., Pérez-Montfort, Tuena de Gómez-Poyou, M., Gómez-Puyou, A. (2003). Control of the reactivation kinetics of homodimeric triosephosphate isomerase from unfolded monomers. *Biochemistry*, 42(11), 3311-3318.

## Anexos

**Anexo 1:** Identificador de secuencias utilizadas para la reconstrucción. Se colocó el reino al que pertenecen. También aparecen las 3 secuencias utilizadas como grupos externos.

|    |  |        |
|----|--|--------|
| 1  | >XP_011811537.1 PREDICTED: triosephosphate isomerase isoform X1 [Colobus angolensis palliatus]                       | Animal |
| 2  | >AAB59511.1 (EC5.3.1.1)[Homosapiens]   | Animal |
| 3  | >gi 308082020 ref NP_001183983.1  [Canis lupus familiaris] PE=1  | Animal |
| 4  | >sp P00939.1 TPIS_RABIT RecName: Full=Triosephosphate isomerase; Short=TIM; AltName: Full=Triose-phosphate isomerase | Animal |
| 5  | >>gi 54855 emb CAA37420.1  [Mus musculus]  | Animal |
| 6  | >gi 80971510 ref NP_001032228.1  [Sus scrofa] PE=2   | Animal |
| 7  | >gi 45382061 ref NP_990782.1  [Gallus gallus] PE=1   | Animal |
| 8  | >XP_014349210.1 PREDICTED: triosephosphate isomerase [Latimeria chalumnae]   | Animal |
| 9  | >gi 166795913 ref NP_001107706.1  [Xenopus (Silurana) tropicalis] PE=2   | Animal |
| 10 | >Q90XG0.1 [Danio rerio] B  | Animal |
| 11 | >XP_004078032.1 PREDICTED: triosephosphate isomerase [Oryzias latipes]   | Animal |
| 12 | >NP_001133174.1 triosephosphate isomerase B [Salmo salar]  | Animal |
| 13 | >EDO41453.1 predicted protein [Nematostella vectensis]   | Animal |
| 14 | >XP_015773003.1 PREDICTED: triosephosphate isomerase-like [Acropora digitifera]                                      | Animal |
| 15 | >XP_017967287.1 PREDICTED: triosephosphate isomerase isoform X2 [Drosophila navojoa]                                 | Animal |
| 16 | >gi 10946 emb CAA40804.1  [Drosophila melanogaster] PE=1   | Animal |
| 17 | >XP_017482926.1 PREDICTED: triosephosphate isomerase [Rhagoletis zephyria]   | Animal |
| 18 | >XP_013107065.1 PREDICTED: triosephosphate isomerase [Stomoxys calcitrans]   | Animal |
| 19 | >XP_019556133.1 PREDICTED: triosephosphate isomerase [Aedes albopictus]  | Animal |
| 20 | >gi 156691 gb AAA73976.1  [Culex tarsalis] PE=1  | Animal |
| 21 | >AHB50481.1 triose-phosphate isomerase [Mayetiola destructor]  | Animal |
| 22 | >tr D2A042 D2A042_TRICA OS=Tribolium castaneum GN=GLEAN_07346 PE=3 SV=1  | Animal |
| 23 | >gi 22090453 emb CAD43178.1  [Tenebrio molitor] PE=1   | Animal |
| 24 | >XP_017779943.1 PREDICTED: triosephosphate isomerase [Nicrophorus vespilloides]                                      | Animal |
| 25 | >AEE62729.1 unknown [Dendroctonus ponderosae]  | Animal |
| 26 | >XP_018327468.1 PREDICTED: triosephosphate isomerase [Agrilus planipennis]   | Animal |
| 27 | >JAT99180.1 putative triosephosphate isomerase [Amblyomma aureolatum]  | Animal |
| 28 | >JAA59214.1 Putative triosephosphate isomerase [Rhipicephalus pulchellus]  | Animal |
| 29 | >gi 118175962 gb ABK76308.1  [Rhipicephalus microplus] PE=1  | Animal |
| 30 | >JAP72957.1 putative triosephosphate isomerase [Ixodes ricinus]  | Animal |
| 31 | >KPI96167.1 [Papilio xuthus]   | Animal |

|    |   |        |
|----|---|--------|
| 32 | >gi 52222824 gb AAU34185.1  [Bombyx mori] PE=1  | Animal |
| 33 | >JAA81080.1 [Pararge aegeria]   | Animal |
| 34 | >XP_011548378.1 PREDICTED: triosephosphate isomerase [Plutella xylostella]                | Animal |
| 35 | >gi 357631080 gb EHJ78785.1  [Danaus plexippus] PE=3                                      | Animal |
| 36 | >tr A0A0L7KX23 A0A0L7KX23_9NEOP OS=Operophtera brumata GN=OBRU01_19537 PE=3 SV=1          | Animal |
| 37 | >gi 308512821 gb ADO33064.1  [Biston betularia] PE=2                                      | Animal |
| 38 | >gi 506968283 gb AGM32485.1  [Coptotermes formosanus] PE=2                                | Animal |
| 39 | >tr E0VJK8 E0VJK8_PEDHC OS=Pediculus humanus subsp. corporis GN=Phum_PHUM247280 PE=3 SV=1 | Animal |
| 40 | >gi 442565872 gb AGC56216.1  [Dermatophagoides farinae]                                   | Animal |
| 41 | >XP_015781890.1 PREDICTED: triosephosphate isomerase B-like [Tetranychus urticae]         | Animal |
| 42 | >XP_013790695.1 PREDICTED: triosephosphate isomerase-like [Limulus polyphemus]            | Animal |
| 43 | >KZS06051.1 Triosephosphate isomerase [Daphnia magna]                                     | Animal |
| 44 | >XP_011867461.1 PREDICTED: triosephosphate isomerase [Vollenhovia emeryi]                 | Animal |
| 45 | >XP_014215284.1 PREDICTED: triosephosphate isomerase isoform X2 [Copidosoma floridanum]   | Animal |
| 46 | >XP_011502456.1 PREDICTED: triosephosphate isomerase [Ceratosolen solmsi marchali]        | Animal |
| 47 | >gi 238477329 gb ACR43476.1  [Crangon crangon] PE=2                                       | Animal |
| 48 | >XP_018024838.1 PREDICTED: triosephosphate isomerase B-like [Hyaella azteca]              | Animal |
| 49 | >gi 225714114 gb ACO12903.1  [Lepeophtheirus salmonis]                                    | Animal |
| 50 | >ELT97198.1 hypothetical protein CAPTEDRAFT_172464 [Capitella teleta]                     | Animal |
| 51 | >XP_013407822.1 PREDICTED: triosephosphate isomerase B-like [Lingula anatina]             | Animal |
| 52 | >NP_001171739.1 triosephosphate isomerase 1 [Saccoglossus kowalevskii]                    | Animal |
| 53 | >gi 946725746 gb ALM30773.1  [Amphioctopus fangsiao]                                      | Animal |
| 54 | >gi 167643881 gb ABZ90135.1  [Dugesia japonica] PE=2                                      | Animal |
| 55 | >gi 333449422 gb AEF33397.1  [Crassostrea ariakensis] PE=2                                | Animal |
| 56 | >XP_002117956.1 hypothetical protein TRIADDRAFT_38509 [Trichoplax adhaerens]              | Animal |
| 57 | >ODM96190.1 Triosephosphate isomerase B [Orchesella cincta]                               | Animal |
| 58 | >tr H2ZAA8 H2ZAA8_CIOSA [Ciona savignyi] PE=3   | Animal |
| 59 | >JAP02759.1 putative triosephosphate isomerase [Triatoma dimidiata]                       | Animal |
| 60 | >XP_009025315.1 hypothetical protein HELRODRAFT_185065 [Helobdella robusta]               | Animal |
| 61 | >EGT48267.1 hypothetical protein CAEBREN_21749 [Caenorhabditis brenneri]                  | Animal |
| 62 | >gi 1036784 gb AAA79846.1  [Caenorhabditis elegans] PE=1                                  | Animal |
| 63 | >gi 313509547 gb ADR66027.1  [Haemonchus contortus] PE=2                                  | Animal |
| 64 | >gi 324523262 gb ADY48215.1  [Ascaris suum] PE=2  | Animal |

|     |  |        |
|-----|--|--------|
| 65  | >AAZ57433.1 triose-phosphate isomerase [ <i>Schistosoma turkestanicum</i> ]                | Animal |
| 66  | >sp P48501 TPIS_SCHMA [ <i>Schistosoma mansoni</i> ] PE=1                                  | Animal |
| 67  | >gi 1255223 gb AAC47393.1  [ <i>Schistosoma japonicum</i> ] PE=2                           | Animal |
| 68  | >KER21742.1 hypothetical protein T265_10017 [ <i>Opisthorchis viverrini</i> ]              | Animal |
| 69  | >gi 480328213 gb AGJ83762.1  [ <i>Fasciola hepatica</i> ]                                  | Animal |
| 70  | >gi 358332321 dbj GAA50994.1  [ <i>Clonorchis sinensis</i> ]                               | Animal |
| 71  | >XP_005106755.1 PREDICTED: triosephosphate isomerase-like [ <i>Aplysia californica</i> ]   | Animal |
| 72  | >GAV05401.1 hypothetical protein RvY_15542 [ <i>Ramazzottius varieornatus</i> ]            | Animal |
| 73  | >JAQ04232.1 Triosephosphate isomerase [ <i>Lygus hesperus</i> ]                            | Animal |
| 74  | >gi 1046537011 gb OCK35149.1  [ <i>Taeniasaginata</i> ]                                    | Animal |
| 75  | >sp Q9GTX8 TPIS_TAESO [ <i>Taenia solium</i> ] PE=2  | Animal |
| 76  | >AFX72997.1 Triosephosphate isomerase [ <i>Spirometra erinaceieuropaei</i> ]               | Animal |
| 77  | >KRZ25238.1 Triosephosphate isomerase [ <i>Trichinella pseudospiralis</i> ]                | Animal |
| 78  | >gi 669220680 emb CDW58011.1  [ <i>Trichuristrichiura</i> ]                                | Animal |
| 79  | >gi 184186117 ref NP_001116981.1  [ <i>Strongylocentrotus purpuratus</i> ] PE=3            | Animal |
| 80  | >CDG70638.1 <i>Hydra vulgaris</i> Triosephosphate isomerase [ <i>Hydra vulgaris</i> ]      | Animal |
| 81  | >XP_003744176.1 PREDICTED: triosephosphate isomerase [ <i>Metaseiulus occidentalis</i> ]   | Animal |
| 82  | >EYE92977.1 triosephosphate isomerase [ <i>Aspergillus ruber</i> CBS 135680]               | Hongo  |
| 83  | >gi 9955867 dbj BAB12233.1  [ <i>Aspergillus oryzae</i> ]                                  | Hongo  |
| 84  | >gi 218723467 gb EED22884.1  [ <i>Talaromyces stipitatus</i> ATCC 10500]                   | Hongo  |
| 85  | >DAA73246.1 TPA_exp: Triosephosphate isomerase [ <i>Trichophyton benhamiae</i> CBS 112371] | Hongo  |
| 86  | >gi 392869932 gb EAS28473.2  [ <i>Coccidioides immitis</i> RS]                             | Hongo  |
| 87  | >gi 226295407 gb EEH50827.1  [ <i>Paracoccidioides brasiliensis</i> Pb18]                  | Hongo  |
| 88  | >gi 226284964 gb EEH40530.1  [ <i>Paracoccidioides</i> sp. 'lutzii' Pb01]                  | Hongo  |
| 89  | >EHK42315.1 triosephosphate isomerase [ <i>Trichoderma atroviride</i> IMI 206040]          | Hongo  |
| 90  | >GAP90047.1 putative triosephosphate isomerase [ <i>Rosellinia necatrix</i> ]              | Hongo  |
| 91  | >EAQ91292.1 hypothetical protein CHGG_03227 [ <i>Chaetomium globosum</i> CBS 148.51]       | Hongo  |
| 92  | >OCW38390.1 triosephosphate isomerase [ <i>Diaporthe helianthi</i> ]                       | Hongo  |
| 93  | >KKF97092.1 Triosephosphate isomerase [ <i>Ceratocystis platani</i> ]                      | Hongo  |
| 94  | >OIW33876.1 triosephosphate isomerase [ <i>Coniochaeta ligniaria</i> NRRL 30616]           | Hongo  |
| 95  | >XP_007733361.1 triosephosphate isomerase [ <i>Capronia epimyces</i> CBS 606.96]           | Hongo  |
| 96  | >ETN44907.1 triosephosphate isomerase [ <i>Cyphellophora europaea</i> CBS 101466]          | Hongo  |
| 97  | >ERF71517.1 Triosephosphate isomerase [ <i>Endocarpon pusillum</i> Z07020]                 | Hongo  |
| 98  | >KKY23942.1 putative triosephosphate isomerase [ <i>Phaeomoniella chlamydospora</i> ]      | Hongo  |
| 99  | >EDU48675.1 triosephosphate isomerase [ <i>Pyrenophora tritici-repentis</i> Pt-1C-BFP]     | Hongo  |
| 100 | >OCL05558.1 Triosephosphate isomerase [ <i>Glonium stellatum</i> ]                         | Hongo  |

|     |  |       |
|-----|--|-------|
| 101 | >KIJ21893.1 triosephosphate isomerase [ <i>Phialocephala scopiformis</i> ]   | Hongo |
| 102 | >EME85715.1 hypothetical protein MYCFIDRAFT_52839 [ <i>Pseudocercospora fijiensis</i> CIRAD86]   | Hongo |
| 103 | >OMP83717.1 Triosephosphate isomerase [ <i>Diplodia seriata</i> ]  | Hongo |
| 104 | >KXL46612.1 hypothetical protein FE78DRAFT_89239 [ <i>Acidomyces richmondensis</i> ]   | Hongo |
| 105 | >KZF26607.1 Triosephosphate isomerase [ <i>Xylona heveae</i> TC161]  | Hongo |
| 106 | >KEQ73708.1 Triosephosphate isomerase [ <i>Aureobasidium namibiae</i> CBS 147.97]  | Hongo |
| 107 | >KJX95887.1 triosephosphate isomerase like protein [ <i>Zymoseptoria brevis</i> ]  | Hongo |
| 108 | >EPQ63062.1 Triose phosphate isomerase abundant glycolytic enzyme mRNA half-life [ <i>Blumeria graminis</i> f. sp. <i>tritici</i> 96224] | Hongo |
| 109 | >OBT98701.1 triosephosphate isomerase [ <i>Pseudogymnoascus verrucosus</i> ]   | Hongo |
| 110 | >ODQ55084.1 triosephosphate isomerase [ <i>Saitoella complicata</i> NRRL Y-17804]  | Hongo |
| 111 | >gi 351646267 gb EHA54127.1  [ <i>Magnaporthe oryzae</i> 70-15]  | Hongo |
| 112 | >KHJ31133.1 putative triosephosphate isomerase [ <i>Erysiphe necator</i> ]   | Hongo |
| 113 | >CCC08460.1 unnamed protein product [ <i>Sordaria macrospora</i> k-hell]   | Hongo |
| 114 | >gi 28920460 gb EAA29827.1  [ <i>Neurospora crassa</i> OR74A]  | Hongo |
| 115 | >EPS44531.1 hypothetical protein H072_1444 [ <i>Dactylellina haptotyla</i> CBS 200.50]   | Hongo |
| 116 | >CEL59915.1 triosephosphate isomerase (TIM) [ <i>Rhizoctonia solani</i> AG-1 IB]   | Hongo |
| 117 | >KIS68731.1 putative triose-phosphate isomerase TPI1 [ <i>Ustilago maydis</i> 521]   | Hongo |
| 118 | >KOS14050.1 triosephosphate isomerase [ <i>Malassezia pachydermatis</i> ]  | Hongo |
| 119 | >CCU98972.1 unnamed protein product [ <i>Malassezia sympodialis</i> ATCC 42132]  | Hongo |
| 120 | >KLT39016.1 Triosephosphate isomerase [ <i>Cutaneotrichosporon oleaginosus</i> ]   | Hongo |
| 121 | >AAA79845.1 triosephosphate isomerase [ <i>Coprinopsis cinerea</i> ]   | Hongo |
| 122 | >gi 169856984 ref XP_001835145.1  [ <i>Coprinopsis cinerea</i> okayama7#130]   | Hongo |
| 123 | >EDR10156.1 triose phosphate isomerase [ <i>Laccaria bicolor</i> S238N-H82]  | Hongo |
| 124 | >KYQ32176.1 Triosephosphate isomerase [ <i>Hypsizygus marmoreus</i> ]  | Hongo |
| 125 | >AFJ04710.1 TPI protein [ <i>Volvariella volvacea</i> ]  | Hongo |
| 126 | >ESK94747.1 triosephosphate isomerase [ <i>Moniliophthora roreri</i> MCA 2997]   | Hongo |
| 127 | >KIJ57895.1 hypothetical protein HYDPIDRAFT_120199 [ <i>Hydnomerulius pinastri</i> MD-312]   | Hongo |
| 128 | >OAX42777.1 Triosephosphate isomerase [ <i>Rhizopogon vinicolor</i> AM-OR11-026]   | Hongo |
| 129 | >EMD36600.1 hypothetical protein CERSUDRAFT_115642 [ <i>Gelatoporia subvermispora</i> B]   | Hongo |
| 130 | >KIO13709.1 hypothetical protein M404DRAFT_993252 [ <i>Pisolithus tinctorius</i> Marx 270]   | Hongo |
| 131 | >KZT70547.1 Triosephosphate isomerase [ <i>Daedalea quercina</i> L-15889]  | Hongo |
| 132 | >EKM53210.1 hypothetical protein PHACADRAFT_259404 [ <i>Phanerochaete carnosa</i> HHB-10118-sp]  | Hongo |
| 133 | >OJT13468.1 Triosephosphate isomerase [ <i>Trametes pubescens</i> ]  | Hongo |

|     |  |       |
|-----|--|-------|
| 134 | >KIY73615.1 Triosephosphate isomerase [Cylindrobasidium torrendii FP15055 ss-10]         | Hongo |
| 135 | >ETW75212.1 hypothetical protein HETIRDRAFT_461226 [Heterobasidion irregulare TC 32-1]   | Hongo |
| 136 | >EIM83693.1 triose phosphate isomerase [Stereum hirsutum FP-91666 SS1]                   | Hongo |
| 137 | >KIJ16202.1 triose-phosphate isomerase [Paxillus involutus ATCC 200175]                  | Hongo |
| 138 | >EIN05592.1 triose phosphate isomerase [Punctularia strigosozonata HHB-11173 SS5]        | Hongo |
| 139 | >KIM74228.1 hypothetical protein PILCRDRAFT_828385 [Piloderma croceum F 1598]            | Hongo |
| 140 | >KDQ23406.1 hypothetical protein PLEOSDRAFT_1090849 [Pleurotus ostreatus PC15]           | Hongo |
| 141 | >KIY44319.1 triose phosphate isomerase [Fistulina hepatica ATCC 64428]                   | Hongo |
| 142 | >KII85507.1 hypothetical protein PLICRDRAFT_31704 [Plicaturopsis crispa FD-325 SS-3]     | Hongo |
| 143 | >KDQ53926.1 hypothetical protein JAAARDRAFT_38896 [Jaapia argillacea MUCL 33604]         | Hongo |
| 144 | >EJF63635.1 triose phosphate isomerase [Dichomitus squalens LYAD-421 SS1]                | Hongo |
| 145 | >KXN89180.1 Triosephosphate isomerase [Leucoagaricus sp. SymC.cos]                       | Hongo |
| 146 | >EPQ51247.1 Triosephosphate isomerase [Gloeophyllum trabeum ATCC 11539]                  | Hongo |
| 147 | >EFI91124.1 hypothetical protein SCHCODRAFT_71461 [Schizophyllum commune H4-8]           | Hongo |
| 148 | >KLO13415.1 triose phosphate isomerase [Schizopora paradoxa]                             | Hongo |
| 149 | >EJD40324.1 triose phosphate isomerase [Auricularia subglabra TFB-10046 SS5]             | Hongo |
| 150 | >CDS12634.1 hypothetical protein LRAMOSA04820 [Lichtheimia ramosa]                       | Hongo |
| 151 | >CEJ01066.1 Putative Triosephosphate isomerase [Rhizopus microsporus]                    | Hongo |
| 152 | >AFJ73484.1 triose phosphate isomerase [Neocallimastix frontalis]                        | Hongo |
| 153 | >OAQ31557.1 triosephosphate isomerase [Mortierella elongata AG-77]                       | Hongo |
| 154 | >ESA13669.1 hypothetical protein GLOINDRAFT_176549 [Rhizophagus irregularis DAOM 181602] | Hongo |
| 155 | >KXS20846.1 Triosephosphate isomerase [Gonapodya prolifera JEL478]                       | Hongo |
| 156 | >KNE63655.1 triose-phosphate isomerase [Allomyces macrogynus ATCC 38327]                 | Hongo |
| 157 | >GAA96249.1 hypothetical protein E5Q_02913 [Mixia osmundae IAM 14324]                    | Hongo |
| 158 | >KDE02765.1 triosephosphate isomerase [Microbotryum lychnidis-dioicae p1A1 Lamole]       | Hongo |
| 159 | >gi 401626311 gb EJS44263.1  [Saccharomyces arboricola H-6]                              | Hongo |
| 160 | >gi 151942040 gb EDN60396.1  [Saccharomyces cerevisiae YJM789]                           | Hongo |
| 161 | >gi 52783442 sp Q6FRI3.1 CANGA RecName: Candida glabrata CBS 138                         | Hongo |
| 162 | >gi 12276170 gb AAG50278.1  [Zygosaccharomyces bailii]                                   | Hongo |
| 163 | >gi 442570059 sp Q750Y8.2 ASHGO RecName: Eremothecium gossypii                           | Hongo |
| 164 | >CDO96232.1 unnamed protein product [Kluyveromyces dobzhanskii CBS 2104]                 | Hongo |

|     |   |        |
|-----|---|--------|
| 165 | >gi 199425098 emb CAG98632.2  KLLA0F18832p [Kluyveromyces lactis]                               | Hongo  |
| 166 | >gi 34850299 emb CAE12106.1  [Kluyveromyces marxianus]  | Hongo  |
| 167 | >CCH62862.1 hypothetical protein TBLA_OI02040 [Tetrapisispora blattae CBS 6284]                 | Hongo  |
| 168 | >KSA00848.1 Triosephosphate isomerase [Debaryomyces fabryi]                                     | Hongo  |
| 169 | >gi 50424135 ref XP_460653.1  DEHA2F06754p [Debaryomyces hansenii CBS767]                       | Hongo  |
| 170 | >gi 146450505 gb EDK44761.1  [Lodderomyces elongisporus NRRL YB-4239]                           | Hongo  |
| 171 | >gi 6841084 gb AAF28895.1 AF124845_1 [Candida albicans]   | Hongo  |
| 172 | >ODV63945.1 triosephosphate isomerase [Ascoidea rubescens DSM 1968]                             | Hongo  |
| 173 | >CAY71012.1 Triose phosphate isomerase, abundant glycolytic enzyme [Komagataella phaffii GS115] | Hongo  |
| 174 | >ESW96569.1 Triosephosphate isomerase [Ogataea parapolyomorpha DL-1]                            | Hongo  |
| 175 | >ODV75700.1 Triosephosphate isomerase [Cyberlindnera jadinii NRRL Y-1542]                       | Hongo  |
| 176 | >OEJ80747.1 Triosephosphate isomerase [Hanseniaspora osmophila]                                 | Hongo  |
| 177 | >gi 125747035 gb ABN55898.1  [Candida glycerinogenes]   | Hongo  |
| 178 | >gi 49650893 emb CAG77830.1  YALIOF05214p [Yarrowia lipolytica CLIB122]                         | Hongo  |
| 179 | >gi 212003358 gb EEB09018.1  [Schizosaccharomyces japonicus yFS275]                             | Hongo  |
| 180 | >gi 13432251 sp P07669.4 SCHPO RecName: Schizosaccharomyces pombe 972h                          | Hongo  |
| 181 | >EFP74082.1 triosephosphate isomerase [Puccinia graminis f. sp. tritici CRL 75-36-700-3]        | Hongo  |
| 182 | >XP_007874541.1 triose-phosphate isomerase [Pneumocystis murina B123]                           | Hongo  |
| 183 | >KTW32704.1 triose-phosphate isomerase [Pneumocystis jirovecii RU7]                             | Hongo  |
| 184 | >EOQ99348.1 Triosephosphate isomerase [Wallemia ichthyophaga EXF-994]                           | Hongo  |
| 185 | >EFH52561.1 predicted protein [Arabidopsis lyrata subsp. lyrata]                                | Planta |
| 186 | >gi 7076787 emb CAB75902.1  cytosolic triosephosphatisomerase [Arabidopsis thaliana]            | Planta |
| 187 | >gi 602590 emb CAA58230.1  triosephosphate isomerase [Petunia x hybrida]                        | Planta |
| 188 | >CAC14917.1 triosephosphatisomerase [Triticum aestivum]   | Planta |
| 189 | >CAA81487.1 triosephosphate isomerase [Secale cereale]  | Planta |
| 190 | >AAB41052.1 cytosolic triosephosphate isomerase [Hordeum vulgare]                               | Planta |
| 191 | >ADM86861.1 triosephosphate isomerase [Oryza sativa Japonica Group]                             | Planta |
| 192 | >AAB81110.1 triosephosphate isomerase 1 [Zea mays]  | Planta |
| 193 | >KMZ66002.1 Triosephosphate isomerase [Zostera marina]  | Planta |
| 194 | >OEL26018.1 Triosephosphate isomerase, cytosolic [Dichanthelium oligosanthes]                   | Planta |
| 195 | >EFJ37925.1 hypothetical protein SELMODRAFT_73442 [Selaginella moellendorffii]                  | Planta |
| 196 | >EFN53775.1 triosephosphate isomerase cytoplasmic type [Chlorella variabilis]                   | Planta |
| 197 | >AAU93945.1 triose phosphate isomerase [Helicosporidium sp. ex Simulium jonesi]                 | Planta |

|     |   |          |
|-----|---|----------|
| 198 | >EIE20135.1 Triosephosphate isomerase [ <i>Coccomyxa subellipsoidea</i> C169]   | Planta   |
| 199 | >gi 1351282 sp P48497.1 TPIS_STELP RecName: Full=Triosephosphate isomerase, cytosolic; Short=TIM; Short=Triosephosphate isomerase <i>Stellaria longipes</i> | Planta   |
| 200 | >gi 37811994 gb AAR04016.1  cytosolic triosephosphate isomerase [ <i>Euglena gracilis</i> ]   | Protista |
| 201 | >gi 906958815 gb KNC86924.1  triosephosphate isomerase, cytosolic [ <i>Sphaeroforma arctica</i> JP610]  | Protista |
| 202 | >gi 301103109 ref XP_002900641.1  triosephosphate isomerase [ <i>Phytophthora infestans</i> T30-4]  | Protista |
| 203 | >gi 970635754 gb KUF80580.1  Triosephosphate isomerase [ <i>Phytophthora nicotianae</i> ]   | Protista |
| 204 | >gi 574112779 gb ETV78271.1  triose-phosphate isomerase [ <i>Aphanomyces astaci</i> ]   | Protista |
| 205 | >EJY68713.1 Triosephosphate isomerase [ <i>Oxytricha trifallax</i> ]  | Protista |
| 206 | >EGR31095.1 triosephosphate isomerase, putative [ <i>Ichthyophthirius multifiliis</i> ]   | Protista |
| 207 | >gi 146146531 gb EAR88549.2  triosephosphate isomerase (macronuclear) [ <i>Tetrahymena thermophila</i> SB210]   | Protista |
| 208 | >gi 954185857 gb KRW98733.1  Triosephosphate isomerase [ <i>Pseudocohnilembus persalinus</i> ]  | Protista |
| 209 | >XP_001434096.1 hypothetical protein (macronuclear) [ <i>Paramecium tetraurelia</i> strain d4-2]  | Protista |
| 210 | >gi 820693221 tpe CEL68962.1  TPA: Triosephosphate isomerase [ <i>Neospora caninum</i> Liverpool]   | Protista |
| 211 | >gi 557137694 emb CDJ37485.1  triosephosphate isomerase [ <i>Eimeria tenella</i> ]  | Protista |
| 212 | >gi 1069840529 gb OEH77147.1  triosephosphate isomerase [ <i>Cyclospora cayetanensis</i> ]  | Protista |
| 213 | >CDJ28937.1 triosephosphate isomerase, putative [ <i>Eimeria mitis</i> ]  | Protista |
| 214 | >gi 608679923 gb EZG72708.1  triosephosphate isomerase [ <i>Gregarina niphandrodes</i> ]  | Protista |
| 215 | >gi 1036546631 emb SBS86859.1  triosephosphate isomerase (TIM) [ <i>Plasmodium ovale curtisi</i> ]  | Protista |
| 216 | >gi 160706 gb AAA18799.1  triosephosphate isomerase [ <i>Plasmodium falciparum</i> ]  | Protista |
| 217 | >gi 113967769 gb ABI49147.1  triosephosphate isomerase [ <i>Paratrimastix pyriformis</i> ]  | Protista |
| 218 | >gi 470247469 ref XP_004357462.1  triose phosphate isomerase [ <i>Dictyostelium fasciculatum</i> ]  | Protista |
| 219 | >gi 281208449 gb EFA82625.1  triose phosphate isomerase [ <i>Polysphondylium pallidum</i> PN500]  | Protista |
| 220 | >EGC40028.1 hypothetical protein DICPUDRAFT_44767 [ <i>Dictyostelium purpureum</i> ]  | Protista |
| 221 | >gi 60472175 gb EAL70128.1  triose phosphate isomerase [ <i>Dictyostelium discoideum</i> AX4]   | Protista |
| 222 | >CCC53086.1 putative triosephosphate isomerase [ <i>Trypanosoma vivax</i> Y486]   | Protista |
| 223 | >gi 10543 emb CAA27559.1  triosephosphate isomerase [ <i>Trypanosoma brucei</i> ]   | Protista |



|     |  |                   |
|-----|--|-------------------|
| 224 | >gi 1289358 gb AAB58349.1  triosephosphate isomerase [Trypanosoma cruzi]   | Protista          |
| 225 | >gi 322504827 emb CBZ14613.1  triosephosphate isomerase [Leishmania braziliensis MHOM/BR/75/M2904]                             | Protista          |
| 226 | >gi 471326 emb CAA52804.1  triosephosphate isomerase [Leishmania mexicana]   | Protista          |
| 227 | >EPY30378.1 triosephosphate isomerase [Strigomonas culicis]  | Protista          |
| 228 | >CCW62522.1 unnamed protein product [Phytomonas sp. isolate EM1]   | Protista          |
| 229 | >CUG78480.1 triosephosphate isomerase, putative [Bodo saltans]   | Protista          |
| 230 | >gi 209881640 ref XP_002142258.1  triosephosphate isomerase [Cryptosporidium muris RN66]                                       | Protista          |
| 231 | >gi 1005910191 gb JAP92511.1  Triosephosphate isomerase [Trepomonas sp. PC1]   | Protista          |
| 232 | >gi 558598229 gb EST45141.1  Triosephosphate isomerase [Spiroplasma salmonicida]   | Protista          |
| 233 | >gi 549096 sp P36186.1 TPI1_GIAIN RecName: Full=Triosephosphate isomerase; Short=TIM; AltName: Full=Triose-phosphate isomerase | Protista          |
| 234 | >gi 123416314 ref XP_001304868.1  triosephosphate isomerase [Trichomonas vaginalis G3]   | Protista          |
| 235 | >OHT14411.1 Triosephosphate isomerase, cytosolic [Trichomonas foetus]  | Protista          |
| 236 | >gi 165896730 gb EDR24065.1  triosephosphate isomerase [Entamoeba dispar SAW760]   | Protista          |
| 237 | >gi 2143243 emb CAA73817.1  triosephosphate isomerase [Entamoeba histolytica]  | Protista          |
| 238 | >gi 71028912 ref XP_764099.1  triosephosphate isomerase [Theileria parva strain Muguga]  | Protista          |
| 239 | >EDO06081.1 triosephosphate isomerase protein, putative [Babesia bovis]  | Protista          |
| 240 | >pdb 4K6A A Chain A, Triosephosphate Isomerase [Escherichia coli]  | Outgroup-bacteria |
| 241 | >VEA33046.1 triosephosphate isomerase [Salmonella enterica subsp. enterica]  | Outgroup-bacteria |
| 242 | >AEK00495.1 triosephosphate isomerase [Klebsiella pneumoniae KCTC 2242]  | Outgroup-bacteria |

**Anexo 2:** Secuencia de nucleótidos de las proteínas ancestrales. Las secuencias ML son las reconstruidas por máxima verosimilitud, las secuencias BAY son las secuencias reconstruidas por inferencia bayesiana y las secuencias Alt se refieren a las secuencias de peor caso plausible. En negritas se indica la secuencia de la proteína, flanqueadas por las secuencias de los sitios de restricción.

ML

CATATG

ATGGCGCGTAAGTTCTTTGTTGGTGGCAACTGGAAAATGAACGGTACCATCGAGAGCATTAAAGAGCCTGGTGGAAACC  
CTGAACAACGCGGAGCTGGACGATAACACCGAAGTGGTTGTTGCTCCGCCGCGATCTACCTGCCGCTGGTTCGTC  
GAGCCTGCGTAAGGACATTCAGGTGGCGGCGCAAACTGCTATACCAAAGCGAGCGGTGCGTTCACCGGTGAAATCA  
GCGCGGAAATGCTGAAAGATCTGGGTATCCCGTGGGTTATTCTGGGCCACAGCGAGCGTCTCAGATTTTCGGCGAG  
AGCGACGAATTTGTTGCGGAAAAGACCAAATACGCGCTGGATAACGGTCTGAAGGTGATCCTGTGCATGGCGAGAC  
CCTGGAGGAACGTGAAGCGGGCAAAACCATGGACGTTGTGACCCGTCAACTGAAGGCGATCGCGGACAAAATTAGCG  
AGAGCGATTGGAGCAAGGTTGTGATCGCGTACGAACCGGTTTGGGCGATTGGTACCGGTAAGTGGCGACCCGGAG  
CAGGCGCAAGAAGTTCACGCGGAGATCCGTAAGTGGCTGGCGGAAAACGTTAGCGCGGAAGTGGCGGAAAGCACCC  
GTATCATTATGGTGGCAGCGTGAACGGTGGCAACTGCAAGGAGCTGGCGAAACAGCCGGACATCGATGGCTTCCTG  
GTTGGTGGCGGAGCCTGAAGCCGGAATTTGTGGATATCATTAACGCGCGTCAAAAA

TAACTCGAG

BAY

CATATG

ATGGCGCGTAAGTTCTTTGTTGGTGGCAACTGGAAAATGAACGGTAGCAAGGCGAGCATCAAAAAGCCTGGTGGAAAC  
CCTGAACGCGAGCAAGCTGGACAACAAAACCGAGGTGGTTGTTGCTCCGCCGCGATCTACCTGCCGAGGTTCGTG  
AAGCGCTGCGTAAGGATATTCAGGTGGCGGCGCAAAACTGCTATACCAAAGCGAGCGGTGCGTTCACCGGTGAAATC  
AGCCCGAAAATGCTGAAGGACCTGGGTATCCCGTGGGTTATTCTGGGCCACAGCGAACGTCGTCACATTTTCGGCGA  
GAGCGACGAATTTGTTGCGGAGAAGGTGAAATACGCGCTGGATCAGGGTCTGAAAGTTATCGCGTGCATTGGCGAGA  
CCCTGGAGGAACGTGAAGCGGGCAAGACCATGGAAGTGTGCGCGCGTCAACTGAAAGCGATCGCGAAGAAAATTAGC  
GAAAGCGATTGGAGCAACGTTGTGATCGCGTACGAGCCGGTTTGGGCGATTGGTACCGGCAAGGTGGCGACCCCGGA  
GCAGGCGCAAGAAGTTCACGCGGCGATCCGTAATGGCTGGCGGAAAACGTTAGCGCGGAAGTGGCGGAAAAGCACCC  
GTATCATTTATGGTGGCAGCGTGAACGGTGGCAACTGCGCGGAACTGGCGAAGCAGCCGGACATTGATGGCTTCCTG  
GTTGGTGGCGGAGCCTGAAGCCGGAGTTTGTGGACATCATTAACGCGCGTAAACAATAACTCGAG

AM

CATATG

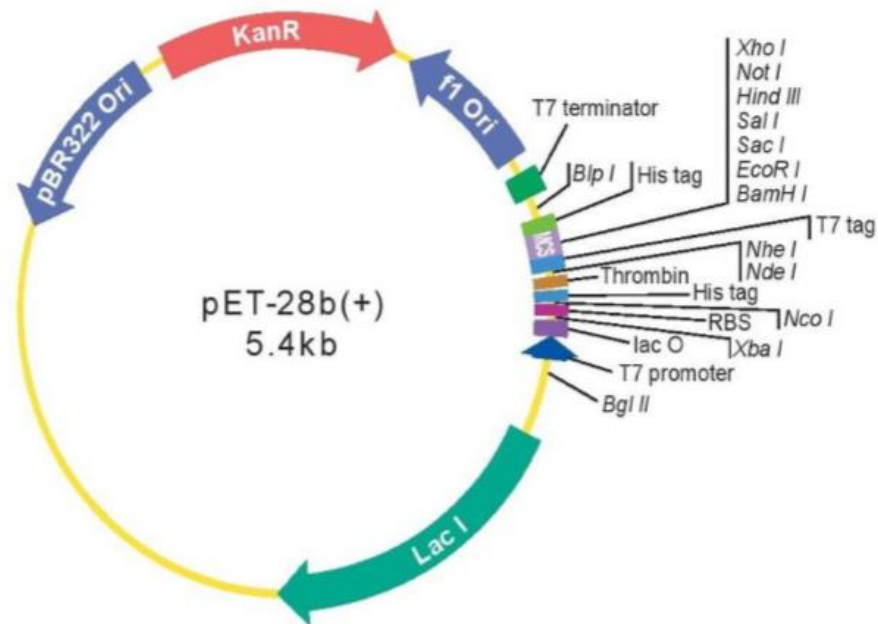
ATGGCGCGTAAGTTCTTTGTGGGTGGCAACTGGAAAATGAACGGTAGCGTGGAGAGCATCAAAAAGCCTGGTTGAAAC  
CCTGAACAACGCGGAGCTGGACCCGAACGTGGAAGTGGTTATTGCTCCGCCGCGGTGTACCTGCCGCTGGTTCGTG  
AAGCGCTGCGTAAGGATATTCAGGTTCGCGGCGCAAAACTGCTACACCAAAGCGAGCGGTGCGTATAACCGGCGAGATC  
AGCGCGAAAATGCTGAAGGACCTGGGTATCCCGTGGGTTATTCTGGGCCACAGCGAGCGTTCGTCACATTTTCGGCGA  
GAGCGACGAAGTGGTGGCGGAAAAGACCAAATATGCGCTGGATAGCGGTCTGAGCGTTATCCTGTGCATTGGCGAGA  
CCCTGGAGGAACGTGAAGCGGGCAAAACCATGGATGTGGTTGCGCGTCAAGTGAAGGCGATCGCGGACAAAATTACC  
GAGAACGATTGGAGCAAGGTGGTTATCGCGTACGAACCGGTGTGGGCGATTGGTACCGGTAAAAGTTGCGACCCCGGA  
GCAGGCGCAAGAAGTGCATGCGGAGATCCGTAAGTGGCTGGCGGAGAACGTGAGCGCGGAAGTGGCGGAAAAGCACCC  
CGTATCATTTATGGTGGCAGCGTTAACGGTGGCAACTGCAAGGAGCTGGCGAAAACAGCCGGACATTGATGGCTTCCT  
GGTGGGTGGCGGAGCCTGAAGCCGGAATTTGTTGACATCATTAACGCGAAAACAAGCGTAACTCGAG

AB

CATATG

ATGGCGCGTAAGTTCTTTGTGGGTGGCAACTGGAAAATGAACGGTACCCAGGAAAGCATCAAGAGCCTGGTGGAGAC  
CCTGAACGCGAACAACCTGGACAACAAAGTTGAAAGTGGTTATTGCTCCGCCGCGGTGTACCTGCCGCAAGTTCGTCA  
GCAACTGCGTAAGGATATTCAGGTGGCGGCGCAAAACTGCTATACCAAAGGTAGCGGTGCGTTCACCGGCGAGATCA  
GCCCGGAAATGCTGAAGGACCTGGGTATCCCGTGGGTTATTCTGGGCCACAGCGAGCGTTCGTAACATTTTCGGCGAG  
AGCGATGAACTGGTGGCGGAAAAGGTTAAACACGCGCTGGAGGCGGGTCTGAAAAGTTATCTTTTGCATTGGCGAGAC  
CCTGGAGGAACGTGAAGCGGGCAAGACCACCGAGGTGGTTGCGCGTCAAGTGAAGGCGATCGCGAAGAAAATTAGCG  
AAAGCGACTGGAGCAAGGTGGTTATCGCGTACGAGCCGGTGTGGGCGATTGGTACCGGTAAAAGTTGCGACCCCGGAG  
CAGGCGCAAGAAGTGCATGCGGCGATCCGTAAGTGGCTGGCGAGCAACGTGAGCGCGGAAGTGGCGGAAAAGCACCC  
GTATCATTTATGGTGGCAGCGTTAACGGTAGCAACTGCGCGGAACTGGCGAAGCAGCCGGACATTGATGGCTTCCTG  
GTGGGTGGCGGAGCCTGAAACCGGAGTTTGTGATATCATTAACGCGAAGAAAACAATAACTCGAG

### Anexo 3: Plásmido utilizado para la sobreexpresión



*Bgl II* T7 promoter lac operator *Xba I* rbs  
 -- AG ATC TCG ATC CCG CGA AAT TAA TAC GAC TCA CTA TAG GGG AAT TGT GAG CGG ATA ACA ATT CCC CTC TAG AAA TAA TTT TGT TTA ACT TTA AGA AGG AGA  
*Nco I* His tag *Nde I* *Nhe I* T7 tag  
 TAT ACC ATG GGC AGC AGC CAT CAT CAT CAT CAC AGC AGC GGC CTG GTG CCG CGC GGC AGC CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT  
 M G S S H H H H H S S G L V P R G S H M A S M T G G E E M G  
 thrombin site  
*BamH I* *EcoR I* *Sac I* *Sal I* *Hind III* *Not I* *Xho I* His tag  
 CCG GAT CCG AAT TCG AGC TCC GTC GAC AAG CTT GCG GCC GCA CTC GAG CAC CAC CAC CAC CAC TGA GAT CCG GCT GCT AAC AAA GCC CGA AAG GAA  
 R D P N S S S V D K L A A A L E H H H H H H H stop  
*Bsp I* T7 terminator  
 GCT GAG TTG GCT GCT GCC ACC GCT GAG CAA TAA CTA GCA TAA CCC CTT GGG GCC TCT AAA CGG GTC TTG AGG GGT TTT TTG --

### Anexo 4: Secuencias de aminoácidos de las secuencias ancestrales en formato fasta.

Las secuencias que se muestran a continuación son: M, la secuencia reconstruida por máxima verosimilitud; BAY, la secuencia reconstruida por inferencia bayesiana; AM, la secuencia peor caso plausible reconstruida por máxima verosimilitud y AB, la secuencia peor caso plausible reconstruida por inferencia bayesiana.

>ML

MARKFFVGGNWKMNGTIESIKSLVETLNNaelDDNTEVVVAPPaiYlPLVRQSLRKDIQVAAQN  
 CYTKASGAFTGEISAEMlKDLGIPWVILGHsERRQIFGESDEFVAEKTKYALDNGlKVILCIGETL  
 EEREAGKTMDVVTRQLKAIADKISESDWSKVViAYEPVWAIGTGKVATPEQAQEVHAEIRKWL  
 AENVSAEVAESTRIIYGGSVNGGNCKELAKQPDIDGFLVGGASLKPEFVDIINARQK

>BAY

MARKFFVGGNWKMNgsKASIKSLVETLNASKLDNKTEVVVAPPaiYlPQVRQALRKDIQVAAQ  
 NCYTKASGAFTGEISPEMLKDLGIPWVILGHsERRHIFGESDEFVAEKVKYALDQGLKVIACIGET  
 LEEREAGKTMEVCARQLKAIAKKISESDWSNVViAYEPVWAIGTGKVATPEQAQEVHAAIRKWL  
 AENVSAEVAESTRIIYGGSVNGGNCAELAKQPDIDGFLVGGASLKPEFVDIINARKQ

>AM

MARKFFVGGNWKMNGSVESIKSLVETLNN AELDPNVEVVIAPPAVYLPVREALRKDIQVAAQ  
NCYTKASGAYTGEISAEMLKDLGIPWVILGHSERRHIFGESDELVAEKT KYALDSGLSVILCIGET  
LEEREAGKTM DVVARQLKAIADKITENDWSKV VIA YEPVWAIGTGKVATPEQAQEVHAEIRKW  
LAENVSAEVAESTRIIYGGSVNGGNCKELAKQPDIDGFLVGGASLKPEFVDIINAKQA

>AB

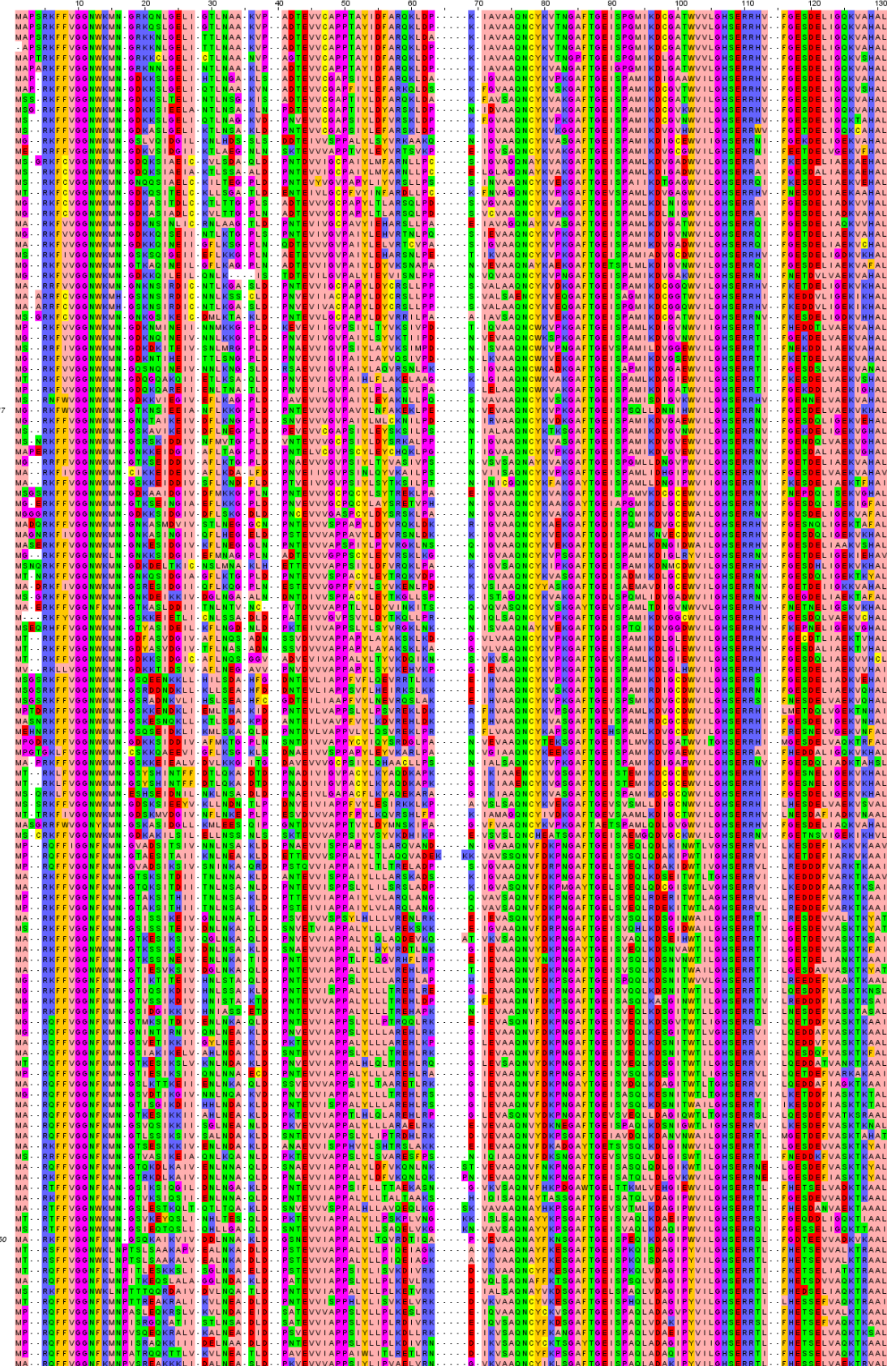
MARKFFVGGNWKMNGTQESIKSLVETLNANNLDNKVEVVIAPPAVYLPQVRQQLRKDIQVAAQ  
NCYTKGSGAFTGEISPEMLKDLGIPWVILGHSERRNIFGESDELVAEKVKHALEAGLKVIFCIGET  
LEEREAGKTTEVVARQLKAI AKKISESDWSKV VIA YEPVWAIGTGKVATPEQAQEVHAAIRKWL  
ASNVSAEVAESTRIIYGGSVNGSNCAELAKQPDIDGFLVGGASLKPEFVDIINAKKQ

**Anexo 5:** Propiedades de las proteínas ancestrales calculadas a partir de la herramienta ProtParam en la página ExPasy con la etiqueta de histidinas.

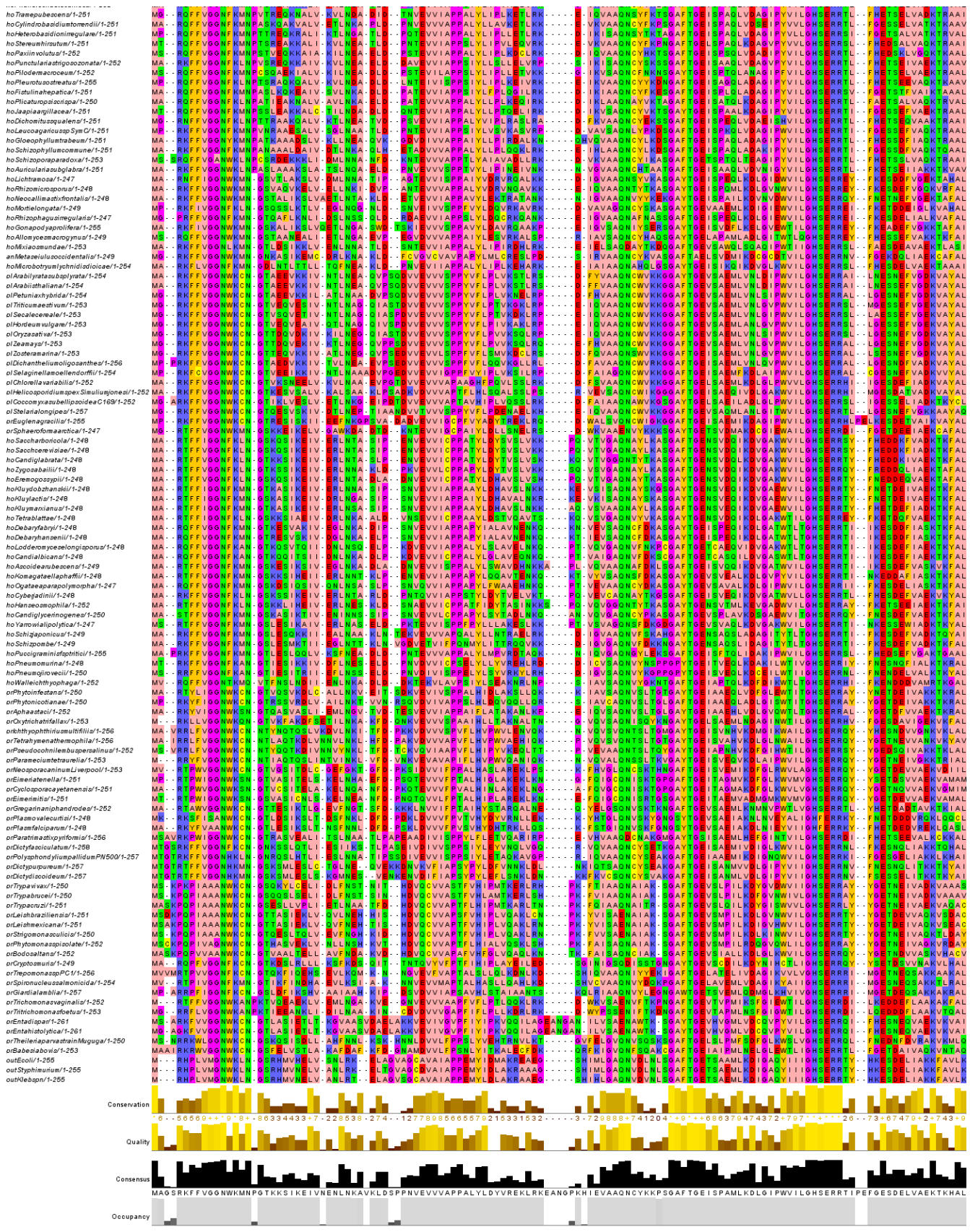
| Proteína ancestral | No. de aa. | MW (Da)  | pI   | Coefficiente de extinción molar |
|--------------------|------------|----------|------|---------------------------------|
| ML                 | 270        | 29560.60 | 5.88 | 34950                           |
| BAY                | 270        | 29364.55 | 7.73 | 34950                           |
| AM                 | 270        | 29346.37 | 5.71 | 36440                           |
| AB                 | 270        | 29380.52 | 7.16 | 33460                           |

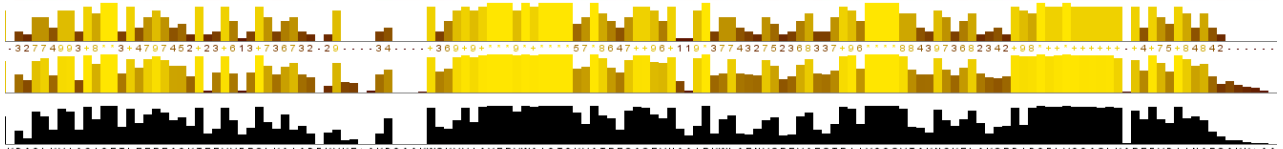
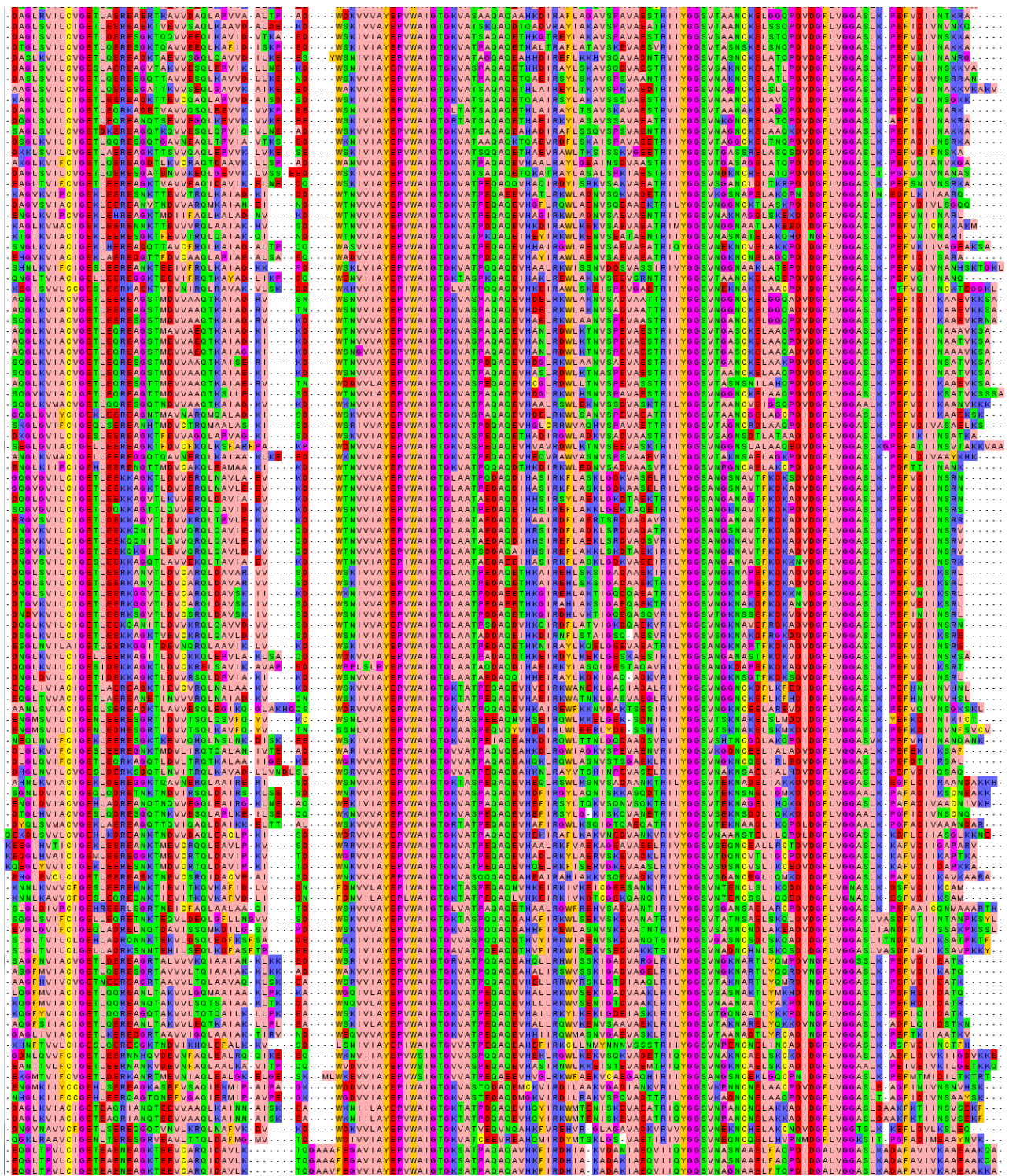
Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M. R., Appel R. D., Bairoch A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

**Anexo 6: Alineamiento múltiple de secuencias de las 239 secuencias de TIM con las 3 secuencias externas. Código de color: Alifáticos-rosa (ILVAM); Aromáticos-naranja (FWY); Positivos-azul(KRH); Negativos-rojo (DE); Hidrofilicos-verde (STNQ); Conformacionalmente especiales-magenta (PG); Cisteína-amarillo(C).**









KDAGLKVIACIGETLEEREAAGTKEVVFRQLKAIADPKVKEAKDGAAKVSKVVIAYEPVWAIIGTKVATPEQAEVHAAIRKWLAEVNSPEVAESTRIYGGVTAKNCKELAKQDPDIQDFLVGGASLKAFEPVDIINARQAKKAA



**Anexo 7:** Aminoácidos ambiguamente reconstruidos en las secuencias ancestrales. En la columna de máxima verosimilitud y de inferencia bayesiana, se encuentran los aminoácidos con mayor probabilidad posterior. En las columnas de peor caso plausible se encuentran los segundos mejores aminoácidos que fueron colocados en la secuencia para cada uno de los métodos de reconstrucción. Las probabilidades posteriores se encuentran entre paréntesis. Los sitios resaltados en rosa son cambios drásticos en la fisicoquímica del aminoácido. Se determinaron sitios que, por un método eran ambiguos, mientras que por el otro no lo eran. En negritas se muestra dichos sitios para su comparación.

| Sitio | Máxima verosimilitud | Peor caso plausible máxima verosimilitud | Inferencia Bayesiana | Peor caso plausible Inferencia Bayesiana |
|-------|----------------------|--|----------------------|--|
| 16    | T(0.638)             | S (0.362)                                | S (0.532)            | T (0.467)                                |
| 17    | I (0.426)            | V (0.332)                                | <b>K (0.318)</b>     | <b>Q (0.314)</b>                         |
| 18    | <b>E (0.779)</b>     | <b>E (0.779)</b>                         | A (0.751)            | E (0.229)                                |
| 30    | <b>A (1)</b>         | <b>A (1)</b>                             | S (0.605)            | N (0.391)                                |
| 31    | <b>E (0.702)</b>     | <b>E (0.702)</b>                         | K (0.601)            | N (0.305)                                |
| 34    | <b>D (0.369)</b>     | P (0.295)                                | <b>N (0.995)</b>     | <b>N (0.995)</b>                         |
| 36    | T (0.749)            | V (0.250)                                | <b>T (0.560)</b>     | <b>V (0.439)</b>                         |
| 40    | V (0.768)            | I (0.232)                                | V (0.796)            | I (0.204)                                |
| 45    | I (0.434)            | V (0.397)                                | I (0.776)            | V (0.208)                                |
| 52    | <b>Q (0.449)</b>     | E (0.338)                                | <b>Q (0.827)</b>     | <b>Q (0.827)</b>                         |
| 53    | S (0.348)            | A (0.344)                                | A (0.306)            | Q (0.202)                                |
| 69    | <b>A (0.626)</b>     | <b>A (0.626)</b>                         | A (0.664)            | G (0.224)                                |
| 73    | F (0.580)            | Y (0.420)                                | <b>F (0.924)</b>     | <b>F (0.924)</b>                         |
| 99    | <b>Q (0.459)</b>     | H (0.439)                                | H (0.419)            | N (0.282)                                |
| 107   | <b>F (0.774)</b>     | L (0.225)                                | F (0.658)            | L (0.323)                                |
| 114   | <b>Y (0.961)</b>     | <b>F (0.924)</b>                         | Y (0.337)            | H (0.353)                                |
| 117   | <b>D (0.871)</b>     | <b>D (0.871)</b>                         | D (0.492)            | E (0.332)                                |
| 118   | <b>N (0.371)</b>     | S (0.276)                                | Q (0.626)            | A (0.334)                                |
| 121   | <b>K (0.434)</b>     | S (0.315)                                | <b>K (0.434)</b>     | <b>K (0.434)</b>                         |
| 124   | <b>L (0.654)</b>     | <b>L (0.654)</b>                         | A (0.569)            | F (0.421)                                |
| 139   | <b>M (0.905)</b>     | <b>M (0.905)</b>                         | M(0.449)             | T (0.266)                                |
| 142   | <b>V (0.999)</b>     | <b>V (0.999)</b>                         | C (0.562)            | V (0.430)                                |
| 143   | <b>T (0.582)</b>     | A (0.430)                                | <b>A (0.950)</b>     | <b>A (0.950)</b>                         |
| 154   | S (0.653)            | T (0.205)                                | <b>S (0.942)</b>     | <b>S (0.942)</b>                         |
| 156   | <b>S (0.438)</b>     | N (0.320)                                | <b>S (0.867)</b>     | <b>S (0.867)</b>                         |
| 160   | <b>K (0.997)</b>     | <b>K (0.997)</b>                         | N (0.731)            | K (0.269)                                |
| 195   | <b>E (0.766)</b>     | <b>E (0.766)</b>                         | E (0.766)            | S (0.225)                                |
| 215   | <b>G (0.741)</b>     | <b>G (0.741)</b>                         | G (0.360)            | S (0.306)                                |
| 248   | R (0.505)            | K (0.388)                                | R (0.556)            | K (0.362)                                |
| 250   | <b>K (0.364)</b>     | A (0.202)                                | <b>Q (0.364)</b>     | <b>Q (0.364)</b>                         |

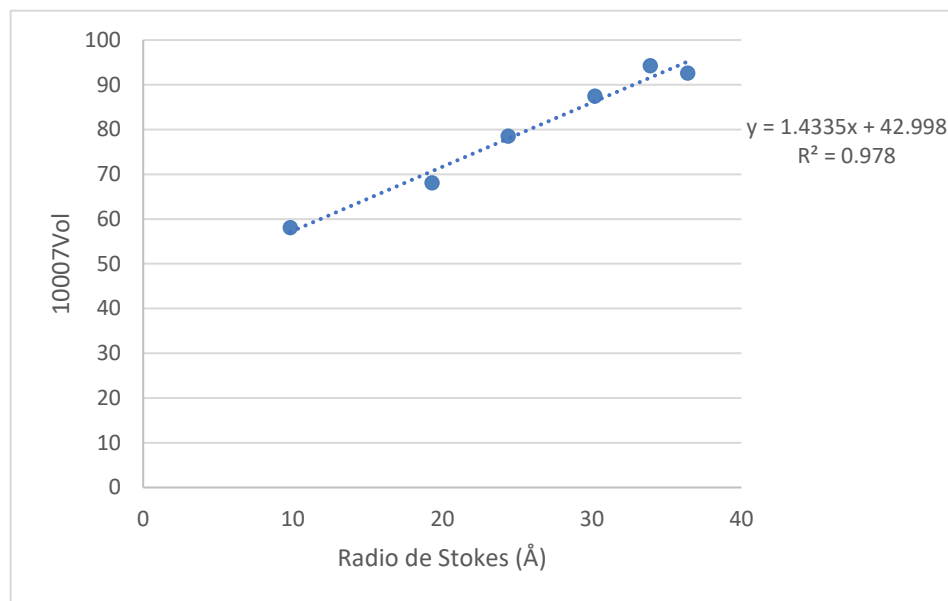
**Anexo 8:** Gráfico de estándares para radio hidrodinámico con sus respectivos valores y volúmenes y curva patrón con ecuación.

|                                  | MW (kDa) | Rs (Å) | Volumen de elución (mL) |
|----------------------------------|----------|--------|-------------------------|
| Aprotitina <sup>b</sup>          | 6.5      | 9.83   | 17.233                  |
| Ribonucleasa <sup>a</sup>        | 13.7     | 19.3   | 14.699                  |
| Anhidrasa carbonica <sup>a</sup> | 28       | 24.4   | 12.735                  |
| Ovoalbumina <sup>a</sup>         | 43.5     | 30.2   | 11.441                  |
| Albumina bovina <sup>a</sup>     | 66.3     | 33.9   | 10.614                  |
| Conalbumina <sup>c</sup>         | 75       | 36.4   | 10.801                  |

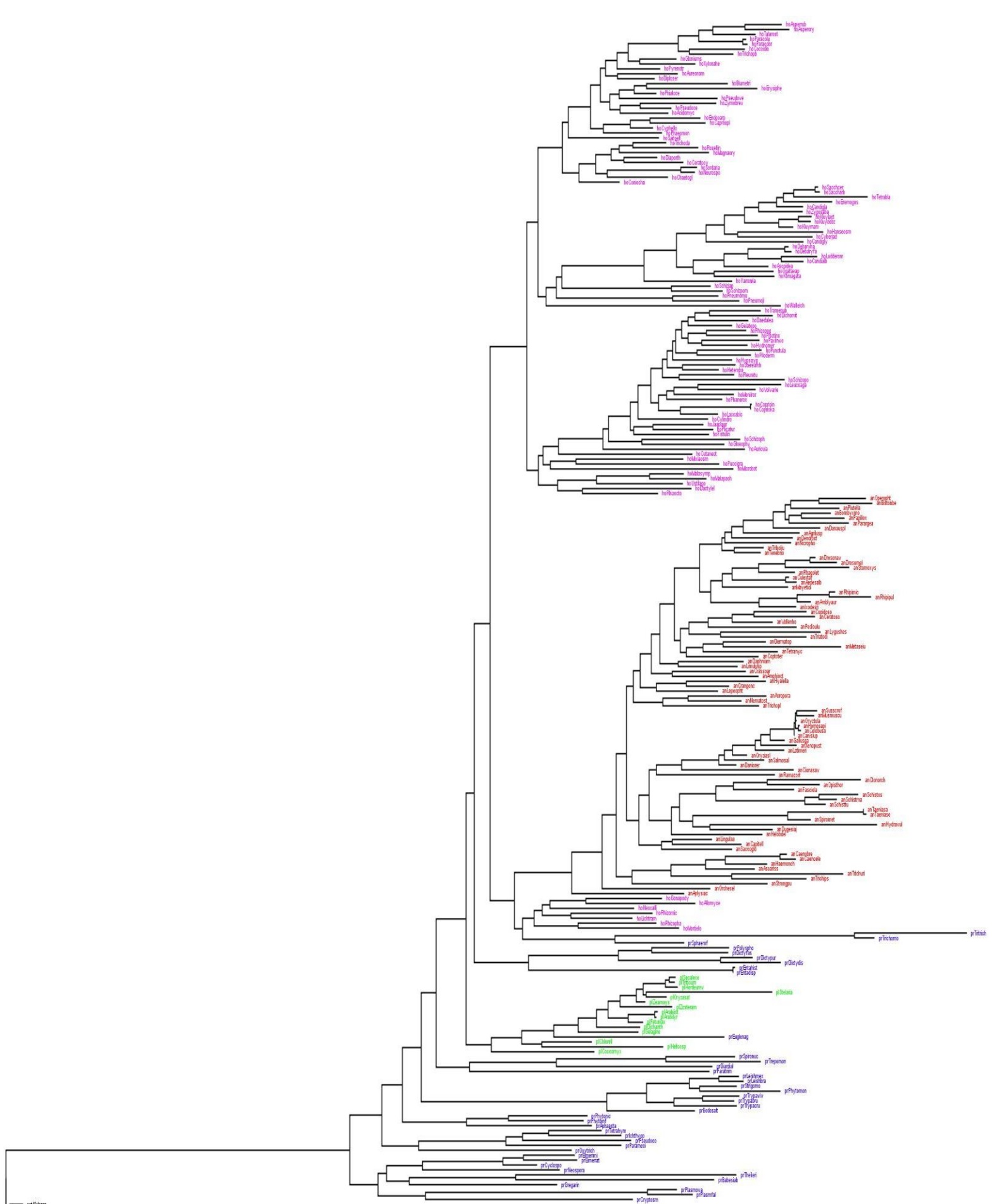
<sup>a</sup>Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, 32(48), 13288-13298.

<sup>b</sup>Carlson, J. C. T., Kanter, A., Thudupathy, G. R., Cody, V., Pineda, P. E., McIvor, R. S., & Wagner, C. R. (2003). Designing Protein Dimerizers: The Importance of Ligand Conformational Equilibria. *Journal of the American Chemical Society*, 125(6), 1501–1507.

<sup>c</sup>Quistgaard, E. M., Löw, C., Moberg, P., Guettou, F., Maddi, K., & Nordlund, P. (2013). Structural and Biophysical Characterization of the Cytoplasmic Domains of Human BAP29 and BAP31. *PLoS ONE*, 8(8).



**Anexo 9:** Árboles filogenéticos con la longitud reales de las ramas obtenidos por máxima verosimilitud e inferencia bayesiana (respectivamente) de las 239 secuencias de TIM con las 3 secuencias externas



0.01  
substitutions  
per site

