

OBTENCIÓN DE UNA ENZIMA CONSENSO DE GLICEROL QUINASA

SARA ISABEL VARELA HURTADO

**UNIVERSIDAD ICESI
FACULTAD DE CIENCIAS NATURALES
DEPARTAMENTO DE CIENCIAS QUÍMICAS
PROGRAMA DE QUÍMICA CON ÉNFASIS EN BIOQUÍMICA
SANTIAGO DE CALI
2015**

OBTENCIÓN DE UNA ENZIMA CONSENSO DE GLICEROL QUINASA

SARA ISABEL VARELA HURTADO

**TRABAJO DE GRADO PARA OPTAR AL TITULO DE:
PREGRADO EN QUIMICA CON ENFASIS EN BIOQUIMICA**

**TUTOR: DIEGO FERNANDO MEJIA CARMONA, Ph.D
COTUTOR: ARAM JOEL PANAY ESCOBAR, Ph.D**

**SANTIAGO DE CALI
2015**



APROBADO POR:

James Fairbanks Weston

Evaluador Externo.

Armando Lucumí Moreno

Evaluador Externo.

Diego Fernando Mejía Carmona

Director del Proyecto.

Aram Joel Panay Escobar

Co-Director del Proyecto.

AGRADECIMIENTOS

Quiero aprovechar estas líneas para agradecer a todas las personas que me han ayudado y me han apoyado a lo largo de estos años de estudio en la universidad Icesi.

En primer lugar quiero agradecer el apoyo recibido por parte de toda mi familia, a mis padres, que siempre han estado apoyándome desde que empecé a estudiar química con énfasis en bioquímica en esta universidad y a mis hermanos, que por tener experiencia en la vida universitaria, me han sabido guiar en los momentos que más me ha costado para poderme mantener al ritmo de ésta carrera.

También quiero dar agradecimientos a mi tutor de proyecto, el profesor Diego Fernando Mejía, por todo el apoyo brindado, por su calidad humana, por instruirme y guiarme a realizar este proyecto que tanto me llena de entusiasmo y orgullo. Gracias al profesor Aram Joel Panay por ser partícipe de esta defensa.

Finalmente, no quería pasar por alto la oportunidad de agradecer a todos los profesores que he tenido durante mi vida académica, pues es gracias a ellos que he logrado forjar todas las bases y conocimientos que tengo hoy en día, no sólo en el campo de la química, sino también que han influido en la formación de mi carácter personal.

TABLA DE CONTENIDO

LISTA DE TABLAS	6
RESUMEN DEL PROYECTO	8
1. INTRODUCCIÓN	9
2. DESCRIPCIÓN DEL PROYECTO	11
2.1. PLANTEAMIENTO Y JUSTIFICACIÓN.....	11
2.2. MARCO TEÓRICO Y ESTADO DEL ARTE	12
2.3. OBJETIVOS	19
2.3.1. OBJETIVO GENERAL.....	19
2.3.2. OBJETIVOS ESPECÍFICOS	19
2.6 DISCUSIÓN	28
2.7. CONCLUSIONES.....	31
2.8 RECOMENDACIONES	32
3. REFERENCIAS.....	33

LISTA DE TABLAS

Tabla 1: Parámetros cinéticos de enzima WildType de glicerol quinasa y 1GLJ. .24
Tabla 2: Valores de rmsd obtenidos para los alineamientos estructurales.....26
Tabla 3: Propiedades fisicoquímicas de la secuencia consenso y semilla.27

LISTA DE FIGURAS

Fig 1: reacciones químicas de la transesterificación.	12
Fig 2: Conversión del glicerol en L-glicerol-3-fosfato por la glicerol quinasa.	14
Fig 3: Resumen del flujo del algoritmo de MUSCLE.	16
Fig 4: Algoritmo para selección de secuencias.	24
Fig 5: Sobreposición de un monómero de la estructura tridimensional de a) 1GLJ vs modelo 1GLJ, b) 1GLJ vs WildType (semilla), c) 1GLJ vs Consenso y d) Semilla vs Consenso.	25
Fig 6: Representación de las cadenas laterales (rojo), de las 34 mutaciones presentes en la secuencia consenso.	27

RESUMEN DEL PROYECTO

La generación de biocombustibles ha sido un gran avance en la ciencia, pues proporciona una alternativa muy útil para reemplazar los combustibles derivados del petróleo, los cuales a diferencia de los biocombustibles no son renovables (Dufey, 2006).

La producción de biocombustibles presenta un reto que lo hace económicamente menos viable, y es que como derivado del proceso, se están obteniendo grandes cantidades de glicerol crudo, lo cual es una grave problemática para las refinerías. Este Glicerol tiene un alto costo de purificación, porque, cuenta con contaminantes tales como metanol, sales y ácidos grasos libres, los cuales no sólo impiden que pueda ser desechado libremente, sino también que sea útil para generar más productos fácilmente (João R M Almeida, 2012).

En éste proyecto se busca crear una secuencia primaria consenso para una enzima Glicerol quinasa (GK). Este consenso se obtendrá del alineamiento de múltiples secuencias de varias enzimas glicerol quinasa presentes en distintos organismos.

Para llevar a cabo esto, se emplearán herramientas computacionales y bases de datos para realizar la selección del grupo de enzimas.

Se realizará un alineamiento múltiple de las secuencias seleccionadas, a partir del cual se reunirá la información necesaria para construir la secuencia consenso de una enzima glicerol quinasa, con la idea de obtener una secuencia no existente en la naturaleza, que se busca usar como punto de partida para mejorar la conversión de glicerol en moléculas de mayor interés comercial. Una vez obtenida la secuencia deseada de la enzima, será obtenida su estructura tridimensional por modelamiento por homología.

1. INTRODUCCIÓN

En Colombia se produjeron 503.337 toneladas de biodiesel a partir de aceite de palma, del cual entre 10 y 40% (p/p) es de glicerol crudo, ésta cifra corresponde al total nacional de la producción de biodiésel en el año 2013, según lo reportó la Federación Nacional de Biocombustibles de Colombia. El porcentaje de glicerol obtenido por la producción del biocombustible es tan alto que ha ocasionado un grave impacto económico, pues su sobreproducción ha generado una sobreoferta en el mercado mundial y ha llevado a que su precio descienda a valores mínimos. Teniendo en cuenta que, para que este compuesto pueda ser empleado en los campos industriales debe ser purificado, pues debe de cumplir con un alto estándar en niveles de calidad y pureza, lo cual requiere de altas inversiones monetarias (Jorge Cardenas Gutierrez, 2003).

Lo que actualmente se busca es encontrar una manera para que pueda ser usado como materia prima, mediante la transformación enzimática la cual, básicamente, requiere de reactivos catalíticos, los cuales actúan bajo pequeñas cantidades y se pueden emplear eficazmente en repetidas ocasiones para obtener productos de mayor valor agregado.

La bioinformática como proceso de vital importancia para el desarrollo de este proyecto, deriva el conocimiento del análisis informático de los datos biológicos. La investigación en bioinformática incluye el desarrollo de métodos para el almacenamiento, recuperación y análisis de los datos. Esta es una rama de rápido desarrollo de la biología y es altamente interdisciplinaria, utilizando técnicas y conceptos de informática, estadística, matemáticas, química, bioquímica, física y la lingüística (Linge, 2000).

Esta investigación busca caracterizar, a nivel computacional, enzimas tipo glicerol quinasa eficientes provenientes de distintos organismos, con el fin de rediseñar una enzima que permita la degradación de glicerina cruda, subproducto de proceso de producción del biodiesel, para facilitar su eliminación o generación de productos de interés comercial. Es por esto que se busca caracterizar un conjunto de enzimas de tipo glicerol quinasa, que permitan seleccionar sus mejores dotes para así crear un ejemplar enzimático que recopile toda ésta información y sea capaz de aumentar la funcionalidad de la clase de enzima mencionada.

El proyecto hace parte de un área de investigación muy activa en la que se busca crear enzimas con características y propiedades nuevas mediante diseño racional, lo cual posee un alto grado de interés en el campo de enzimología y proteómica, porque puede servir como soporte para realizar procesos semejantes que además tengan como función ayudar a mejorar el medio ambiente y/o la estabilidad de los seres vivos.

Este proyecto puede causar un impacto positivo en las ciencias naturales ya que, busca crear el diseño de una enzima modelo a la cual en un futuro se le podrán hacer modificaciones por técnicas computacionales en su estructura primaria, que pueden lograr que ésta exprese una función de un interés biológico como por ejemplo la problemática que se ha expuesto sobre el glicerol.

2. DESCRIPCIÓN DEL PROYECTO

2.1. PLANTEAMIENTO Y JUSTIFICACIÓN

Hoy en día el glicerol se ha convertido en un producto que está causando un grave impacto a nivel industrial, económico y ambiental, pues es generado como subproducto de la elaboración del biodiesel.

Este subproducto a su vez está teniendo una sobreproducción que genera grandes costos a las biorefinerías, dado a las restricciones que tiene para ser desechado por los componentes que lo integran, los cuales impiden al mismo tiempo que se le dé alternativas de uso o comercialización a este subproducto.

Además, eliminar los componentes que lo limitan implica optar por procesos costosos. ¿Es posible encontrar una estrategia que pueda brindar alternativas de salida a la acumulación excesiva de glicerol crudo en las biorefinerías?

Una posible solución, consiste en primero encontrar una solución tipo bioinformática capaz de crear una molécula con la capacidad de darle un valor agregado a este coproducto que por sus propiedades, no genera mayor beneficio y de paso, reducir el costo de producción del biocombustible.

En este proyecto se busca rediseñar, mediante herramientas computacionales la secuencia primaria de una enzima consenso de tipo glicerol quinasa que pueda convertir el glicerol crudo en un producto de interés comercial y con valor agregado.

En la línea de investigación en la que se enmarca esta investigación, se usará la secuencia consenso obtenida aquí, para construir experimentalmente la enzima y evaluarla en el laboratorio.

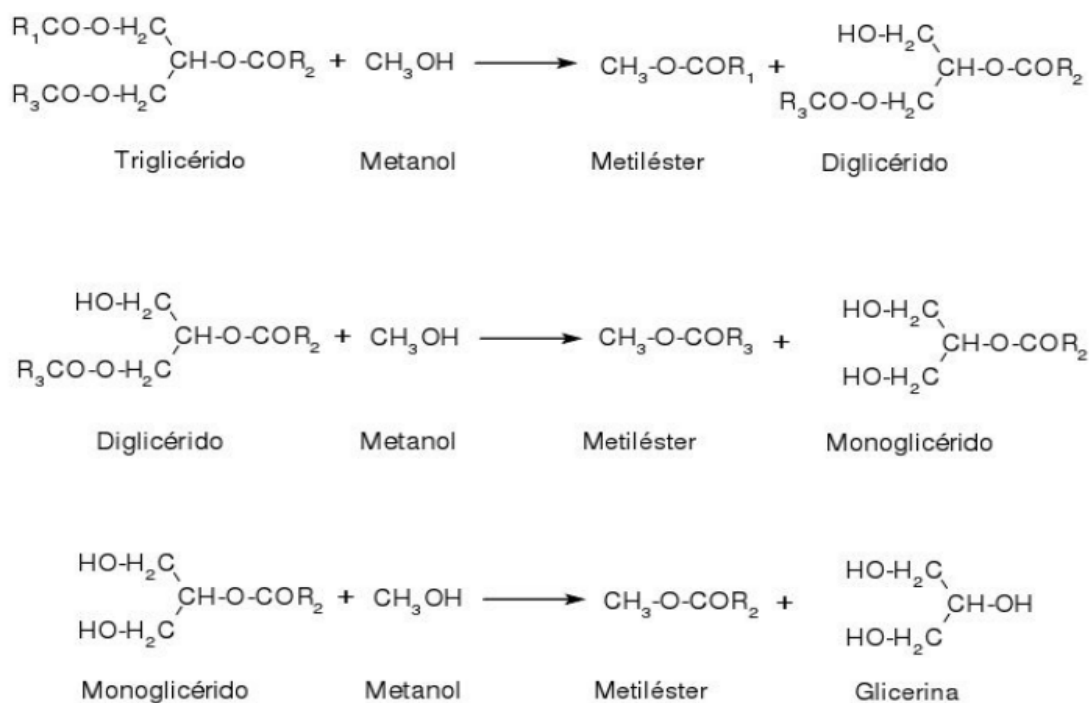
En un futuro, se busca que la enzima consenso pueda suplir las necesidades que demande una empresa que necesite convertir la glicerina cruda en un producto de desecho o en producto de interés industrial de una manera más rápida y efectiva.

2.2. MARCO TEÓRICO Y ESTADO DEL ARTE

La producción de biodiesel hoy en día es llevada a cabo a nivel industrial, usando como materia de producción grasas o aceites vegetales. Para tal propósito, éstos deben ser refinados, por lo cual son sometidos a procesos de desgomado, filtración, neutralización y secado (Valérie Dossata, 1999), con el fin de obtener triacilglicerolos.

Adicionalmente, se requiere de un alcohol ligero o monovalente (p. ej. metanol) que se debe encontrar en exceso y de la presencia de un catalizador básico (esto es un agente que acelere la reacción) que puede ser hidróxido de sodio (Ganduglia, 2009). Todos estos reactivos son sometidos a una transesterificación, es decir, un proceso conformado por tres reacciones reversibles seguidas para formar un éster metílico acompañado de un diglicérido en la primera reacción, un monoglicérido en la segunda reacción y finalmente la glicerina en la tercera reacción (Marchetti, 2007) (ver Fig. 1).

Fig 1: reacciones químicas de la transesterificación.



Fuente: Herguedas, A. I. (2012). BIOMASA, BIOCOMBUSTIBLES Y SOSTENIBILIDAD. Madrid: Centro Tecnológico Agrario y Agroalimentario. ITAGRA.CT.

Una vez terminada la reacción de transesterificación para la producción de biodiesel, el glicerol (GC) que aparece como producto representa un 10% en peso de la producción total del éster metílico (esto es, el biocombustible). En el caso de las biorefinerías, dicho 10% de GC es equivalente a una gran cantidad de galones que deben ser sometidos a procesos adicionales de purificación (García José, 2014).

Las cifras obtenidas por la Federación Nacional de Biocombustibles en Colombia, reportaron el total de toneladas producidas de biodiesel generado a partir de aceite de palma entre los años 2009 y 2014. Dichas cifras evidencian un incremento constante en el tiempo hasta alcanzar un total de 246.158,1 toneladas de GC generado durante estos seis años (Jorge Cardenas Gutierrez, 2003).

Una solución para la problemática que viven las plantas de biocombustibles por excesiva acumulación de GC, es darle una vida útil a este producto de manera que resulte económicamente rentable y que a su vez no tenga efectos adversos para el medio ambiente. Con este propósito, el GC debe ser purificado, es decir, recuperar el glicerol libre de contaminantes. Esto, dado que el GC está compuesto no sólo por aproximadamente 86.95% de glicerol, sino también por 0.03 % de metanol, 3.2 % de cenizas, 3.1% de cloruro de sodio, 0.3% de ácidos grasos y 0.005% de potasio (Muhammad Ayoub, 2012).

A nivel industrial, el proceso de purificación del GC consta, en primera instancia, de su acondicionamiento mediante su dilución en agua. Luego, se procede a neutralizarlo usando ácido sulfúrico, más adelante debe ser filtrado y finalmente, es sometido a una serie de destilaciones a vacío y destilaciones simples para lograr obtener el glicerol puro. Finalmente éste podrá ser sometido a una reacción mediada por una enzima tipo glicerol quinasa. A partir de este proceso, se inicia la generación de un nuevo producto de interés comercial (W. G. Morales, 2010).

Es importante reconocer que la reacción mediada por esta clase de enzimas cataliza la etapa limitante de la velocidad en la principal vía metabólica de utilización de glicerol. Esto se logra mediante la transferencia de un fosfato desde el ATP al glicerol, obteniéndose glicerol 3-fosfato (ver Fig. 2), que es un intermediario importante tanto para el metabolismo energético y la producción de glicolípidos (Cheng Chenga, 2014).

Fig 2: Conversión del glicerol en L-glicerol-3-fosfato por la glicerol quinasa.



Fuente: Müller-Esterl, W. (2008). Bioquímica. Fundamentos para Medicina y Ciencias de la Vida. Barcelona, España: REVERTÉ.

Dentro de las importancias del glicerol-3-fosfato generado por la enzima, es que no sólo participa en la glicólisis, sino que también puede generar mediante una serie de reacciones reversibles, intermediarios del ciclo de Krebs, del cual se pueden obtener ya sea por vía anaeróbica o aeróbica productos de interés industrial tales como los ácidos cítrico, isocítrico, fumárico, málico (Pena, 2004) y entre otros productos de valor agregado.

Por otra parte el glicerol 3-fosfato también sirve como precursor del 1,3 propanodiol, el cual es usado para sintetizar algunos polímeros como el polietilenglicol y el poliéster; los ácidos láctico y acético tienen un alto grado de aplicabilidad en el campo de la industria química, textil, de alimentos, de medicamentos, de plásticos, en la agricultura, entre otros. Por otra parte, el etanol también tiene una amplia gama de aplicaciones de interés industrial en la producción de licores, en la industria farmacéutica, en la síntesis química, en su uso como combustible, entre otros (Duane T. Johnson, 2007).

En esta investigación se busca ofrecer una nueva enzima tipo glicerol quinasa, dada la capacidad que tiene de fosforilar al glicerol para formar al precursor que puede dar inicio a los muchos destinos que puede tener la glicerina.

Gran parte de la importancia de este proyecto se basa en el estudio de distintas secuencias, por esto es importante definir dicho concepto como el orden de distintos aminoácidos unidos por enlaces peptídicos (Ríos, 1985). También se debe diferenciar entre *semilla* o secuencia molde, *consenso* siendo ésta la que muestra cuáles son los elementos comunes en distintas secuencias que cumplen una función común y *mutante* o secuencia que ha sufrido alteraciones genéticas.

Para la fase inicial del proyecto, la base de datos que va a ser consultada es el Protein Data Bank, ya que es una base de datos en línea nutrida por científicos de todo el mundo, que cuenta en su versión de diciembre de 2015 con 10002 secuencias y estructuras de proteínas distintas, las cuales han sido obtenidas bajo

los métodos de cristalografía de rayos X, resonancia magnética nuclear o microscopía electrónica de alta resolución. La principal importancia de esta base de datos es que además de permitir obtener las secuencias de interés, también, brinda los modelos en tercera dimensión, los cuales son necesarios para predecir la estructura 3D de la enzima consenso que se va a diseñar (Helen M. Berman, 2000).

El RCSB PDB es administrado por dos miembros de la Research Collaboratory for Structural Bioinformatic, y operado por la Universidad Rutgers (Universidad Estatal de Nueva Jersey) y el Centro de Supercomputación de San Diego en la Universidad de California (Andrei Kouranov, 2005).

Una vez obtenidas todas las secuencias de las enzimas que hayan cumplido los criterios de selección, se realiza un alineamiento múltiple de secuencias (MSA por sus siglas en inglés), el cual organiza las secuencias de proteínas en una matriz rectangular con el objetivo de que queden en una columna dada los residuos idénticos (derivados de una sola posición en una secuencia ancestral), superponibles (locales rígido en una alineación estructural) o que jueguen un papel funcional (Robert C Edgar, 2006).

La validación de un programa de MSA normalmente utiliza un conjunto de datos de alineaciones de referencia. Un alineamiento producido por el programa se compara con la alineación de referencia correspondiente, dando una puntuación de precisión.

Existen diversos programas que permiten realizar MSA tales como CLUSTALW, MAFFT, MEGA, T-COFFEE y PROBCONS, siendo CLUSTALW el programa más popular, sin embargo no será el elegido para ejecutar las finalidades de este proyecto porque desde mediados de 1994 no se le han realizado mejoras a su algoritmo de ejecución (Edgar, 2004). Debido a esto será empleado MEGA, pues es más rápido y más preciso que CLUSTALW, pues cuenta con una buena compensación de precisión y opciones para correr las secuencias aún más rápido y con una precisión promedio más alta para aplicaciones de alto rendimiento.

MEGA tiene una herramienta conocida como MUSCLE, el cual es un programa informático para realizar múltiples alineamientos de secuencias de proteínas que permite la alineación progresiva mediante una nueva función de perfil llamada *log-expectation score* (Edgar, 2004). Este algoritmo realiza el alineamiento múltiple para un grupo de secuencias de aminoácidos dadas; su estrategia de operación consiste en la ejecución iterativa de tres etapas, a saber: *Draft progressive*, *Improved progressive* y *Refinement*.

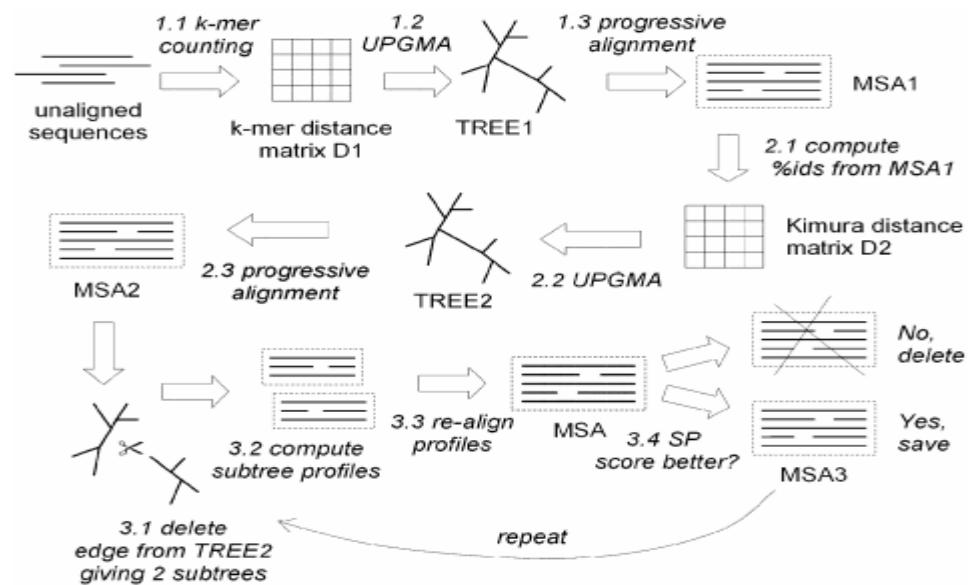
Draft progressive: consiste en realizar un primer alineamiento múltiple haciendo énfasis en la velocidad del cálculo y no en la precisión, donde primero se genera una matriz de distancia creada a partir de las distancias *K-mer* de las entradas (secuencias). A partir de ésta matriz se genera un primer árbol binario que determina el orden de la alineación progresiva.

Improved progressive: realiza un realineamiento de la primera matriz, usando esta vez la distancia *Kimura* (M, 1980), que es más precisa que la distancia *K-mer* de la etapa anterior. Con las distancias calculadas, se genera una segunda matriz de distancia, que a su vez permite la generación de un segundo árbol binario. Como resultado se obtiene un nuevo alineamiento múltiple de secuencias más preciso que el anterior.

Refinement: divide el árbol de la etapa anterior en dos subárboles, mediante un proceso iterativo, que busca mejorar la puntuación en cada iteración, por lo tanto cada iteración produce un realineamiento mejor o igual que el anterior.

En la Fig 3. se resume el algoritmo empleado por MUSCLE para realizar el alineamiento múltiple de secuencias.

Fig 3: Resumen del flujo del algoritmo de MUSCLE.



Fuente: Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research*, 1792-1797.

MEGA (Molecular Evolution Genetic Analysis) es un software informático que contiene instalaciones para la construcción de la secuencia de alineaciones, infiriendo historias filogenéticas, y la realización de análisis evolutivo molecular, además de permitir el uso del programa MUSCLE, también permite la edición y depuración del alineamiento de secuencias biológicas (Koichiro Tamura, 2013).

Otro programa que ejecuta buenas alineamientos de secuencia es MAFFT, pues ofrece diversas estrategias múltiples para la alineación, mediante el uso de tres tipos

de algoritmos tales como, el método progresivo, el método de refinamiento iterativo con la puntuación WSP, y el método de refinamiento iterativo utilizando tanto el PSA y las puntuaciones de consistencia (Kato, 2002).

A continuación, se introducen el mecanismo de acción de estos algoritmos:

Métodos progresivos: FFT-NS-1 y FFT-NS-2

Son métodos sencillos progresivos. FFT-NS-1, primero busca hacer una matriz de distancia aproximada contando el número de tuplas compartidas entre cada par de secuencias, luego, construye una guía de árboles y finalmente alinea las secuencias de acuerdo con el orden de ramificación.

Por otra parte, FFT-NS-2 ejecuta el mismo procedimiento que el FFT-NS-1, pero aumenta dos pasos más, donde el árbol de guía se vuelve a calcular a partir de la alineación FFT-NS-1, y luego, se lleva a cabo una segunda alineación progresiva, la cual es ligeramente superior que la alineación progresiva realizada por FFT-NS-1.

Método de refinamiento iterativo: FFT-NS-i y i NW-NS

Busca mejorar la alineación progresiva explicada anteriormente, donde la alineación realizada por FFT-NS-2 se somete a un proceso de refinamiento iterativo, ésta vez dos algoritmos, FFT-NS-i y i NW-NS. FFT-NS-i, repite el refinamiento iterativo hasta que se llegue a una mejoría en la puntuación del PSA o hasta alcanzar un número de ciclos igual a 1000.

El i NW-NS, realiza sólo dos ciclos más, con el fin de mejorar el procedimiento anterior.

Métodos de refinamiento iterativo utilizando WSP y las puntuaciones de consistencia: L-INS-i, E-INS-i, G-INS-i.

Estas opciones utilizan una nueva función objetivo combinando la puntuación WSP (Gotoh) con la puntuación empleada en T-COFFEE, con el fin de evaluar la coherencia entre una alineación múltiple y alineamientos de pares (Kato et al. 2005).

Una vez realizado el alineamiento tanto en MEGA como en MAFFT, se buscará verificar cuál de estos dos programas brinda un mejor alineamiento, según el mayor número de zonas conservadas que genere cada uno.

Cuando se decida cual alineamiento será empleado, se buscará determinar la secuencia de la enzima tipo consenso, para esto, será empleado Jalview, el cual es un programa gratuito y de fácil acceso, que permite usar un alineamiento de

secuencias múltiples para generar a partir de éste una secuencia consenso (buscar que algoritmo emplea).

Una vez determinada la secuencia consenso de la enzima glicerol quinasa, se procederá finalmente a determinar cuál será su posible estructura 3D usando Swiss-Pdb Viewer, que es una aplicación que permite generar modelos en una plantilla en 3D a partir de una secuencia primaria, donde se espera que tenga una gran similitud con las estructuras en 3D de la secuencia usada para realizar el alineamiento.

Finalmente, se realizará un análisis estructural, para lo cual se usará UCSF chimera, es un programa que cuenta con una herramienta que permite superponer estructuras tridimensionales de distintas proteínas. El algoritmo que emplea éste programa es conocido como MatchMaker, éste comando realiza un ajuste después de identificar automáticamente qué residuos deben ser emparejados, permitiendo superponer estructuras similares o incluso estructuras con baja similitud. Básicamente, éste comando lo que hace es primero generar una alineación de secuencias por pares y luego, realiza la superposición de las estructuras de acuerdo con los alineamientos de pares (Resource for Biocomputing, Visualization, and Informatics , 2015).

La finalidad de este proyecto es que la secuencia consenso optimizada en un futuro sea capaz de someter al glicerol a una reacción de fosforilación y así pueda propiciar la formación de precursores para la producción de compuestos de valor agregado.

2.3. OBJETIVOS

2.3.1. OBJETIVO GENERAL

Generar una secuencia consenso de aminoácidos para una enzima de tipo glicerol quinasa.

2.3.2. OBJETIVOS ESPECÍFICOS

Seleccionar una secuencia a partir de un grupo de enzimas glicerol quinasa del PDB, la cual debe ser determinada por cristalografía de rayos X y tener una resolución menor a 2.5 Å, a partir de esta, identificar una secuencia wildtype que actúe como semilla.

Seleccionar un grupo de glicerol quinasa para hacer un alineamiento múltiple de secuencias en MEGA y MAFFT, e identificar las zonas conservadas y no conservadas para generar una secuencia consenso en JalView.

Realizar un modelado por homología de la secuencia consenso.

Identificar las diferencias estructurales que se puedan presentar entre la estructura de la proteína semilla y la consenso.

2.4. METODOLOGIA

Para llevar a cabo la ejecución de este proyecto, se realizó una búsqueda en el PDB de un grupo de enzimas de tipo glicerol quinasa, las cuales han sido obtenidas por cristalografía de rayos X, método de obtención predilecto, pues, es una técnica muy fiable que además, permite obtener una alta resolución. De los resultados arrojados por esta base de datos, se seleccionó una secuencia reportada como “1GLJ”, con la que, según el artículo que se encuentra enlazado a ésta, es posible conocer la secuencia *WildType (WT)* (Bystrom CE, 1999), donde el término *wiltype* se refiere a cualquier enzima que se encuentra en la naturaleza y no ha sido modificada.

Ésta base de datos es un sitio web de libre acceso y sin ánimo de lucro, que además cuenta con muchas herramientas computacionales y modelos matemáticos similares a las de otras bases de datos, pero, tiene la ventaja de proveer archivos de coordenadas moleculares para un gran número de moléculas que tienen su estructura tridimensional obtenida por los métodos de difracción de rayos-x, resonancia magnética nuclear, microscopía electrónica de alta resolución y otros métodos menos usados.

La secuencia WT de 1GLJ fue introducida en el banco de datos NCBI para seleccionar secuencias con un porcentaje de identidad mayor al 75%, dado que por arriba de este valor, la probabilidad de relación evolutiva es más alta.

Al realizarse la búsqueda de las secuencias, el programa NCBI retornó un total de 1.000 resultados, donde se encontraban secuencias repetidas, secuencias mutadas cuyo aporte al proceso era sustituido por la secuencia original, y secuencias con un porcentaje bajo de similitud. Para filtrar la información se elaboró un algoritmo en JavaScript (lenguaje de programación), el cual seleccionó automáticamente secuencias que cumplan criterios, donde primen la escogencia de distintos organismos, se excluyan modelos y además donde cada una contenga efectivamente su estructura primaria y un artículo enlazado a ellas que afirme que se trata de una secuencia que no ha sufrido modificaciones, es decir, que no se encuentre mutada.

Con las secuencias glicerol quinastas seleccionadas, se realizaron dos alineamientos múltiples. Primero fue empleado MEGA, programa bioinformático que permite hacer alineamientos con MUSCLE (Multiple Sequence Comparison by Log-Expectation), comando que identifica las zonas conservadas y no conservadas de todas las secuencias, después se realizó un segundo alineamiento con el mismo grupo de secuencias en MAFFT (Multiple Alignment using Fast Fourier Transform), que es un programa de alta velocidad y está disponible gratuitamente para uso académico.

Para iniciar con los dos alineamientos tanto en MEGA como en MAFFT, se introdujo la secuencia semilla, la cual es la encargada de definir el tamaño de la enzima

consenso y una vez realizado el AMS, permite identificar las zonas conservadas y las zonas no conservadas presentes en las secuencias seleccionadas.

Para realizar ambos alineamientos se emplearon los mismos parámetros. Los alineamientos fueron realizados usando la matriz BLOSUM62, pues se adopta como una matriz de puntuación por defecto, BLOSUM80, 45, JTT200PAM, 100PAM y Gonnet. También se pidió que se emplearan penalizaciones de huecos o GAPS por defecto para las alineaciones de aminoácidos para el residuo en un hueco un valor de 1,53 y para residuos adicionales en una brecha un valor de 0.123.

Una vez terminados los alineamientos en ambos programas, se optó por emplear el generado por MAFFT, ya que con este se logró obtener uno mejor, pues arrojó un total de 14 residuos alineados más en comparación con el ejecutado por MEGA.

Éste alineamiento fue guardado en formato fasta para luego abrirlo en el programa JalView, que se basa en las predicciones de los seis algoritmos de predicción de estructura secundaria que aprovechan la información evolutiva de múltiples secuencias, mostrando debajo del alineamiento el porcentaje del residuo por columna modal, resultado que es traducido como secuencia consenso.

Ya obtenida la secuencia tipo consenso, se verificó que no fuera idéntica a ninguna de las secuencias del alineamiento, ni que fuera igual a ninguna otra secuencia en el Genbank. Para esto, primero se hizo una búsqueda a modo de comparación entre las 282 secuencias que fueron empleadas para hacer el alineamiento, donde se corroboró que no existían secuencias idénticas. Posteriormente se consultó en el ncbi, donde se encontró que de todas las secuencias arrojadas por el buscador, la más próxima cuenta con un porcentaje de identidad del 95%, valor que corresponde al total de residuos de aminoácidos idénticos de los 501 residuos de la consenso, lo que lleva a inferir que no existen secuencias exactamente iguales reportadas a la fecha en la web, pues ninguna alcanzó un 100% de identidad.

Con respecto a la secuencia consenso generada, se obtuvo una secuencia compuesta por un total de 501 residuos al igual que la semilla, pero incluye 34 mutaciones. Las mutaciones de la consenso con respecto a la secuencia semilla son las siguientes: M21L, I28V, E36T, P42A, E110D, H114K, D122E, S126H, I132V, H147N, S151A, R156E, I170V, E205D, V206A, E212A, S240A, F251Y, E258P, M271L, K278E, K281T, E283K, T295R, A309G, N324S, Q337K, N338D, T339S, A373S, E393D, N456D, A484K, M494Q.

En general no se presentaron numerosos cambios, porque, las mutaciones que generaron cambios significativos representan el 3.4% del total de residuos de la secuencia, razón por la cual se puede inferir que a nivel estructural no se deberían presentar grandes cambios.

Para corroborar el anterior enunciado, se elaboraron los modelos tridimensionales de las secuencias semilla y consenso empleando el programa SWISS-MODEL, donde a simple vista, se logra observar que las estructuras terciarias presentan un alto grado de similitud, sin embargo, se consideró más interesante sobreponer las estructuras entre sí para notar aún mejor los cambios que se encuentren presentes.

Un aspecto que fue considerado en esta instancia, fue que los modelos creados en tercera dimensión debían ser comparados primero con un modelo que haya sido obtenido bajo criterios experimentales, es decir, que se tenga evidencia a partir de un proceso altamente preciso, como lo es la cristalografía de rayos-x para poder hacer una comparación más ajustada a un modelo real.

Una cantidad para expresar la similitud estructural es la distancia de la raíz cuadrática media (rmsd) calculada entre átomos equivalentes en dos estructuras, se encuentra definida como:

$$rmsd = \sqrt{\frac{\sum_i d_i^2}{n}} \quad \text{Ecuación (1)}$$

, donde d es la distancia entre cada uno de los n pares de átomos equivalentes en dos estructuras superpuestas. Si el valor de rmsd es cercano a cero, indica que se trata de estructuras idénticas, mientras que si su valor aumenta, se infiere que las dos estructuras se hacen más diferentes. Los valores de rmsd se consideran como indicadores fiables de la variabilidad estructural (Pongor, 2001).

Se realizaron cuatro sobreposiciones: Fig. 5, 1GLJ vs modelo 1GLJ, 1GLJ vs semilla, 1GLJ vs consenso, semilla vs consenso, a las cuales, se les determinó el valor de rmsd empleando chimera ver Tabla 2, donde, se encontró que la sobreposición de la semilla vs consenso, fue el menor valor, éste fue igual a 0.023 Å, lo cual indica que estas dos secuencias son muy similares.

En un alineamiento de secuencias, es importante determinar el porcentaje de identidad que se refiere, al número de coincidencias de los mismos residuos de aminoácidos entre las dos secuencias alineadas, por esto se determinó dicho valor para la consenso vs la semilla, donde el porcentaje obtenido fue de 93.2%, lo cual indica que son altamente parecidas entre sí.

En referencia a la secuencia consenso, resulta importante exponer de manera visual las diferentes mutaciones resultantes; para esto se generó una representación tridimensional de la estructura, donde las 34 posiciones mutadas se muestran con color rojo, ver Fig. 6.

Finalmente, se determinaron las propiedades que podría presentar la enzima consenso. Para esto, se acudió a expasy, programa que cuenta con un calculador

que permitió determinar las propiedades fisicoquímicas de la secuencia consenso, ver tabla 3.

2.5 RESULTADOS

En primera medida se seleccionó una enzima semilla, ésta fue tomada del artículo Crystal structures of Escherichia coli glycerol kinase variant S58-->W in complex with nonhydrolyzable ATP analogues reveal a putative active conformation of the enzyme as a result of domain motion, porque es una enzima que cuenta con altos parámetros cinéticos, (Bystrom CE, 1999):

Tabla 1: Parámetros cinéticos de enzima WildType de glicerol quinasa y 1GLJ.

Secuencia	V _{máx} (units/mg)	K _{ATP} (μ m)	K _{glicerol} (μ M)	K _{i,ATP} (μ M)
Glicerol quinasa <i>WildType</i>	15.3	8.4	4.9	86
1GLJ	79	23	14	108

Fuente: Bystrom CE, Pettigrew DW, Branchaud BP, O'Brien P, Remington SJ.(1999). Crystal structures of Escherichia coli glycerol kinase variant S58-->W in complex with nonhydrolyzable ATP analogues reveal a putative active conformation of the enzyme as a result of domain motion. Oregon, USA.

Una vez obtenida la secuencia semilla, se realizó una búsqueda en NCBI, la cual arrojó mil resultados, de los que se necesitaba seleccionar un grupo de secuencias aptas para realizar el alineamiento. Debido a que la realización manual de este proceso resulta muy extenuante y propensa a equivocaciones, fue empleado un algoritmo que facilitará dicha selección:

Fig 4: Algoritmo para selección de secuencias.

```
1. Function Select
2.   sequences =: new array
3.   For each sequence s IN results do:
4.     IF NOT(sequences.contains(s)) AND (NOT(isMutated(s)) OR similarity(s) > 95%)
5.     then
6.       sequences.add(s)
7.     End
8.   End
9.   MarkInTable(sequences)
10. End
```

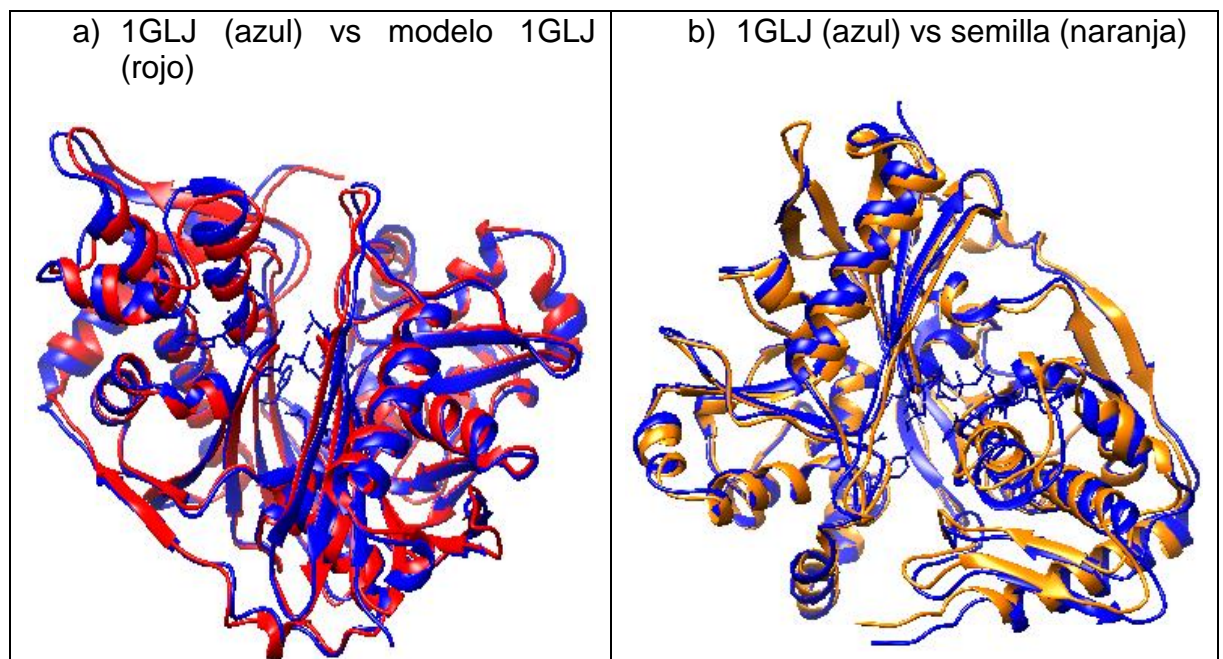
Fuente: autoría propia.

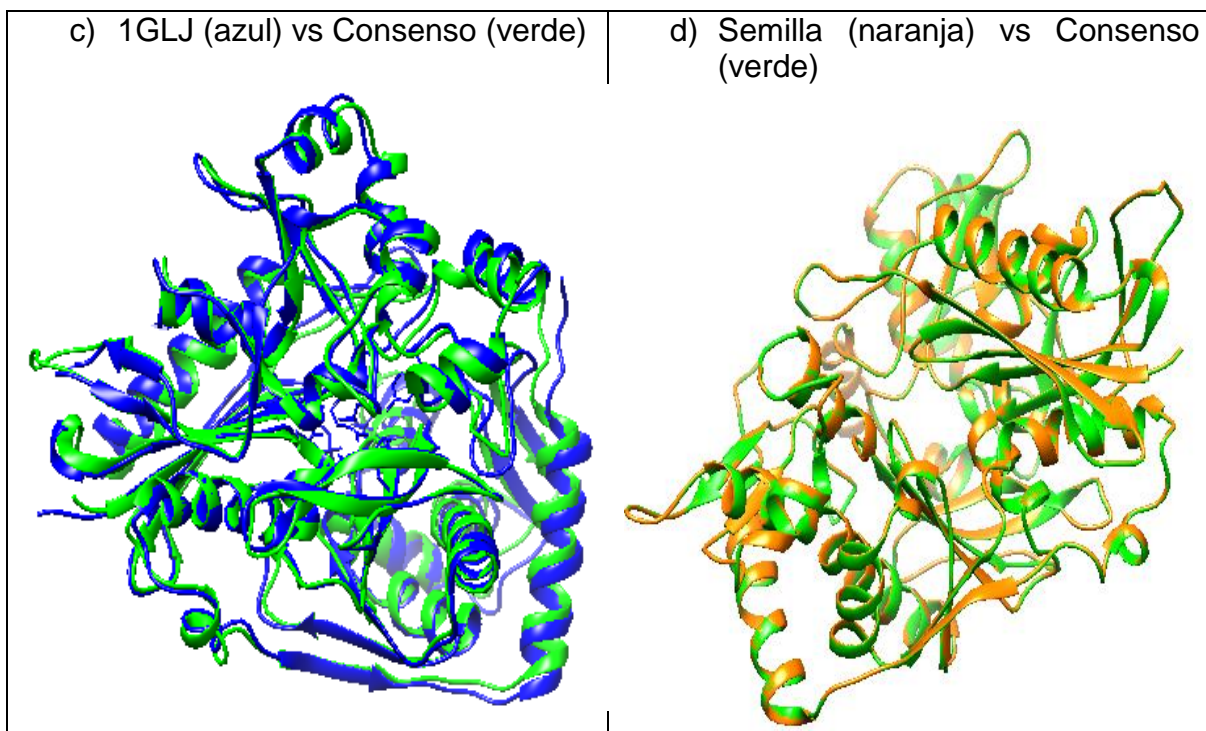
Éste algoritmo hizo una selección sistemática de secuencias que representan a cada uno de los distintos organismos que se encontraban en los resultados, verificó que efectivamente la secuencia se encontrará incluida y finalmente buscó palabras claves en los artículos vinculados a cada una de las secuencias tales como *wildtype* o *no mutant*, con el fin de garantizar que las secuencias seleccionadas no hayan tenido modificaciones genéticas y así lograr obtener un alineamiento con mayor contenido de secuencias naturales conservadas.

Dicho algoritmo permitió escoger un grupo selecto de secuencias, con lo cual se procedió a ejecutar los alineamientos múltiples de secuencias primero en Muscle (herramienta de MEGA), con el fin de realizar una depuración de las secuencias, para luego sí hacer un alineamiento en MAFFT.

Más adelante fueron construidos los modelos tridimensionales de la secuencia semilla, consenso y de la 1GLJ en Swissmodel. Una vez hecho esto, se hizo un alineamiento estructural o superposición de las estructuras 3d.versus la estructura tridimensional de la 1GLJ obtenida directamente del PDB, y que fue resuelta experimentalmente, mediante cristalografía de Rayos X. A continuación son mostradas las comparaciones de las estructuras mencionadas:

Fig 5: Sobreposición de un monómero de la estructura tridimensional de a) 1GLJ vs modelo 1GLJ, b) 1GLJ vs WildType (semilla), c) 1GLJ vs Consenso y d) Semilla vs Consenso.





Fuente: autoría propia

Es importante resaltar que aquí se realizó la superposición de los monómeros en lugar de los dímeros, con el fin de simplificar la visualización de las estructuras tridimensionales y la conservación de las mismas.

En el mismo programa se calcularon los valores de rmsd obtenidos en los alineamientos estructurales mostrados en la Fig. 4, se encuentran enmarcados en la tabla a continuación:

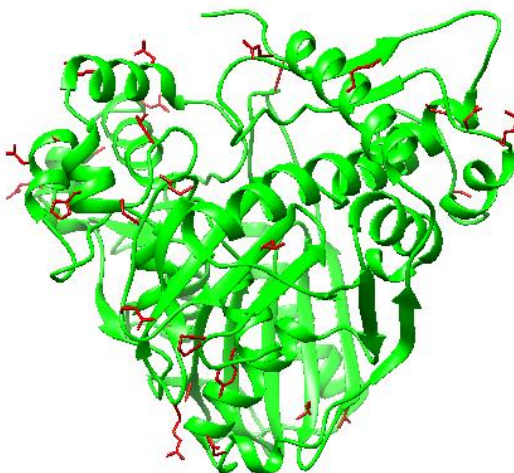
Tabla 2: Valores de rmsd obtenidos para los alineamientos estructurales.

Secuencia	Modelo 1GLJ	Semilla	consenso
1GLJ	0.577Å	0.578 Å	0.577 Å
Semilla	-	-	0.023 Å

Fuente: autoría propia

Ya obtenidos los valores de rmsd para los alineamientos estructurales, se generó un representación visual en chimera de la posición que ocupa cada una de las mutaciones, ver Fig. 6.

Fig 6: Representación de las cadenas laterales (rojo), de las 34 mutaciones presentes en la secuencia consenso, representadas en la estructura 3d de un monómero de la proteína 1GLJ.



En la anterior figura han sido representados los residuos (rojo) que fueron reemplazados de la secuencia semilla, para formar la estructura de la secuencia consenso.

Finalmente fueron determinadas las propiedades fisicoquímicas de la enzima consenso y semilla, para esto fue empleado la herramienta protparam de ExPASy y son mostradas a continuación:

Tabla 3: Propiedades fisicoquímicas de la secuencia consenso y semilla.

Propiedades de la secuencia consenso y semilla		
Propiedad	Consenso	Semilla
Peso molecular	55825.2 g/mol	56099.5 g/mol
Punto isoeléctrico	5.42	5.35
Coeficiente de extinción	280 M ⁻¹ cm ⁻¹	280 M ⁻¹ cm ⁻¹
Índice de inestabilidad	25.76	28.89
Índice alifático	85.2	84.51

Fuente: autoría propia

2.6 DISCUSIÓN

Un paso crucial en el desenvolvimiento de este proyecto fue la escogencia de la secuencia que se iba a tomar como semilla, porque se usó como secuencia inicial para basar en ella la búsqueda de secuencias similares, que fueron usadas en el alineamiento múltiple. Para esto se realizó una búsqueda en el PDB, donde sólo se obtuvieron secuencias que ya habían sufrido modificaciones genéticas. La secuencia semilla a su vez, se empleó como el principal punto de referencia para la obtención de la consenso, porque, permite identificar más fácilmente los cambios que fueron introducidos en la consenso.

Debido a que se trabajó con la secuencia de una enzima wiltype de *Escherichia coli*, que no cuenta con una estructura tridimensional obtenida por cristalografía de rayos-X, se optó por acudir a la secuencia 1GLJ que sólo difiere en el cambio de una Serina por un Triptófano en la posición 58 y sí se tiene la estructura 3D que fue obtenida por dicho método y además tiene altos parámetros cinéticos que permiten especular sobre el nivel de eficacia que podría tener la secuencia consenso para llevar a cabo la fosforilación del glicerol que permite obtener glicerol 3 fosfato. En síntesis, la escogencia de la secuencia 1GLJ, está basada en el hecho que tiene características similares a la semilla. La Tabla 1 muestra los parámetros cinéticos de la semilla y la 1GLJ.

Con esta secuencia se buscaron las secuencias con las cuales se realizó el MSA, el cual en bioinformática es una forma de representar y comparar dos o más estructuras primarias proteicas. De este modo, al compararlas, se resaltan sus zonas de similitud, las cuales podrían indicar relaciones funcionales o evolutivas entre los genes o proteínas consultados. El grupo seleccionado para realizar el MSA debía tener un porcentaje de identidad mayor al 75% para así garantizar que éstas tengan un alto grado de homología (es normal en ortólogos) y así obtener una secuencia consenso que conserve información común a todas las secuencias seleccionadas.

Dentro de los parámetros determinados para llevar a cabo los alineamientos en MEGA y en MAFFT se trabajó con la matriz BLOSUM, ya que es una matriz de sustitución utilizada para el alineamiento de secuencias de proteínas y realiza puntuaciones en los alineamientos entre secuencias de proteínas evolutivamente divergentes, además se basa en alineamientos locales, que generan alineamientos de las zonas más similares de las secuencias. También se asignaron penalizaciones por abrir y extender GAPs de 1,53 y 0.123 respectivamente. La razón por la que se escogió el alineamiento de MAFFT fue porque según una revisión de literatura en la que se comparaban los diferentes programas para alineamiento de secuencias, se encontró que éste es el más confiable y preciso para el alineamiento múltiple de secuencia, además logró alinear 169 residuos de 501, mientras que MEGA alineó sólo 155

Una vez realizado el alineamiento, se usó el programa JalView, para producir una secuencia consenso a partir del MSA. Jalview genera la secuencia consenso a partir del alineamiento, considerando las frecuencias de aparición de cada residuo de aminoácido en cada posición de la proteína, y asigna un puntaje de probabilidad de aparición de cada residuo en la secuencia consenso

Una vez obtenida la consenso, se identificaron 34 mutaciones con respecto a la secuencia que se usó como semilla (la wildtype de E. coli), de los cuales, diecisiete de ellos son conservativas dado que los aminoácidos que fueron reemplazados en la consenso presentan las mismas propiedades que los aminoácidos que se encontraban presentes en la secuencia semilla original, mientras que, las otras diecisiete mutaciones corresponden a aminoácidos con características diferentes a su correspondiente residuo en la misma posición de la secuencia semilla, tal es el caso de las mutaciones H147N, R156E, K278E, E283K, Q337K, que cambiaron propiedades ácidas por básicas o viceversa. También se presentan casos donde las mutaciones M21L, M271L y M494Q, reemplazaron residuos azufrados, otra mutación como la S126H cambia un residuo polar por una básico, lo cual se debe considerar de gran importancia si tal residuo hace parte del corazón hidrofóbico y en general se lograron observar otros cambios en los residuos alifáticos e hidroxilados.

Al comparar dos secuencias en un alineamiento las no coincidencias se interpretan como mutaciones puntuales, debido a la procedencia de un ancestro común, mientras que los huecos o GAPS son mutaciones por adición o delección de residuos introducidas en las secuencias a través del tiempo que haya transcurrido desde que divergieron.

En un MSA, el grado de similitud entre los aminoácidos en posiciones determinadas en la secuencia, se pueden interpretar como una medida aproximada de conservación en una región determinada. Es por esto que cuando se da una sustitución o no sustitución de residuos con propiedades químicas de características semejantes, las secuencias indican que la zona tiene importancia de tipo estructural o funcional.

De las mencionadas mutaciones, se identificó que no se generaron cambios en los aminoácidos del centro activo donde se presenta la actividad catalítica de la enzima, tales residuos son principalmente la His232 que es el encargado de llevar a cabo la fosforilación (Joanne I. Yeh, 2009), y Arg84, Trp104, Tyr136 y Asp246, los cuales permiten que se dé la unión del glicerol mediante la formación de enlaces de hidrógeno (Hagspiel, SF). En general, las mutaciones introducidas en la consenso no son de carácter catalítico, pues se introdujeron residuos que no interaccionan directamente con el sustrato, por tanto se puede inferir que dichos cambios podrían formar parte del andamiaje de la enzima.

Con respecto a la creación de los modelos en 3D, primero se generó el modelo de la secuencia 1GLJ, para compararlo con la estructura obtenida bajo cristalografía de rayos-X que brinda una estructura real que se toma como referencia comparativa para verificar la predicción que se genera con el modelo, éste resultado mostró que, la predicción presenta características muy similares a la estructura real obtenida por medios experimentales. Del mismo modo se procedió con las secuencias semilla y consenso, para comparar tales modelos con una estructura determinada bajo criterios experimentales. Esto se ejecutó con el único fin de comparar los modelos generados con uno que permita inferir si realmente el resultado obtenido es acorde con los ya existentes.

Para hacer un análisis de los resultados mostrados en la Fig.4, primero se iniciará explicando que en el literal a, se muestra la sobreposición de la estructura 3D del modelo reportado por el PDB con el modelo generado por swissmodel, esto con el fin de identificar la proximidad de predicción del modelo creado por el programa, con respecto a la estructura obtenida por un método experimental muy preciso como el ya mencionado. En el literal b, se ha sobre puesto la secuencia mutada 1GLJ encontrada en el PDB con la secuencia semilla, este resultado es un tanto peculiar, dado a que la secuencia 1GLJ sólo tiene una mutación con respecto a la semilla y corresponde al reemplazo en la posición 58 de una Serina de carácter polar por un Triptófano que es aromático de carácter medianamente apolar, sin embargo a nivel estructural se perciben varios pequeños cambios con respecto a los bucles presentes.

Por otra parte en el literal c se compara la estructura de la 1GLJ con el modelo generado de la secuencia consenso, en este caso, ocurre un cambio muy similar al presentado en el literal b, aunque la consenso posea más mutaciones con respecto a la 1GLJ, finalmente, en el literal d, se muestra una perfecta sobreposición de la secuencia consenso con la que usamos como semilla (wt de E. coli).

Seguido a esto, fueron calculadas las propiedades fisicoquímicas de la secuencia consenso y semilla usando la plataforma Expasy, como se puede notar en la tabla 3, no se presentan grandes diferencias en dichas secuencias, lo que nos da confianza en la proteína consenso obtenida, y nos ofrece parámetros a tener en cuenta tanto en los ensayos enzimáticos, como en la purificación de la proteína, una vez se pase a la fase de experimentación en wet-lab.

Finalmente El proyecto en las fases venideras pasa a un campo aplicativo, donde se va a realizar la síntesis y la expresión de la enzima Glicerol Quinasa, para llevar a cabo la evaluación experimental de la actividad y las propiedades fisicoquímicas de la misma.

2.7. CONCLUSIONES.

Se realizaron dos alineamientos múltiples de secuencias glicerol quinasa en MEGA y en MAFFT de los cuales se escogió el alineamiento realizado por MAFFT, pues, arrojó un mayor número de residuos alineados.

Se obtuvo una secuencia tipo consenso generada a partir de un alineamiento de múltiples secuencias de enzimas glicerol quinasa, que presenta 34 mutaciones con respecto a la semilla (wildtype).

Se modeló la estructura tridimensional de la secuencia tipo consenso, la cual como se había predicho, tuvo un alto grado de similitud con la secuencia empleada como semilla.

Las secuencias semilla y consenso presentaron un porcentaje de identidad del 93.2%, lo cual representa un grado de coincidencias de los aminoácidos de ambas secuencias, e implica 34 residuos de aminoácido diferentes entre la secuencia de la enzima wildtype y nuestra enzima consenso.

2.8 RECOMENDACIONES

En esta investigación se han empleado herramientas computacionales, que han permitido la ejecución de los objetivos de la misma. Dado a que todos los comandos que se emplearon son de carácter estándar y sólo son modificados cada que sale una versión actualizada, es de esperar que si se emplean los mismos parámetros empleados para la culminación del proyecto, es de esperar que los resultados obtenidos sean fácilmente replicables, incluso pueden ser modificados si los estudios demuestran que la generación de cambios en los parámetros, se puede lograr obtener resultados aún mejores.

Se espera que se construya experimentalmente la enzima que obtuvimos computacionalmente como proteína consenso, para evaluar sus propiedades fisicoquímicas y cinéticas, con el fin de considerar su aplicación en la transformación del glicerol.

Como medio de obtención computacional de los parámetros cinéticos de la enzima consenso, se sugiere la base de datos Netlogo MODELING COMMON, que sirve como simulador para estudiar la cinética enzimática de un sistema Michaelis de único sustrato.

Si en un futuro se logra obtener la enzima por medio experimentales de laboratorio, y se determina que sus propiedades cinéticas pueden ser superadas si se reemplazan ciertos residuos determinantes en la actividad catalítica, es posible realizar el diseño a nivel computacional y crear una nueva secuencia que tenga un mejor grado de eficacia.

3. REFERENCIAS

- Andrei Kouranov, L. X. (2005). The RCSB PDB information portal for structural genomics. *Nucleic Acids Research*, D302-D305.
- Bystrom CE, P. D. (1999). Crystal structures of Escherichia coli glycerol kinase variant S58-->W in complex with nonhydrolyzable ATP analogues reveal a putative active conformation of the enzyme as a result of domain motion. *Biochemistry*, 3508-18.
- Cheng Chenga, F. Y. (2014). Identification of the glycerol kinase gene and its role in diapause embryo restart and early embryo development of Artemia sinica. *Gene*, 51–62.
- Devlin, T. M. (2004). *Bioquímica: libro de texto con aplicaciones clínicas*. Barcelona, España: REVERTÉ, S.A.
- Duane T. Johnson, K. A. (2007). The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environmental Progress and sustainable energy*, 338–348.
- Dufey, A. (2006). Producción y comercio de biocombustibles y desarrollo sustentable: los grandes temas. 1-2.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research*, 1792-1797.
- Ganduglia, F. (2009). *Manual de Biocombustibles*. Uruguay: ARPEL.
- García José, G. M. (2014). Glycerol based solvents: synthesis, properties and applications. *Green Chemistry*, 16.
- Hagspiel, L. (SF). <https://collab.itc.virginia.edu>. Recuperado el 02 de 11 de 2015, de <https://collab.itc.virginia.edu>:
https://collab.itc.virginia.edu/access/content/group/f85bed6c-45d2-4b18-b868-6a2353586804/P/Ch22_Hagspiel_L_Glycerol_kinase_complex_glycerol_--Enterococcus_casseliflavus_-_/Ch22_Hagspiel_L_Glycerol_kinase_complex_glycerol_--Enterococcus_casseliflavus_-_G
- Hall, T. A. (1999). BioEDit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium*, 95-98.

- Helen M. Berman, J. W. (2000). The Protein Data Bank. *Nucleic Acids Research*, 235–242.
- Herguedas, A. I. (2012). *BIOMASA, BIOCOMBUSTIBLES Y SOSTENIBILIDAD*. Madrid: Centro Tecnológico Agrario y Agroalimentario. ITAGRA.CT.
- Jan Koolman, K.-H. (2004). *Bioquímica: texto y atlas*. Madrid, España: Panamericana.
- Joanne I. Yeh, R. K. (2009). Structural Characterizations of Glycerol Kinase: unraveling phosphorylation-induced long-range activation. *Biochemistry*, 346–356.
- João R M Almeida, L. C. (2012). Biodiesel biorefinery: opportunities and. *Biotechnology for Biofuels*, 1.
- Jorge Cardenas Gutierrez, J. B. (2003). *Federación Nacional de Biocombustibles*. Bogotá D. C.
- José I. García, H. G.-M. (2014). Glycerol based solvents: synthesis, properties and applications. *Green Chemistry*, 16.
- Katoh, K. (21 de Agosto de 2002). *MAFFT*. Recuperado el 18 de noviembre de 2015, de MAFFT: <http://mafft.cbrc.jp/alignment/software/algorithms/algorithms.html>
- Koichiro Tamura, G. S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 2725-2729.
- M, K. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. . *Journal of Molecular Evolution*, 16:111-120.
- Marchetti, J. M. (2007). Possible methods for biodiesel. *Renewable and Sustainable Energy Reviews*, 1300-1311.
- Muhammad Ayoub, A. Z. (2012). Critical review on the current scenario and significance of crude glycerol resulting from biodiesel industry towards more sustainable renewable energy industry. *Renewable and Sustainable Energy Reviews*, 2671–2686.
- Müller-Esterl, W. (2008). *Bioquímica. Fundamentos para Medicina y Ciencias de la Vida*. Barcelona, España: REVERTÉ.
- Ocampo CB; Ferro C; Cadena H; Marín D; Lozano L; Ramírez CA; Munstermann L. . (2013). *jerajeraejjsdfjao. acta tropica*, 27- 30.

- Pena, A. (2004). *Bioquímica*. México D.F: Limusa S.A.
- Pongor, O. C. (2001). A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Science*, 1470–1473.
- Resource for Biocomputing, Visualization, and Informatics . (23 de julio de 2015). *UCSF Chimera*. Recuperado el 18 de noviembre de 2015, de UCSF Chimera: <https://www.cgl.ucsf.edu/chimera/current/docs/UsersGuide/framecore.html>
- Robert C Edgar, S. B. (2006). Multiple sequence alignment. *Current Opinion in Structural Biology*, 368–373.
- saldarriaga, c; Valderrama, B; . (2014). jraj fjaoidsjiadu. xxxxx, 37-45.
- Teresa K. Attwood, D. J.-S. (2002). *Introducción a la bioinformática*. Madrid, España: Pearson Educación.
- Valérie Dossata, D. C. (1999). Continuous enzymatic transesterification of high oleic sunflower oil in a packed bed reactor: influence of the glycerol production. *ELSEVIER*, 194-195.
- W. G. Morales, N. P. (2010). PURIFICACIÓN DE GLICEROL OBTENIDO A PARTIR DE SÍNTESIS DE BIODIESEL DE ACEITE. *Avances en Energías Renovables y Medio Ambiente* , 6.25-6.27.