

**UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**COLEGIO DE CIENCIAS BIOLÓGICAS Y AMBIENTALES**

**Caracterización del proteoma del virus de mosaico del  
babaco**

Proyecto de Investigación

**Brigitte Aracelly Garzón Proaño**

**Ingeniería en Procesos Biotecnológicos**

Trabajo de titulación presentado como requisito  
para la obtención del título de  
Ingeniera en Procesos Biotecnológicos

Quito, 11 de diciembre de 2018

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ  
COLEGIO DE CIENCIAS BIOLÓGICAS Y  
AMBIENTALES

HOJA DE CALIFICACIÓN DE TRABAJO DE TITULACIÓN

**Caracterización del proteoma del virus de mosaico del babaco**

**Brigitte Aracelly Garzón Proaño**

Calificación:

Nombre del profesor, Título académico:

Miguel Ángel Méndez, PhD.

Firma del profesor:

\_\_\_\_\_

Quito, 11 de diciembre del 2018

## Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma de la estudiante: \_\_\_\_\_

Nombres y Apellidos: Brigitte Aracelly Garzón Proaño

Código: 00112007

Cédula de identidad: 1723269393

Lugar y Fecha: Quito, 11 de diciembre del 2018

## **DEDICATORIA**

Para Marlene, Fabián y Santiago.

## **AGRADECIMIENTOS**

A Dios por bendecirme en cada paso que he dado desde que empecé este sueño.

A mis padres y hermano por su amor y apoyo incondicional a lo largo de mi vida. Ustedes son y siempre serán mi inspiración, y el tesoro más lindo que tengo. Los amo mucho.

A todos los profesores que han contribuido en mi formación académica, en especial a Miguel Ángel por todos los conocimientos impartidos y por brindarme su amistad.

A todos mis amigos con quienes he compartido momentos especiales los cuales han hecho que se conviertan en personas importantes en mi vida: Jorge Álvarez, Andrea Argudo, Carolina Armijos, Ma. José Baquero, Rebeca Cortez, Diana Mollocana, Daniela Orbe, Miguel Orellana, Juan Salazar y Carlos Vintimilla.

## RESUMEN

El virus del mosaico babaco (VMB) fue recientemente reportado gracias a estudios realizados a nivel de secuencia permitiéndonos conocer la similitud que tiene con el virus del mosaico de la papaya. Sin embargo, no hay información sobre la similitud que hay a nivel de estructura, es por ello que en este estudio se procedió a determinar su modelo estructural tridimensional y caracterizar las funciones de las proteínas del VMB (CP, RdRp, Helicasa, Metiltransferasa, AlkB, TGB1, TGB2 y TGB3), mediante el uso de herramientas bioinformáticas. Para esto, para cada proteína se realizó una predicción del modelo de estructura tridimensional y un análisis de confianza de su calidad. Luego, se hizo un análisis de desorden para ver si las proteínas tenían regiones inestables y, de esta manera, tener información sobre la fluctuación espontánea de estas regiones entre diferentes rangos conformacionales en la dinámica molecular, lo que nos da una idea de la flexibilidad de las proteínas. De la misma manera, se analizó la coevolución para saber si las proteínas tenían residuos que mutaban simultáneamente. Después, se usó el paquete Gromacs para llevar a cabo la dinámica molecular de cada una de las proteínas, ya que es un paso preparatorio para la predicción de funciones y también para confirmar los datos de flexibilidad del modelo. Sexto, se realizó un análisis de la estructura para refinar dichas estructuras y comparar los modelos estructurales anteriores y los obtenidos después de la dinámica molecular. Finalmente, la predicción de las funciones se realizó por similitud estructural utilizando herramientas como Dalí, UniProt y PDBe. Obtuvimos que, estructuralmente, las proteínas tienen similitud con las proteínas no solo de virus sino también de bacterias. Un resultado de esto, fue la proteína AlkB, la cual tenía similitud a nivel estructural con el dominio Cupin de *Bordetella pertussis* (2fqp).

**Palabras clave:** Virus de mosaico del babaco, Proteína de cobertura (CP), ARN-polimerasa-dependiente de ARN(RdRp), Helicasa, Metiltransferasa, AlkB, Triple Gene Block (TGB).

## ABSTRACT

The babaco mosaic virus (BMV) was recently reported thanks to studies at the level of sequence, resulting in similarity with the Papaya mosaic virus. However, in this study we proceeded to determine the three-dimensional structural model and characterize the functions of BMV proteins (CP, RdRp, Helicase, Methyltransferase, AlkB, TGB1, TGB2 and TGB3) through the use of bioinformatics tools. For this, for each protein a prediction of the three-dimensional structure model and confidence analysis of their quality was made. Then, a disorder analysis was performed to see if the proteins had unstable regions and in this way have information about the spontaneous fluctuation of these regions between different conformational ranges in the molecular dynamics, giving us an idea about the flexibility of the proteins. In the same way, a coevolution analysis was done to know if the proteins had residues that mutate simultaneously. In addition, the Gromacs package was used to carry out the molecular dynamics of each of the proteins since it is a preparatory step for the prediction of functions and also to confirm model flexibility data. Sixth, a structure analysis was done to refine structures and compare structural models before and those obtained after molecular dynamics. Finally, the function prediction was made by structural similarity using tools such as Dali, UniProt and PDBe. We obtained that, structurally, proteins have similarity with proteins not only of viruses but also of bacteria. One result of this was the AlkB protein, which was found to have structural similarity to the Cupin domain of *Bordetella pertussis* (2fqp).

**Key words:** Babaco mosaic virus, Coverage protein (CP), RNA-polymerase-dependent RNA (RdRp), Helicase, Methyltransferase, AlkB, Triple Gene Block (TGB).

# TABLA DE CONTENIDOS

<b>RESUMEN</b> .....	<b>6</b>
<b>ABSTRACT</b> .....	<b>7</b>
<b>INTRODUCCIÓN</b> .....	<b>12</b>
<b>ÁREA DE ESTUDIO</b> .....	<b>18</b>
<b>JUSTIFICACIÓN</b> .....	<b>19</b>
<b>OBJETIVOS</b> .....	<b>20</b>
<b>MATERIALES Y MÉTODOS</b> .....	<b>21</b>
<b>Secuencia y estructura</b> .....	<b>21</b>
<b>Análisis de Desorden</b> .....	<b>21</b>
<b>Análisis de Coevolución</b> .....	<b>22</b>
<b>Análisis de calidad de estructuras</b> .....	<b>22</b>
ERRAT .....	22
Verify3D.....	22
Procheck .....	23
<b>Dinámica molecular</b> .....	<b>23</b>
<b>Análisis de estructura</b> .....	<b>24</b>
<b>Predicción de la función de las proteínas en base a su similitud estructural</b> .....	<b>24</b>
<b>RESULTADOS</b> .....	<b>26</b>
<b>Secuencia y estructura</b> .....	<b>26</b>
<b>Análisis de Desorden</b> .....	<b>27</b>
<b>Análisis de Coevolución</b> .....	<b>27</b>
<b>Dinámica molecular</b> .....	<b>28</b>
RMSD.....	28
RMSF .....	28
Cálculo de Energía .....	28
<b>Clusters</b> .....	<b>29</b>
<b>Análisis de calidad</b> .....	<b>29</b>
ERRAT .....	29
Verify3D.....	29
Procheck .....	30
<b>Análisis de estructura</b> .....	<b>30</b>
<b>Predicción de la función en base a la similitud estructural</b> .....	<b>31</b>
Proteína de Cobertura (CP) .....	31
Dominio RdRp.....	31
TGB1 .....	31
Helicasa .....	31
AlkB- domain .....	32
<b>DISCUSIÓN</b> .....	<b>33</b>
<b>Secuencia y estructura</b> .....	<b>33</b>
<b>Análisis de Desorden</b> .....	<b>34</b>
<b>Dinámica molecular</b> .....	<b>34</b>

RMSD.....	34
RMSF .....	35
<b>Clusters de las configuraciones obtenidos mediante la Dinámica Molecular.....</b>	<b>35</b>
<b>Análisis de calidad de los modelos.....</b>	<b>36</b>
ERRAT .....	36
Verify3D.....	37
Procheck .....	38
<b>Análisis de Coevolución.....</b>	<b>38</b>
<b>Predicción de la función de las proteínas en base a su similitud estructural.....</b>	<b>39</b>
<b>CONCLUSIONES .....</b>	<b>42</b>
<b>RECOMENDACIONES .....</b>	<b>43</b>
<b>REFERENCIAS .....</b>	<b>44</b>

## ÍNDICE DE TABLAS

Tabla No. 1 Valores de la fiabilidad y calidad de los modelos predichos de la estructura tridimensional de cada una de las proteínas del virus de Mosaico del Babaco .....	51
Tabla No. 2 Análisis de Desorden de las proteínas del Virus de Mosaico del Babaco empleando la herramienta DISOPRED3 del del grupo bioinformático Bloomsbury de la UCL .....	52
Tabla No. 3 Análisis de coevolución de las proteínas del virus del babaco .....	54
Tabla No. 4 Análisis de conservación de las secuencias de las proteínas del Virus de Mosaico del Babaco .....	55
Tabla No. 5 RMSF de las diferentes proteínas del virus de mosaico del babaco .....	57
Tabla No. 6 Cálculo de energía para cada una de las proteínas del virus de mosaico del babaco durante la dinámica molecular.....	57
Tabla No. 7 Análisis de Calidad Errat para las proteínas del BMV .....	57
Tabla No. 8 Análisis de calidad de las distintas proteínas de BMV a partir del indicador Verify 3D .....	58
Tabla No. 9 Análisis de calidad de las distintas proteínas de BMV a partir del indicador Procheck .....	59
Tabla No 10. Análisis de calidad Procheck mediante diagramas Ramachadan de las proteínas del Virus en Mosaico del Babaco antes y después de la dinámica molecular .....	59
Tabla No. 11 Visualización de los modelos de cada una de las proteínas del BMV antes y después de la dinámica molecular .....	62
Tabla No. 12 Alineamiento entre las proteínas del virus de mosaico del babaco antes y después de realizar la dinámica molecular empleando la plataforma TM-align.....	67
Tabla No. 13 Superposición de las proteínas del virus de babaco del masaico antes y después de realizar la dinámica molecular.....	67

## ÍNDICE DE FIGURAS

Figura No. 1 Diagrama PSIPRED para la proteína TGB2 donde muestra la predicción de la estructura secundaria. ....	70
Figura No. 2 Diagrama PSIPRED para la proteína TGB3 donde muestra la predicción de la estructura secundaria. ....	70
Figura No. 3 Diagrama PSIPRED para la proteína Metiltransferasa donde muestra la predicción de la estructura secundaria.....	71
Figura No. 4 Diagrama PSIPRED para la proteína TGB2-helix donde muestra la predicción de la estructura secundaria. ....	71
Figura No. 5 Diagrama PSIPRED para la proteína TGB3-helix donde muestra la predicción de la estructura secundaria. ....	72
Figura No. 6 Análisis de desorden intrínseco de la proteína TGB2 .....	72
Figura No. 7 Análisis de desorden intrínseco de la proteína TGB3 .....	73
Figura No. 8 Análisis de desorden intrínseco de la proteína TGB2-helix .....	73
Figura No. 9 Análisis de desorden intrínseco de la proteína TGB3-helix .....	74
Figura No. 10 Desviación cuadrática media de los diferentes frames comparada en diferentes tiempos para las proteína CP, RDRP, TGB1, Helicasa y Alk-domain después de la dinámica molecular .....	75
Figura No. 11 Análisis de Calidad ERRAT para cada una de las proteínas del virus de Mosaico del Babaco antes y después de realizar la Dinámica Molecular. ....	75
Figura No. 12 Análisis de Calidad Verify 3d para cada una de las proteínas del virus de Mosaico del Babaco antes y después de realizar la Dinámica Molecular. ....	76

## INTRODUCCIÓN

El babaco (*Vasconcellea* × *heilbornii*) es una planta nativa de la zona subtropical del Ecuador, principalmente de la provincia de Loja. Puede crecer incluso a alturas que sobrepasan los 2000 m.s.n.m. y se caracteriza por ser la más tolerante al frío en comparación a las otras especies. Es un híbrido natural entre el toronche (*Vasconcellea stipulata*) y el chamburo (*Vasconcellea cundinamarcensis*), por lo que es un arbusto que no sobrepasa los 8 metros de altura y su fruto se diferencia de la papaya por ser más angosto (Soria & Viteri, 1999). El cultivo del babaco tiene importancia para el sector agrícola ya que no necesita de amplias superficies y tiene un alto rendimiento. El fruto del babaco se caracteriza principalmente por su sabor, alto valor nutricional y se ha demostrado sobre su contenido de enzimas proteolíticas y lipolíticas las cuales son de interés industrial debido a que se pueden usar como catalizadores en diversos procesos metabólicos (Proaño, 2007).

En nuestro país, el cultivo de babaco se realiza en varias provincias tales como Pichincha, Cotopaxi, Tungurahua y Azuay, pero debido a su versatilidad ahora su distribución geográfica se extiende incluso en países europeos como Reino Unido (Robles *et al.*, 2016). Sin embargo, la calidad de este cultivo se ha visto afectada por la infección de fitopatógenos, incluyendo nemátodos, bacterias, hongos y virus; lo cual ha provocado significativas pérdidas. Los virus han llamado la atención pues actualmente no hay suficiente información de la forma de prevenirlos ni tratarlos y son los que más afectan a la producción de la fruta. Se ha reportado que pertenecen a los géneros *Potexvirus*, *Tombusvirus* y *Tobavirus* (Guillén, 2017).

Los *Potexvirus* pertenecen a una nueva familia de virus (Alphaflexiviridae) que infectan a plantas. Su virión se caracteriza por ser flexible, filamentoso, no tiene envoltura, tiene una longitud de 470- 1000 nm. o más y un diámetro de 12-13 nm. (Leiva, 2015). Respecto a su material genético, se caracterizan por ser virus de ARN lineal monocatenario de sentido positivo, aproximadamente de 5.9 - 7 Kb de longitud, con un extremo Cap-5' y otro extremo 3' poliadenilado (Loebenstein & Lecoq, 2012). Su genoma se encuentra organizado en cinco marcos de lectura abiertos (Open Reading Frames), los cuales son evolutivamente conservados, codifican para diferentes proteínas y se encuentran flanqueados por dos regiones no traducidas (UTRs) (García *et al.*, 2014). El primer cuadro de lectura abierto codifica la ARN polimerasa putativa dependiente de ARN (RdRp), la cual a su vez contiene cuatro

dominios bien conservados: metiltransferasa, AlkB, helicasa, y RdRp (Álvarez *et al.*, 2017). El dominio metiltransferasa se ubica cerca al extremo N de la replicasa y actúa como un catalizador del RdRp formando parte del mecanismo de protección del ARN contra exonucleasas (Álvarez *et al.*, 2017). Después se encuentra el dominio AlkB el cual está implicado en la reparación del material genético viral y protección del ARN viral contra los compuestos metilantes del medio ambiente (Bratlie & Drabløs, 2005). A continuación, está el dominio de la helicasa la cual interviene en la replicación del ARN viral debido a su función motora (Jankowsky & Fairman, 2007). El dominio RdRp se encuentra cerca del extremo C de la replicasa y su función es catalizar la replicación del ARN a partir de un template de ARN (Venkataraman *et al.*, 2018). El segundo, tercer y cuarto cuadro de lectura abierto abarca las proteínas bloque de triple gen (TGB) las cuales se superponen parcialmente y codifican para: TGB1, TGB2 y TGB3. Se encuentran involucradas en el movimiento del virus de una célula a otra y a larga distancia. TGB1 contiene una región de dominio similar a la helicasa con tres residuos de arginina conservados que son necesarios para la actividad ATPasa de dicha proteína, unión de ARN y el movimiento de célula a célula. Se encuentra ubicada en el citoplasma y en el núcleo, e interactúa con la proteína de cobertura (CP) para encapsular el ARN viral. En cambio, TGB2 y TGB3 son proteínas integrales de la membrana pues se encuentran asociadas con el retículo endoplasmático y gránulos móviles (Verchot *et al.*, 2010). En el último cuadro de lectura abierta se encuentra la proteína de cobertura (CP) la cual protege el genoma viral de la degradación durante la transmisión y replicación del virus, interviene en el ensamblaje del virus e incluso activa o desactiva los mecanismos de defensa de la planta hospedera (Bol, 2008).

Los *Potexvirus* poseen un virión infeccioso que sirve como genoma y ARN mensajero viral. La replicación de estos virus consiste en que infectan a la planta penetrando en la célula huésped y liberando el vARN en el citoplasma (Lough *et al.*, 2006). Después, el vARN (+) se traduce en mARN monocistrónico y producen RdRp codificada por el virus usando maquinaria de traducción del hospedador. La replicación ocurre en fábricas virales reclutadas por TGB1. A partir de ssARN (+) se sintetiza un dsARN. El genoma de dsARN se replica, proporcionando así mARN y nuevos ssARN (+). Posteriormente, se utilizan promotores subgenómicos internos para que el sgARN se traduzca en proteínas de la cápside (CP) y del movimiento, y se ensamblan las nuevas partículas víricas (Park, 2014). Después de la

replicación, el (+) vARN pasa a ser parte de un complejo de movimiento de ribonucleoproteína (RNP) el cual también está conformado por las proteínas TGB1 y CP para el transporte del material genético. Dicho (+) vARN se encuentra total o parcialmente encapsidado por subunidades CP, y TGB1 se une a su extremo 5'. Posteriormente, la proteína TGB1 interactúa con TGB2 la cual induce a la formación de vesículas para el complejo RNP. Algunos estudios recientes sugieren que las vesículas son producidas por el retículo endoplasmático y se mueven a lo largo de la red de actina hacia la periferia de la célula (Verchot & Bamunusighe, 2007). TGB2 emite un mensaje para que TGB3 pueda inducir el transporte del RNP desde el retículo endoplasmático perinuclear al cortical. Finalmente, el complejo RNP conducido por TGB1 dirige los viriones hacia el poro del plasmodesmo (Tilsner *et al.*, 2009). Mediante experimentos de microinyección se ha demostrado que solo TGB1, pero no CP, puede desencadenar una activación del plasmodesmo, lo que sugiere que estas proteínas actúan juntas dentro de estos canales para promover la expansión del poro y permitir el transporte del virus de célula a célula (Verchot & Bamunusighe, 2007).

Dentro de los *Potexvirus*, se ha reportado que hay un virus que infecta al babaco y provoca varios síntomas en las plantas tales como mosaicos, manchas en forma de anillo, enanismo e incluso hay casos asintomáticos. Hasta el momento, en una investigación realizada por Álvarez *et al.* (2017) han llamado a este fitopatógeno virus de mosaico del babaco (secuencia del genoma y anotaciones completas están disponibles en NCBI GenBank: número de acceso MF978248) y se ha demostrado que tiene similitud a nivel de secuencia con el virus de mosaico de la papaya. Además, se ha reportado que este virus conserva las ocho proteínas propias de los *Potexvirus* repartidas en los cinco cuadros abiertos de lectura. Sin embargo, todavía no se cuenta con información computacional ni experimental de todas las proteínas, como se puede verificar en la base de datos UniProt, no se encuentran anotadas computacionalmente las proteínas metiltransferasa, helicasa y AlkB. (Karplus & McCammon, 2002).

La replicación de los *Potexvirus* es conservada y por ende para el virus de mosaico del babaco ocurre de una manera similar en donde la infección de la planta hospedera se origina por una serie de interacciones entre las proteínas virales y del huésped (Song *et al.*, 2007). Dichas interacciones se encuentran relacionadas no solo por las moléculas inmunes de la planta huésped y los compuestos antivirales, sino también por las mutaciones en proteínas

virales. Entonces, para comprender de una manera más amplia, la forma en la que dichas interacciones proceden mecánicamente y cómo están influenciadas por mutaciones, es necesario conocer las estructuras y la dinámica de las proteínas (Tarasova & Nerukh, 2018). La dinámica molecular (MD) es un método de simulación computacional empleado para analizar los movimientos de proteínas a escala atómica a través de principios teóricos y empíricos en química física. Los recientes avances en el hardware y software para la simulación biomolecular han mejorado rápidamente la precisión y el rendimiento de esta técnica. La dinámica molecular se está extendiendo rápidamente y actualmente tiene bastante aplicabilidad en el campo de la biología, ayudando a revelar características únicas de estructuras de proteínas que serían difíciles de obtener solo con métodos experimentales (Epa & Tran, 2012).

Las proteínas fluctúan espontáneamente en solución. La evidencia acumulada indica que dichas fluctuaciones juegan un papel clave en las funciones específicas de las proteínas, como las reacciones catalíticas de las enzimas (Voth, 2016). A lo largo de los años han existido varios métodos experimentales que se han empleado para caracterizar la dinámica de las proteínas, como la microscopía de fuerzas atómica, transferencia de energía de resonancia de Förster, radiación infrarroja y resonancia magnética nuclear. Sin embargo, llegaron a ser técnicas que necesitaban complementarse con otras para poder tener un conocimiento más detallado de los movimientos de las proteínas a escalas atómicas (Ode *et al.*, 2012).

La simulación computacional actualmente nos permite investigar la dinámica estructural de las proteínas en fracciones de tiempo de nanosegundos a microsegundos, y probablemente en un futuro se podrá trabajar en milisegundos (Henzler *et al.*, 2007). Esta técnica es ampliamente utilizada en el campo de la biología estructural. Entre las ventajas que presenta la dinámica molecular se encuentran que es útil para refinar las estructuras tridimensionales (3D) de las proteínas determinadas experimentalmente. Además, es beneficiosa para la construcción de estructuras 3D de proteínas previamente no descritas en combinación con técnicas de modelado de homología, cuando la estructura de un homólogo está disponible. (Dodson *et al.*, 2008). También, proporciona una herramienta única para ejecutar la dinámica estructural de las proteínas, es decir, permite conocer la evolución de las conformaciones en solución en escalas de tiempo de nanosegundos a microsegundos. Las

imágenes estructurales obtenidas durante la simulación son útiles para describir las características estructurales únicas de las proteínas (Karplus & McCammon, 2002).

Algunas proteínas no tienen una estructura 3D estable bajo condiciones fisiológicas y se las conocen como proteínas intrínsecamente desordenadas. Es posible que se sometan a una transición estructural de un estado desordenado a un estado ordenado al unirse a moléculas diana, como proteínas y material genético (Dunker *et al.*, 2002). Las proteínas virales o determinadas regiones de estas a menudo están intrínsecamente desordenadas, como es el caso de la proteína VPg ligada al genoma de los virus de plantas, Tat del VIH-1, proteínas Vif, y nucleoproteínas y fosfoproteínas de paramixovirus (Grzela *et al.*, 2008) (Shojania & D O'Neil, 2010) (H Reingewertz & Friedler, 2010) (Habchi & Longhi, 2012). Se ha propuesto que la estructura desordenada es beneficiosa para que los virus obtengan múltiples funciones en el ciclo de vida viral. El uso de dinámica molecular ha permitido entender mejor el plegamiento de las proteínas virales y comprender los principios estructurales mediante los cuales dichas proteínas ejecutan múltiples funciones en el ciclo de vida de los virus (Dunker *et al.*, 2002)

Dentro del campo de la virología, ha sido importante entender mejor la estructura de los virus para comprender los mecanismos de infección viral y de ensamblaje. La dinámica molecular se utiliza para abordar este problema. Freddolino *et al.* (2006) realizó una simulación de dinámica molecular de proteoma completo del virus satélite de mosaico del tabaco, reportada como una de las más grandes. Se simularon 1 millón de átomos y se evidenció los cambios ocurridos cada 50 ns. demostrando la estabilidad de todo el virión y del núcleo de ARN. Se analizaron las propiedades físicas de la partícula viral simulada, incluyendo el potencial electrostático, la distribución radial de los componentes virales y los patrones de movimiento correlacionados, así como las implicaciones para el ensamblaje y el mecanismo de infección del virus. De la misma manera, Larsson *et al.* (2012), usaron la dinámica molecular a gran escala para estudiar la cápside del virus de la necrosis del tabaco por satélite donde se analizaron los movimientos atómicos de 1,2 millones de partículas, y se vieron los cambios producidos por estas cada 1  $\mu$ s. Encontraron que la cápside se expandía en las simulaciones y que la permeabilidad del agua aumentaba 10 veces al eliminar los iones estructurales de calcio.

Gracias a estos estudios, actualmente se cuenta con una base estructural para identificar los reguladores clave del ensamblaje y las infecciones, y tener más certeza de cómo funcionan

mecánicamente. Aunque la dinámica molecular de viriones compuesta por un gran número de átomos es aún difícil de llevar a cabo en la mayoría de los casos, el progreso en el hardware y software para la simulación, junto con la acumulación de información biológica y fisicoquímica sobre viriones, ayudará a superar estas limitaciones para seguir avanzando en el campo de la virología (Arkhipov *et al.*, 2006).

## ÁREA DE ESTUDIO

Los virus han sido estudiados durante mucho tiempo, no solo por su patología y las enfermedades que provocan, sino también como sistemas modelo para procesos moleculares y como herramientas para identificar importantes proteínas y vías reguladoras celulares (Alfonso *et al.*, 2004). Los avances recientes en los métodos de espectrometría de masas, junto con el desarrollo de enfoques proteómicos, han facilitado enormemente la detección de componentes virales, las interacciones de proteínas en las células infectadas y los cambios inducidos por los virus en el proteoma celular, lo que ha permitido la comprensión más completa de la infección viral (Viswanathan & Früh, 2007). Además, se han incrementado los reportes de estructuras proteicas virales de alta calidad las cuales han proporcionado información valiosa sobre su mecanismo de acción, y han contribuido en el diseño y la comprensión de inhibidores específicos de virus. (Ince *et al.*, 2015).

A pesar de ello, el estudio del proteoma exige cada día una investigación más profunda y ya no solo se necesitan de procesos experimentales, sino actualmente se la complementa con la bioinformática. La dinámica molecular es un método computacional que permite conocer el movimiento de las proteínas a una escala atómica, mediante una simulación de condiciones biológicas reales. El virus de mosaico del babaco es un fitopatógeno que ha provocado importantes pérdidas agrícolas en nuestro país. Sin embargo, falta complementar la información que se tiene sobre el virus a nivel de secuencia con datos a nivel de estructura para lo cual en este proyecto de investigación se ha decidido caracterizar su proteoma mediante el uso de herramientas bioinformáticas (Guillén, 2017).

## JUSTIFICACIÓN

El Ecuador se caracteriza no solo por su diversidad biológica sino también por la variedad de microclimas los cuales permiten la producción de cultivos de importancia agrícola y económica. Por varios años se ha exportado varios productos de cultivos, principalmente banano, cacao y café. Sin embargo, no se ha podido expandir nuestro mercado a pesar de que nuestro país cuenta con cultivos no tradicionales o de especies “exóticas”, los cuales tienen apertura en los mercados extranjeros (Soria, 1999).

Uno de los cultivos no tradicionales con los que cuenta el Ecuador es el babaco. En los años 90, la producción de esta fruta cumplía con todos los requerimientos de calidad por lo que se lo exportaba a países como Estados Unidos, Canadá, México, Colombia, Holanda, Alemania, Italia, Reino Unido, Suecia y Suiza. Según datos reportados, desde el año 1996 hasta el 2002 se exportaba un promedio de 3270 toneladas anuales. Sin embargo, a inicios del año 2003, se dejó de comercializar al mercado internacional por incumplimiento de estándares fitosanitarios y falta de conocimiento por parte de los agricultores para manejar adecuadamente los cultivos de esta planta. (Tobar, 2008).

Los principales implicados en este problema son los virus, en especial los pertenecientes al género de los *Potexvirus*, pero aún no se tiene suficiente información sobre los mecanismos que usa para dañar a los cultivos y su forma de disseminación. En Pichincha se han reportado casos en los que las plantas de babaco han presentado síntomas tales como clorosis, ampollamiento y deformación de hojas; y mediante microscopía electrónica y técnicas serológicas se ha reportado que se trata de un virus con similitud con el virus de mosaico de la Papaya (Guillén, 2017) (Pisi & Vicchi, 1989).

Sin embargo, aún hay falta de conocimiento sobre los agentes fitopatógenos que afectan al babaco, en especial por parte de los agricultores quienes confunden las infecciones virales con daños producidos por otros factores como climáticos. Es por ello que se aspira que al solucionar este problema se logre implementar campañas de control del virus y de esta manera mejorar la productividad y la calidad de los cultivos. Una forma de conseguir esto es complementando estudios moleculares y técnicas de secuenciación con análisis bioinformáticos.

## **OBJETIVOS**

### **Objetivo general**

- Determinar los modelos estructurales tridimensionales y caracterizar las propiedades de las proteínas del virus de mosaico del babaco usando herramientas bioinformáticas.

### **Objetivos específicos**

- Predecir la función de las proteínas del virus de mosaico del babaco mediante similitud a nivel estructural con otras proteínas de otros virus.
- Establecer la estructura de cápside del virus de mosaico del babaco a partir de datos proteómicos.

## MATERIALES Y MÉTODOS

### Secuencia y estructura

Las secuencias fueron obtenidas por el Laboratorio de Biotecnología de la ESPE, quienes identificaron molecularmente el virus de mosaico del babaco en hojas de esta planta que presentaban la sintomatología y que fueron recolectadas en varios puntos de la provincia de Pichincha. (Secuencia del genoma y anotaciones completas están disponibles en NCBI GenBank: número de acceso MF978248)

A partir de estas secuencias, en la USFQ se procedió a analizar cada secuencia de aminoácidos con los servidores Phyre2 (Protein Homology Analogy Recognition Engine) y I-Tasser (Iterative Threading Assembly Refinement) los cuales permitieron predecir el modelo de la estructura tridimensional de las proteínas y para corroborar su calidad se usó como referencia los valores de fiabilidad que cada una de estas interfaces brinda. En el caso de Phyre2, se analizó los residuos de la secuencia y se obtuvo un porcentaje de confianza, el cual mientras más alto sea significaba que el modelo obtenido era más confiable y por ende tenía una mejor calidad. (Kelley *et al.*, 2015). Mientras que en I-Tasser se usó la métrica C-score el cual es un valor de confianza para estimar la calidad del modelo predicho; varía entre -5 y 2, pero al obtener un C-score  $> -1.5$  es indicativo de que el modelo tiene una topología global correcta. (Roy *et al.*, 2010).

### Análisis de Desorden

Se realizó un análisis de las regiones desordenadas de las proteínas, esto debido a que dichas regiones representan alrededor de una quinta parte de un proteoma típico, y que un tercio de las proteínas contienen una región desordenada contigua con una longitud superior a treinta aminoácidos, además que las proteínas del virus de mosaico del babaco se encuentran involucradas en procesos regulatorios importantes como la replicación (Bryson *et al.*, 2005).

Para este estudio se usó el método de predicción DISOPRED3 del grupo bioinformático Bloomsbury de la UCL. Las gráficas obtenidas permitieron analizar los niveles de confianza de desorden de los residuos involucrados en las interacciones proteína-proteínas (Buchan *et al.*, 2013).

### **Análisis de Coevolución**

Se procedió a realizar un análisis de la coevolución de todas las proteínas que conforman el virus de mosaico del babaco pues se consideró importante conocer sobre las interacciones existentes entre dos aminoácidos que son indispensables para la estructura y función de las proteínas, ya que pueden imponer restricciones evolutivas en los sets de mutaciones aceptables en los sitios de interacción. Para ello se empleó la plataforma Gremlin el cual es un método para aprender un modelo estadístico que permitió conocer patrones de conservación y coevolución de las proteínas. Se usaron las secuencias mejores alineadas en NCBI blast para cada una de las proteínas del virus objeto de estudio. (Kamisetty *et al.*, 2013).

### **Análisis de calidad de estructuras**

Fue necesario evaluar y analizar la calidad de los modelos obtenidos antes y después de iniciar los cálculos de dinámica molecular pues se deseó prevenir la presencia de errores en las estructuras o que en el caso de que fuesen inestables alteren el proceso. Para esto se utilizó la interfaz denominada “DOE Lab”, de la Universidad de California (UCLA), la cual cuenta con varias herramientas que permitieron validar y verificar los modelos proteína-proteína. Los programas utilizados fueron:

#### **ERRAT**

Permitió diferenciar entre las regiones de estructuras de proteínas determinadas correctamente o incorrectamente en función de la interacción atómica característica. Los errores en la construcción del modelo pueden a distribuciones más aleatorias de los diferentes tipos de átomos, que pueden distinguirse de las distribuciones correctas por métodos estadísticos. Los átomos se clasifican en una de tres categorías: carbono (C), nitrógeno (N) y oxígeno (O). De igual manera, utiliza una función de error cuadrático para caracterizar el conjunto de interacciones de los átomos pares no covalentes en una base de datos de 96 estructuras de proteínas confiables (Colovos & Yeates, 1993).

#### **Verify3D**

Determinó la compatibilidad de una estructura tridimensional (3D) conocida con su propia secuencia de aminoácidos (1D) asignando una clase estructural basada en su ubicación y entorno (alfa, beta, bucle, polar, no polar, etc.), y comparando los resultados con estructuras buenas. Los entornos de los residuos corresponden a tres parámetros: la estructura secundaria

local, el área del residuo que está abarcando en la proteína y la inaccesibilidad al solvente, y la fracción del área de la cadena lateral que está cubierta por los átomos polares (Bowie *et al.*, 1991).

### **Procheck**

Permitió comprobar la calidad estereoquímica de una estructura proteica determinada, evaluando cuando normal o inusual era la geometría de los residuos de dicha estructura, en comparación con los parámetros estereoquímicos derivados de modelos de alta resolución bien refinados. Esto se pudo hacer gracias a los distintos diagramas que se obtuvieron donde se hizo un análisis general y residuo por residuo (Laskowski *et al.*, 1993).

### **Dinámica molecular**

Los estudios que se han realizado hasta hora del virus de mosaico del babaco solo se han basado en análisis a nivel de secuencia, sin embargo, no se cuenta con una investigación sobre la similitud que pueden tener las diferentes proteínas a nivel estructural. Es por ello que se procedió a realizar una dinámica molecular con cada una de las proteínas que, si cumplieron con los parámetros de confianza de calidad antes mencionados, pues este es un paso preparativo importante para predecir la funcionalidad de las proteínas que conforman el virus de mosaico del babaco (Abraham *et al.*, 2015). El programa que se usó fue Gromacs 5.1.4 con el campo de fuerza AMBER-03.

Las condiciones a las cuales se llevó a cabo la dinámica molecular fueron a una fuerza máxima de convergencia del sistema de 10 kJ/mol/nm para la minimización de energía, a una temperatura máxima de 298K para la dinámica de acoplamiento de temperatura (NVT), pues es a esta temperatura a la cual la planta del babaco crece con normalidad; y a una presión de 1 bar para la dinámica de acoplamiento de presión (NPT) (Bhutani *et al.*, 2015).

Para el análisis de estructura se procedió a realizar el clusterizado pues a partir de esto se desea agrupar a las proteínas que tengan estados conformacionales moleculares similares en solución y de esta forma tener una idea más clara sobre las interacciones de la proteína. El proceso que se usó para esto fue un análisis de clúster hecho por Gromacs, se usó la función *gmx cluster* (Zhang *et al.*, 2005).

### **Análisis de estructura**

Se realizaron alineamientos de cada uno de los modelos proteicos estructurales tridimensionales a los cuales se les hizo dinámica molecular bajo condiciones fisiológicas normales con los modelos iniciales que no fueron simulados. Para esto se usó el programa TM-align, el cual mediante un puntaje (TM- score) permitió medir la similitud estructural de dos modelos de proteínas (Zhang & Skolnick, 2005).

### **Predicción de la función de las proteínas en base a su similitud estructural**

Para esto se empleó el servidor Dalí el cual permitió comparar estructuras de proteínas en 3D con un banco de datos con el fin de que se pudiera obtener similitudes biológicas de interés que no son detectables mediante la comparación de secuencias (Holm & Sander, 1995). En el caso de este estudio, se emplearon los archivos en formato PDB obtenidos antes y después de la dinámica molecular, se generó un documento de texto con las estructuras proteicas con las que fueron comparadas y se encontró cierta similitud a nivel estructural, y se brindó información de los siguientes parámetros: código PDB, cadena de la proteína, RMSD, Valor Z, largo del alineamiento (LAlí) y la descripción del PDB.

Posteriormente, se usó la base de datos UniProt la cual es un repositorio universal de datos de secuencias de proteínas e información funcional y brindó detalles sobre la función biológica de las proteínas derivadas de estudios de investigación (Apweiler *et al.*, 2004). Para este trabajo de investigación se procedió a emplear los hits que brindó DALI (código PDB), se los procesó dentro de esta base de datos y las proteínas que tenían similitud fueron exportadas a una hoja de cálculo. Las propiedades que permitió conocer este repositorio fueron: código PDB, código de entrada, nombre de la proteína con la que se comparó y el organismo al que corresponde, longitud, referencia cruzada, cadena a la que perteneció y función.

Finalmente, se usó la plataforma Rapid Miner la cual permite la preparación de datos, machine learning y el despliegue de modelos predictivos (Fayruzov *et al.*, 2010). En el caso de este estudio permitió filtrar los datos obtenidos en las dos bases de datos antes mencionadas y se unificó la información en una tabla. Los códigos PDB obtenidos de las diferentes proteínas permitieron aproximarnos a la función que podrían tener cada una de las proteínas del virus del mosaico del babaco.

Para corroborar esto, también se usó el servidor PDBeFold el cual permitió identificar las estructuras que fueron similares a las proteínas del virus de babaco en base a la identificación

de residuos que ocupan posiciones geométricas equivalentes (Velankar *et al.*, 2010). Para esta investigación se empleó el archivo PDB de cada una de las proteínas cuyas estructuras generamos y los resultados fueron analizados en base a las siguientes características: Alineamiento de calidad (Q), Pdb con el cual tiene similitud (Match), identidad de la secuencia (% SEQ), identidad de la estructura secundaria (%SSE).

## RESULTADOS

### Secuencia y estructura

Como se puede observar en la **Tabla No. 1**, los valores de confianza y calidad para cada uno de los modelos de las estructuras tridimensionales de las diferentes proteínas variaron. Respecto a CP, TGB1, helicasa y AlkB-domain, los valores arrojados por Phyre 2 y I-Tasser se encontraron dentro de los parámetros por lo que calificaron para proceder con la dinámica molecular.

En lo que corresponde a RDRP, su secuencia de aminoácidos abarcó 4619 residuos y excedió la cantidad de aminoácidos que Phyre 2 y I-Tasser podían analizar por lo que se tuvo que seleccionar su dominio usando la plataforma de análisis de secuencias de proteínas y clasificación “Interpro” del Instituto Bioinformático europeo (EMBL-EBI), y se obtuvo el porcentaje de confianza adecuado para llevar a cabo la simulación computacional.

También se obtuvo que TGB2, TGB3 y la metiltransferasa lograron porcentajes de confianza y C-score demasiado bajos por lo que se procedió a realizar la predicción de la estructura secundaria (**Figura No. 1 y 2**) y un análisis de desorden (**Figura No. 5 y 6**) con la plataforma del UCL (Bloomsbury Centre for Bioinformatics) con la finalidad de buscar regiones de las que fuese más factible generar una estructura por homología. Respecto a la predicción de la estructura secundaria, se obtuvo que tanto para TGB2 y TGB3, a lo largo de la secuencia de aminoácidos se obtuvo niveles bajo de confianza y se esperaba que estos sean altos. En los diagramas de desorden para ambas proteínas se pudo observar que hay regiones desordenadas con un alto nivel de confianza y que llamaron nuestra atención porque fue indicador de que carecen de manera total o parcial de una estructura tridimensional correctamente definida. En vista de esto, se procedió a tomar en cuenta las regiones de las hélices alfa de las estructuras secundarias predichas pues tuvieron un nivel alto de confianza y se esperó mejorar el modelo de ambas proteínas. Como se puede observar en las **Figuras No. 3, 4, 7 y 8**, estas secuencias al ser más estables debieron haber tenido una mejor calidad y haber mejorado el modelo. Sin embargo, al correrlas nuevamente en Phyre 2 y I-Tasser, como se puede observar en la **Tabla No. 1**, los valores de confianza no mejoraron. Respecto a la Metiltransferasa, como se puede observar en la **Tabla No. 2**, la mayor parte de la estructura es ordenada, y la predicción de estructura secundaria (**Figura No. 3**) tuvo un alto nivel de

confianza en toda la estructura lo que nos indicó que era estable, a pesar de ello se analizaron las regiones que comprendían a las hélices alfa y no se obtuvo un mejoramiento en la calidad del modelo (**Tabla No. 1**). Por dichas razones, para TGB2, TGB3 y la Metiltransferasa ya no fue posible continuar con un análisis a nivel de estructura sino es recomendable realizarlos solo a nivel de secuencia.

### **Análisis de Desorden**

Como se puede observar en la **Tabla No. 2**, después de realizar los análisis de desorden de cada una de las proteínas, TGB2 y TGB3 tuvieron más residuos que fueron desordenados con un nivel de confianza alto. Mientras que para CP, RdRp, TGB1, helicasa y metiltransferasa, se observó regiones intrínsecamente desordenadas con una fiabilidad superior al 0.5, solo para los primeros y últimos residuos. Respecto al dominio AlkB-B ningún aminoácido presentó un nivel de desorden intrínseco con un nivel de confianza que no sobrepasó 0.2.

### **Análisis de Coevolución**

Como se puede observar en la **Tabla No. 3**, no se logró realizar el análisis de coevolución para todas las proteínas pues según los reportes del programa, TGB1, TGB3, Helicasa, Metiltransferasa y RDRP no contaban con suficientes secuencias alineadas como para que el programa las procese. Sin embargo, para el resto de las proteínas si se logró identificar la covarianza existente entre los residuos. Se seleccionaron los tres pares de aminoácidos con una mayor probabilidad de que estén coevolucionando, dado el puntaje normalizado (scaled score) y la cantidad de secuencias por longitud. De igual manera, se los escogió porque el puntaje normalizado (scaled score) fue mayor que 1 para todos, lo cual fue indicativo de que son más covariables con respecto al promedio.

Gremlin también permitió analizar la conservación de la secuencia mediante el diagrama WebLogo3. Como se puede observar en la **Tabla No. 4**, las diferentes proteínas son bastantes conservadas pues las pilas totales en cada posición en la secuencia obtuvieron bits que se encontraban entre 2 y 4 bits, siendo TGB1 y AlkB-domain las más conservadas. Respecto a los colores de los nucleótidos fue posible determinar la hidrofobicidad, como se puede observar la mayor parte son verde lo cual es indicativo de neutra y negras lo cual significa que es hidrofóbico.

## Dinámica molecular

### RMSD

Dentro de la dinámica molecular, RMSD nos permitió conocer si las estructuras fueron estables o no en la escala de tiempo de las simulaciones. Como se observa en la **Figura No. 10**, el RMSD del frame inicial fue cero y conforme fue pasando el tiempo estos valores fueron incrementando hasta que llegaron a un punto, en el caso de este estudio los máximos, donde el valor de la desviación cuadrática media empezó a decrecer. Este incremento y disminución del RMSD nos sirvió para confirmar que la dinámica molecular llegó al equilibrio y se obtuvieron modelos estructurales con menos variación asegurándonos que las configuraciones obtenidas se acercaron a la realidad (Sargsyan *et al.*, 2017).

### RMSF

La RMSF dentro de la dinámica molecular nos brindó la medida del desplazamiento de los átomos en relación con nuestra estructura de referencia, promediada sobre el número de átomos. Como se puede observar en la **Tabla No. 5**, la fluctuación observada entre los residuos de cada una de las proteínas difirió unas con otras. CP y TGB1 tuvieron los mayores valores lo que indicó un nivel de flexibilidad y motilidad mayor de ambas durante la simulación (Sneha & Doss, 2016).

### Cálculo de Energía

Como se puede observar en la **Tabla No. 6**, las condiciones a las cuales se llevaron la dinámica molecular de cada una de las proteínas fueron las adecuadas. A pesar de que estos valores no nos dijeron nada sobre el nivel de confiabilidad de los modelos estructurales tridimensionales de las proteínas, nos permitieron comprobar que la simulación fue físicamente válida (Payne *et al.*, 1992). Respecto a la energía total, se esperaba y para todas las proteínas se obtuvieron valores negativos lo cual indicó un acoplamiento favorable de las configuraciones en términos energéticos. Mientras que para la temperatura se esperaba que se conserve alrededor a 300 K, pues es a esa temperatura en la que el babaco crece normalmente. En relación a la presión, se esperaba y se obtuvo que se encuentre cercana a 1 Bar lo cual nos indicó que se hizo correctamente la dinámica molecular de acoplamiento de presión. Lo que se refiere al volumen, todas se encontraban en los rangos permitidos. Finalmente, se esperaba que

la densidad se encuentre entre  $1000 \text{ kg/m}^3$ , pues es el valor de la densidad del agua, y cada una de las proteínas cumplieron con este parámetro.

### **Clusters**

El clusterizado para las diferentes proteínas se realizó para conocer la heterogeneidad de conformación en el conjunto de estructuras de proteínas generadas por la simulación. Para la proteína CP se obtuvieron 64 clusters, para RDRP 72, TGB1 30, AlkB-domain 9 y para la helicasa 17.

Como se puede observar en la **Tabla No. 11**, hubo diferencias en los modelos antes y después de llevar a cabo la dinámica molecular. En un inicio, las modelos estructurales de las proteínas se encontraban más amplias, mientras que en los diferentes clusters se las pudo ver más compactas, lo cual coincidió con los resultados esperados. Sin embargo, en RdRp ocurrió lo contrario y se evidenció que la proteína se encontraba más compacta antes que después de realizar la dinámica molecular.

### **Análisis de calidad**

Como se explicó previamente, los análisis de calidad se llevaron a cabo antes y después de la dinámica molecular para observar el efecto de esta última en las distintas proteínas, para ello se usaron los tres primeros clusters de cada una.

### **ERRAT**

Este indicador permitió conocer la calidad de los modelos mediante un porcentaje denominado Overall quality factor, el cual mientras mayor era significaba que tenía una mejor calidad. Como se puede observar en la **Figura No. 11** y en la **Tabla No. 7**, CP pasó de tener un valor de calidad de 76.44% a mayores de 95% para los tres clusters. Lo mismo sucedió con TGB1 y Alkb-domain, en donde hubo una mejoría significativa pasando de una calidad del 20% a 86%, y del 87% a 91%, respectivamente. Sin embargo, en el caso de RdRp ocurrió todo lo contrario, en lugar de mejorar empeoró la calidad, antes de la dinámica se reportó un porcentaje de 95.4%, mientras que después los clusters tuvieron un factor de 85%.

### **Verify3D**

Se esperaba que el porcentaje de residuos sea cercano a 100%, es decir que tengan un puntaje promedio de  $3D-1D \geq 2$ , y por ende iba a ser indicador de que el modelo estaba bien construido y que casi todos los residuos eran válidos en su conformación plegada. Como se

puede observar en la **Figura No. 12** y en la **Tabla No. 8**, para las proteínas CP, TGB1 y Alkb-domain hubo mejoría antes y después de haber llevado a cabo la dinámica, en especial con CP la cual pasó de 55.7% a 97%. Sin embargo, nuevamente para la proteína RdRp no se obtuvieron los resultados esperados y los clusters presentaron un porcentaje menor después de la simulación computacional.

### **Procheck**

Este indicador de calidad brindó una serie de diagramas, entre ellos Ramachandran, el cual permitió verificar la calidad estereoquímica residuo por residuo de una estructura proteica. Para ello, se analizó en base al porcentaje de residuos que se encuentren en regiones favorables. Para considerar un modelo de buena calidad el 90% de los residuos se deben encontrar en regiones favorables. Como se puede observar en las **Tabla No 9 y 10**, para todas las proteínas del BMV hay un incremento de residuos que estaban en regiones favorables al comparar las estructuras, antes y después de la dinámica molecular. Sin embargo, ninguno cumplió con el estándar de calidad.

### **Análisis de estructura**

Se realizó usando la plataforma de I-Tasser, TM-align y permitió conocer el alineamiento estructural entre el modelo de las proteínas del virus de mosaico del tabaco antes y después de realizar la dinámica molecular. Para esto, se analizó en base al puntaje, denominado TM-score, el cual varió entre 0 y 1, siendo 1 una coincidencia perfecta entre ambas estructuras, 0.5 significó que dichas estructuras comparten el mismo pliegue, y cuando fue menor a 0.2 fue indicativo de que no había similitud entre sí. Como se puede observar en la **Tabla No. 12**, ninguna de las proteínas con las cuales se trabajó presentaron un puntaje inferior a 0.5, por lo que todas tuvieron una alineación óptima. El dominio AlkB-B y la helicasa obtuvieron los valores más altos, 0.72 para ambas, lo cual fue indicativo de que estas proteínas tuvieron similitud a nivel estructural entre sus modelos previos y después de la dinámica. De igual forma, como se puede observar en la **Tabla No. 13**, ninguno de los modelos estructurales tridimensionales se superpuso perfectamente por lo que se descartó que sean iguales, sino solo similares entre sí. Esto concuerda con los valores del TM-score obtenidos donde son puntajes cercanos a uno.

## **Predicción de la función en base a la similitud estructural**

### **Proteína de Cobertura (CP)**

En base a los resultados arrojados por Dalí y UniProt, la proteína de cobertura del virus de mosaico del babaco tuvo similitud a nivel estructural con las proteínas de cobertura de los virus de mosaico de otros organismos como la papaya (4dox), pepino (5fn1), bambú (5 a2t) y sandía (5odv). En PDBeFold, se obtuvo que tuvo similitud a nivel estructural de igual manera con el CP del virus de mosaico del bambú (5 a2t), con una calidad de alineamiento de 0.43 (1 es el mejor puntaje), un 20% de identidad de secuencia de aminoácidos entre ambas proteínas, y 73% de identidad de la estructura secundaria.

### **Dominio RdRp**

En base a los resultados arrojados por Dalí y UniProt, la proteína RdRp del virus de mosaico del babaco tiene similitud con la polimerasa de los enterovirus (4ika) y la transcriptasa inversa del VIH-1 (1har). De igual manera, en PDBeFold, se obtuvo que RdRp del virus de mosaico del babaco tuvo similitud también con la RdRp del rinovirus humano (1xr5), la cual también está involucrada en la replicación del ARN. La calidad del alineamiento fue de 0.28, un 16% de identidad de secuencia de aminoácidos entre ambas proteínas, y 48% de identidad de la estructura secundaria.

### **TGB1**

En base a los resultados arrojados por Dalí y UniProt, esta proteína tiene similitud a nivel estructural con la Helicasa del virus de mosaico del tomate (3wry), así como con la helicasa del virus de la hepatitis C (1cu1). Respecto a PDBeFold, se obtuvo similitud con la helicasa del arterivirus (4n0o) que afecta a los equinos, con una calidad de alineamiento de 0.26 (siendo uno el mejor), un 17% de identidad de secuencia de aminoácidos entre ambas proteínas, y 80% de identidad de la estructura secundaria.

### **Helicasa**

En base a los resultados arrojados por Dalí y UniProt, esta proteína tuvo similitud a nivel estructural con la subunidad larga de la replicasa del virus de mosaico del tomate (3wry) y de igual forma con la replicasa (4n0o) del arterivirus que afecta a los equinos. Los resultados obtenidos en PDBeFold nos informaron sobre la similitud a nivel estructural con la helicasa del arterivirus equino la cual tuvo una calidad de alineamiento más alta, 0.27, con helicasa del

arterivirus antes mencionado (4n0o). La identidad de secuencia de aminoácidos entre ambas proteínas correspondió a 12% y la identidad de la estructura secundaria fue de 75%.

### **AlkB- domain**

En base a los resultados arrojados por Dalí y UniProt, esta proteína tuvo similitud con la legumina específica demetilasa del arroz (4igo), así como con en la legumina A del guisante *Pisum sativum* (3ksc). Respecto a PDBeFold, se obtuvo que AlkB-domain tiene una mejor calidad de alineamiento con el dominio cupin de *Bordetella pertussis* (2fqp). El valor de Q fue de 0.4, la identidad de secuencia de aminoácidos entre ambas proteínas correspondió a 16% y la identidad de la estructura secundaria fue de 60%.

## DISCUSIÓN

### Secuencia y estructura

Como se mencionó anteriormente, no todas las modelos de las estructuras tridimensionales obtenidos tuvieron una buena calidad como para iniciar con la dinámica molecular. CP, TGB1, el dominio RdRp, helicasa y Alkb-domain contaron con una confiabilidad excelente.

Sin embargo, la Metiltransferasa, TGB2 y TGB3 no cumplieron con los parámetros establecidos, a pesar de que se llevaron a cabo la predicción de la estructura secundaria y análisis de desorden. Se esperaba que las proteínas adopten conformaciones espaciales únicas en donde logren un estado nativo es decir un estado estable de energía libre mínima, y puedan cumplir su función y su estructura tridimensional este bien definida. (Mazo & Bedoya, 2011) (Cuevas & Covarrubias, 2011). Se decidió trabajar con las regiones de las hélices alfa de cada una de las estructuras secundarias obtenidas pues se supone que son las más estables por los niveles de confianza que obtuvieron, pero a pesar de ello tampoco se logró mejorar la calidad, esto se pudo deber a que el tamaño de la secuencia fue muy pequeño (Cuevas & Covarrubias, 2011). Otro factor que pudo haber incidido es que programas de modelación de proteínas tienen sus limitaciones, en el caso de Phyre2 si no puede detectar homologías entre las secuencias suministradas por los investigadores y una secuencia de estructura conocida, el modelado tiene una calidad muy baja, lo cual refleja un problema en el plegamiento de la proteína. La baja cantidad de homologías encontradas se debe principalmente a que el virus de mosaico del babaco no cuenta con una información completa de las proteínas. Además, aún no existen métodos confiables para predecir la estructura de una proteína únicamente a partir de la secuencia sola sin hacer referencia a estructuras conocidas (Kelley *et al.*, 2015).

En base a lo anterior, no fue posible realizar la dinámica molecular para TGB2, TGB3 y la metiltransferasa, pues la simulación busca caracterizar y predecir el comportamiento de sistemas reales y esto depende netamente de la calidad del modelo usado ya que de esta forma vamos a tener la información estructural que se desea. La dinámica molecular es un proceso complejo que, si bien permite estudiar procesos que son difíciles de llevar a cabo experimentalmente, exige una gran demanda de la precisión de la función de interacción atómica y el modelo que se usa (Daura *et al.*, 2003).

Una de las utilidades de la dinámica molecular de proteínas para comprender las funciones biológicas es el grado en el que las simulaciones muestrean adecuadamente el espacio conformacional de la proteína. Si una propiedad dada se muestre de manera deficiente en las simulaciones de dinámica molecular, los resultados obtenidos van a tener una utilidad limitada. Es por ello que no se pudo seguir usando TGB2, TGB3 y la metiltransferasa para pasos posteriores de la dinámica molecular por lo que se sugiere que se complemente con análisis a nivel de secuencia (Zhang *et al.*, 2003).

### **Análisis de Desorden**

El análisis de desorden nos informó que las proteínas tenían una estructura tridimensional bien definida y estable. Sin embargo, coincidió que, en casi todas, excepto en AlkB-domain, los primeros y últimos residuos si tenían un nivel de desorden alto, dentro de las proteínas es normal que los extremos se encuentren desordenados (Bryson *et al.*, 2005). Hace algunos años no se conocía sobre la importancia de estas regiones, pero ahora nos permite comprender que son fundamentales para que las distintas proteínas del virus de mosaico del tabaco cumplan con sus funciones. La presencia de estas regiones también nos permitió asumir que tienen aminoácidos cargados o polares, y que carecen de aminoácidos hidrofóbicos. Dentro de la dinámica molecular, estas regiones son las que fluctúan entre distintos estados conformacionales en rangos de tiempos muy pequeños. Por tanto, hay una concordancia entre las predicciones realizadas en DISOPRED3 y los resultados obtenidos con la dinámica molecular realizados en Gromacs 5.1.4. Estas fluctuaciones contribuyen a la flexibilidad de las proteínas permitiéndoles interactuar con una serie de moléculas diferentes con las que se unen con una afinidad mínima, pero conservando un alto nivel de especificidad. (Covarrubias & Cuevas, 2011).

### **Dinámica molecular**

#### **RMSD**

A pesar de su aparente simplicidad, el problema de cuantificar las diferencias entre dos estructuras de la misma proteína o complejo continúa evolucionando. (UIUC, 2018). En este estudio, se realizó un análisis de la desviación cuadrática media (RMSD) para obtener una medida de similitud ampliamente utilizada en el análisis de estructuras y dinámicas moleculares. Como se pudo observar en la **Figura No. 10**, en un inicio el RMSD se

encontraba en cero, esto se debió a que las posiciones de los átomos de las proteínas se encontraban fijadas durante ese tiempo en la simulación, para permitir a las moléculas de agua alrededor de la proteína que se ajusten a la superficie de la proteína. Después de eso, la proteína fue liberada, y el RMSD creció rápidamente hasta que llegó a un punto, en el caso de este estudio los puntos máximos, donde las proteínas ya se relajaron y cambió su estructura cristalina inicial a una estructura más estable. (UIUC, 2018).

### **RMSF**

Los RMSF se pueden calcular a partir de cualquier conjunto de conformaciones de proteínas, como las obtenidas a partir de simulaciones de dinámica molecular. Como se pudo observar en la **Tabla No. 5**, hay diferentes valores de RMSF para las proteínas del BMV lo cual quiere decir que sus átomos presentaron diferentes fluctuaciones sobre su posición promedio. En el caso de este estudio, se obtuvo que CP y TGB1 tienen un valor mayor de RMSF, esto fue indicativo de presencia de fluctuaciones y flexibilidad, pero para aseverar esto es necesario complementar con otras técnicas. Según la literatura, las diferentes formas de flexibilidad pueden ser identificadas en un modelo proteico dependiendo de la frecuencia de fluctuaciones y la proporción relativa de los modos de fluctuación contributiva. Es por ello que no es apropiado confiar en un procedimiento experimental único para evaluar la flexibilidad en una macromolécula, en especial cuando es compleja, porque una proteína puede ser rígida en una escala de tiempo microscópica y flexible en una escala de tiempo más larga. Dentro de la misma escala de tiempo, la flexibilidad y la rigidez de los homólogos de proteínas se pueden dividir de manera diferente sobre la matriz de proteínas y pueden involucrar diferentes modos de fluctuación. (Karshikoff *et al.*, 2015).

### **Clusters de las configuraciones obtenidos mediante la Dinámica Molecular**

Como se pudo observar en la **Tabla No. 11**, los modelos proteicos tuvieron diferencias antes y después de haber llevado a cabo la dinámica molecular, e incluso entre los mismos clusters. Esto se debió a la forma en la que cada una de las proteínas se plegaron (Daura *et al.*, 2009).

En un inicio, las cadenas tuvieron una estructura tridimensional abierta, es por ello que como se puede observar en la **Tabla No. 11** antes de la dinámica las estructuras fueron más sueltas, y las fuerzas efectivas promediadas en el tiempo entre los residuos jugaron un rol

importante pues permitieron el plegamiento de la cadena en una forma más compacta (Daura *et al.*, 2009) (Levitt & Warshel, 1975). Una vez que la cadena se tornó compacta tuvo menos libertad de movimiento, las fuerzas interatómicas específicas de corto alcance se volvieron importantes pues forman una conformación precisa siempre que la ganancia resultante en entalpía supere la pérdida de entropía. Dentro de este proceso, es muy importante los valores obtenidos de energía de la molécula pues permite conocer que la simulación se hizo a partir de una conformación correcta minimizando las zonas superpuestas sucesivas de algunos residuos a la vez. (Levitt, 2011).

El uso de un modelo simple basado en las fuerzas efectivas promediadas en el tiempo es bastante útil y aplicable, por ejemplo, en la formación de proteínas con estructura cuaternaria y complejos multienzimáticos, el ensamblaje del virus, entre otros. En cada nivel de complejidad, las fuerzas serían promediadas en el tiempo sobre aquellas subestructuras que son relativamente fijas o que parecen jugar un papel menos importante en el ensamblaje. (Levitt & Warshel, 1975).

### **Análisis de calidad de los modelos**

Actualmente dentro de las diferentes bases de datos de proteínas se han depositado miles de estructuras que han permitido mejorar los métodos cristalográficos y por ende han permitido el desarrollo y optimización de los cálculos para obtener dichas estructuras proteicas. Sin embargo, debido a la magnitud de dichos cálculos y la presencia de errores ha sido necesario incluir estimador de calidad. Como se mencionó anteriormente, los tres evaluadores seleccionados para este estudio fueron Errat, Verify 3D y Procheck, pues son los más usados para análisis comparativos entre modelos proteicos. (Govindaraj *et al.*, 2011).

#### **ERRAT**

Las diferentes estructuras fueron analizadas con este indicador el cual evalúa la distribución de los distintos tipos de átomos entre sí en un modelo proteico y permite verificar la calidad general. (Govindaraj *et al.*, 2011). De igual manera, ayuda a distinguir entre las regiones de las estructuras de las proteínas determinadas correcta e incorrectamente basadas en las interacciones atómicas características. Emplea un factor de calidad general el cual mientras más alto significa que el modelo tiene una mejor calidad. (Hooda *et al.*, 2012).

Como se pudo ver en la **Figura No. 11** y en la **Tabla No. 7**, las proteínas CP, TGB1 y Alkb-domain tuvieron un factor de calidad general diferente y más alto después de la dinámica molecular. Sin embargo, solo CP logró una calidad sobre el 95% y cumplió con el estándar determinado. Cuando un modelo proteico obtiene un factor de calidad general de 95% o mayor hay completa certeza de que se minimizaron los errores durante la construcción de dicho modelo y por ende se evitó distribuciones aleatorias de los diferentes tipos de átomos. (Colovos & Yeates, 1993).

En el caso de RdRp, el factor de calidad general no mejoró lo cual significó que los errores presentes en la estructura proteica no lograron corregirse durante la simulación y por ende la estructura proteica no pudo refinarse. De igual manera es indicador de que después de la dinámica molecular, las regiones incorrectas del modelo estructural tuvieron más distribuciones aleatorias (Colovos & Yeates, 1993).

### **Verify3D**

Este método de análisis de calidad permite conocer la compatibilidad de un modelo proteico con su secuencia de aminoácidos medido por un perfil tridimensional (Hou *et al.*, 2014). El perfil 3D de una estructura proteica es calculado a partir de las coordenadas atómicas de la estructura. Cuando las proteínas son buenos modelos, los perfiles tridimensionales coinciden con sus propias secuencias de aminoácidos y obtienen puntuaciones altas. La posición de cada residuo en el modelo 3D es caracterizada por su entorno y está representada por una fila de 20 números en el perfil. Estos números se conocen como puntuación 3D-1D y son las preferencias estadísticas de cada uno de los 20 aminoácidos para este entorno. Los entornos de los residuos se definen por tres parámetros: el área del residuo que está cubriendo, la fracción del área de la cadena lateral que está cubierta por átomos polares (oxígeno y nitrógeno) y la estructura secundaria local (Grotthuss *et al.*, 2003)

Como se puede observar en la **Figura No. 12**, CP, TGB1 y Alkb-domain presentaron un modelo proteico final mejor después de la dinámica molecular pues un mayor porcentaje de residuos coincidieron con sus perfiles tridimensionales y por ende obtuvieron un puntaje 3D-1D más alto. Esto igual es indicativo de que la conformación obtenida es más estable respecto a la inicial (Lakhlili *et al.*, 2015).

Respecto a la proteína RdRp, hubo un mayor porcentaje de residuos que tenían una puntuación 3D-1D promedio  $\geq 0.2$ , antes de la dinámica molecular que después. Esto se pudo haber debido a que el modelo proteico final no tuvo una buena compatibilidad de su perfil 3D con su secuencia de aminoácidos. Sin embargo, al evaluar la estructura inicial obtuvo que tan solo el 55% de los residuos tenían una puntuación 3D-1D promedio  $\geq 0.2$ , por lo que sería bueno analizar si existen problemas locales (Eisenberg & Bowie, 1997)

### **Procheck**

Proporciona una validación detallada de la estereoquímica de la estructura de una proteína (Laskowski *et al.*, 1993). Para ello emplea diagramas Ramachadran el cual es un gráfico de ángulos de torsión de los aminoácidos contenidos en una proteína (Menyhárd *et al.*, 2017). Los ángulos de torsión de cada residuo en un péptido definen la geometría de su unión a sus dos aminoácidos adyacentes colocando su enlace peptídico planar en relación con los dos enlaces peptídicos planares adyacentes, por lo que los ángulos torsionales determinan la conformación de los residuos y el péptido. Muchas de las combinaciones de ángulos, y por lo tanto las conformaciones de los aminoácidos, no son posibles debido al impedimento estérico (Ramachadran *et al.*, 1963).

Para que un modelo sea de buena calidad, como mínimo el 90% de los residuos se deben encontrar dentro de regiones favorables. Como se puede observar en la **Tabla No. 9 y 10**, ninguna de las proteínas logró cumplir con este parámetro de calidad, y como resultado la calidad estereoquímica no es tan buena. Esto nos permite asumir que hay la presencia de algún defecto en la distribución espacial y geometría de los átomos que conforman las proteínas lo cual puede influir en las propiedades físicas, químicas o la reactividad de las mismas (Zhou *et al.*, 2013). A pesar de ello, es importante resaltar que los valores obtenidos no son errores como tal, sino son características inusuales que pueden haber sido resultado de la unión de algún ligando al sitio activo de la proteína y pueden haber provocado distorsiones (Laskowski *et al.*, 1993).

### **Análisis de Coevolución**

Determinar qué residuos de una proteína contribuyen a una función específica es difícil. Los análisis de la coevolución de aminoácidos dentro de una familia de proteínas pueden servir como una guía útil al identificar residuos que están acoplados funcionalmente y también

para identificar cuáles residuos trabajan juntos para llevar a cabo una función. Para este estudio, las proteínas que se pudieron hacer un análisis de coevolución fueron CP, TGB2 y AlkB-domain, esto debido a que contaban con las secuencias alineadas suficientes como para que la plataforma las analice (Ashenberg & Laub, 2013). Como se pudo ver en la **Tabla No. 3**, los pares de residuos obtenidos para cada una de las proteínas fueron los más importantes y los que permitieron asumir que su presencia interviene en la selección evolutiva, es decir al haber una mutación en uno de estos puede provocar cambios en la funcionalidad de estas proteínas, como: promover plegamientos, interacciones proteína-proteína o contribuir a una actividad enzimática (Kamisetty *et al.*, 2013). (Wang *et al.*, 2017).

### **Predicción de la función de las proteínas en base a su similitud estructural**

Actualmente se tiene conocimiento sobre las funciones de las proteínas de diversas especies de virus que se encuentran dentro de los *Potexvirus*, sin embargo, para el virus de mosaico del babaco no se tenían datos aproximados. De acuerdo a UniProt, tan solo hay información oficial de las proteínas CP, TGB1, TGB2, TGB3 y RdRp; pero no hay de la Helicasa, metiltransferasa y AlkB. Mediante este estudio fue posible predecir una posible función de cada una de las proteínas que codifican para el virus de mosaico del babaco analizando su similitud a nivel estructural.

La proteína de cobertura o CP en general para los *Potexvirus* es la única proteína estructural del virus altamente conservada. En el caso de virus de mosaico del babaco conserva el motivo entre los aminoácidos en las posiciones 139-149 (Álvarez *et al.*, 2017). Se comprobó que tiene cierta similitud a nivel de secuencia con el virus de mosaico de la papaya, y a nivel estructural con el virus de mosaico del bambú, sandía y pepino, lo cual permitió inferir que la CP del BMV tiene la capacidad de auto-ensamblarse alrededor del ARN viral, y al comparar los modelos tridimensionales del CP PaMV y del BMV se puede asumir que lo hace en una vaina helicoidal flexible pero se debería confirmar esto con otras técnicas experimentales como con microscopía electrónica. (Yang *et al.*, 2012). De igual manera, los aminoácidos finales (dentro de los cuales podrían estar la arginina y lisina) en el extremo N-terminal podrían estar involucrados en la unión del ARN y por ende en la interacción entre la proteína y el ARN viral. (DiMaio *et al.*, 2015).

La RdRp es una proteína esencial que codifica para todos los virus de ARN. Dentro de los *Potexvirus* forma parte del primer cuadro abierto de lectura. En base a la similitud estructural encontrada, se confirma su papel en la replicación del ARN viral para generar nuevos templates de ARN viral y ARNm viral (Álvarez *et al.*, 2017). Se asume que este proceso se realiza directamente en el citoplasma de las células y que RdRp podría tener una replicación eficiente de su genoma debido a que reconoce el dominio D de la región UTR 3', el cual se ha reportado que posee hexanucleótidos potexvíricos altamente conservados, así como un pseudonudo, el cual también es conservado dentro de los *Potexvirus*, y podría comprender al menos 13 residuos de adenilato de la cola 3' poli (A) terminal (Tsai *et al.*, 1999). (Yeh *et al.*, 2004). De igual manera, se asume que la presencia de esta polimerasa permite que las actividades de unión y polimerización del ARN sean bastante cooperativas, lo cual sugiere que las interacciones polimerasa-polimerasa son importantes para llevar a cabo la función. (Hansen *et al.*, 1997). Además, se ha comprobado que las interacciones entre TGB1 y RdRp pueden variar la replicación del virus y la producción de síntomas en la planta infectada. (Lim *et al.*, 2010).

La proteína TGB1 forma parte del proceso del transporte del virus de célula a célula. En base a la similitud a nivel estructural que tuvo con las proteínas de otros virus se ha podido inferir que esta proteína contiene un dominio de helicasa mediante transferencia horizontal de genes. (Morozov & Solovyev, 2015). Este dominio contiene a su vez siete motivos típicos de una helicasa general, esto ha hecho que se considere a TGB1 como una proteína multifuncional con actividades importantes para el movimiento del virus a lo largo de las fibras de actina del retículo endoplasmático, induce el aumento del tamaño del poro del plasmodesmo, el motivo II es responsable de la unión de ATP y Mg<sup>2+</sup> a los sitios correspondientes en proteínas de unión a ATP y suprimen el silenciamiento de ARN viral (Deng *et al.*, 2013) (Verchot *et al.*, 2013). (Morozov & Solovyev, 2015).

La proteína helicasa, de la cual no había anotaciones computacionales previamente a este trabajo, se encuentra también en el primer cuadro abierto de lectura de los *Potexvirus*. De igual forma, al compararla en las distintas bases de datos y analizar su similitud a nivel estructural se pudo asumir que esta proteína motora interviene y asiste a RdRp en el proceso de replicación desenrollando la estructura de dsARN. Cabe recalcar que, aunque TGB1 tiene un motivo de helicasa, estas contienen una homología que no es superior al 20% a pesar de

que sus secuencias fueron similares en regiones que contenían motivos putativos I y II (Wang, 2012). (Morozov & Solovyev, 2015). Se han hecho varios estudios de fitomejoramiento en distintas plantas infectadas con *Potexvirus*, intentado insertar genes de resistencia en el genoma de las plantas que se unan a las proteínas de replicación y por lo tanto inhiban la formación de complejos de replicación, sin embargo, los virus al mutar a tasas elevadas superaron esta resistencia debido a sustituciones de aminoácidos en el dominio helicasa del primer cuadro abierto de lectura. (Ishibashi *et al.*, 2014).

Finalmente, el dominio de la proteína AlkB, del cual tampoco se tenía información computacional descrita anteriormente, forma parte de las proteínas de replicación del virus de mosaico del babaco. Al compararla estructuralmente se ha podido asumir que protegen al ARN contra el sistema de silenciamiento génico post-transcripcional que tienen las plantas como mecanismo de defensa. Este dominio también tiene motivos de sitios activos lo cual indica que tienen una alta tasa de mutación donde los dominios no esenciales se pierden rápidamente. Por lo tanto, la conservación de este dominio demuestra que son funcionales. (Bratlie & Drabløs, 2005). De igual manera, se ha comprobado que esta proteína tiene una similitud estructural con la familia de proteínas cupin perteneciendo ambas a la superfamilia de la oxigenasa dependiente de 2-oxoglutarato y Fe (II) y siendo proteínas oxidativas protegiendo al ARN del virus de compuestos metilantes. También se ha encontrado que estas dos familias de proteínas conservan en su estructura un double-strain-beta-torn (Uberto & Moomaw, 2013).

## CONCLUSIONES

La dinámica molecular permitió tener modelos estructurales tridimensionales de calidad y de esta forma fue posible conocer la similitud a nivel estructural que tiene con las proteínas de otros organismos. De igual forma, al analizar las similitudes existentes se pudo tener un conocimiento más específico sobre las propiedades que tiene cada una de las proteínas del virus de mosaico del babaco. También se concluye sobre la importancia de complementar los análisis de proteínas en base a similitud a nivel estructural y secuencial pues nos permiten obtener resultados más cercanos a la realidad, en el caso de este estudio nos informó que a pesar de ser proteínas virales pueden tener similitud a nivel de estructura incluso con proteínas de otros microorganismos, sin embargo, es necesario un muestreo más sistemático para tener datos más certeros.

## **RECOMENDACIONES**

A pesar de que la diferencia de los modelos estructurales tridimensionales para cada una de las proteínas no haya sido drástica, al ser este el primer estudio a nivel de similitud estructural que se hace con las proteínas del virus de mosaico del babaco es importante cualquier mejora que se tenga, así sea mínima, porque va a servir de referencia para otros estudios de virus que infectan al babaco. Al ser una de las vías de transmisión de este virus mediante el agua de riego, estamos seguros que hay más fitopatógenos virales que causan daños a los cultivos de esta planta, que no son caracterizados todavía y este estudio va a facilitar su caracterización.

## REFERENCIAS

- Alfonso, P., Rivera, J., Hernáez, B., Alonso, C., & Escribano, J. M. (2004). Identification of cellular proteins modified in response to African swine fever virus infection by proteomics. *Proteomics*, 4(7), 2037-2046.
- Alvarez-Quinto, R. A., Cornejo-Franco, J. F., & Quito-Avila, D. F. (2017). Characterization of a not so new potexvirus from babaco (*Vasconcellea x heilbornii*). *PloS one*, 12(12), e0189519.
- A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5: 725-738, 2010.
- Arkhipov, A., Freddolino, P. L., & Schulten, K. (2006). Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, 14(12), 1767-1777.
- Ashenberg, O., & Laub, M. T. (2013). Using analyses of amino acid coevolution to understand protein structure and function. In *Methods in enzymology* (Vol. 523, pp. 191-212). Academic Press.
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... & Martin, M. J. (2004). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 32(suppl\_1), D115-D119.
- Bhutani, I., Loharch, S., Gupta, P., Madathil, R., & Parkesh, R. (2015). Structure, dynamics, and interaction of Mycobacterium tuberculosis (Mtb) DprE1 and DprE2 examined by molecular modeling, simulation, and electrostatic studies. *PloS one*, 10(3), e0119771.
- Bol, J. F. (2008). Role of capsid proteins. In *Plant Virology Protocols* (pp. 21-31). Humana Press.
- Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016), 164-170.
- Bratlie, M. S., & Drabløs, F. (2005). Bioinformatic mapping of AlkB homology domains in viruses. *BMC genomics*, 6(1), 1.
- Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*. 41 (W1): W340-W348.
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, 2(9), 1511-1519.

- Cuevas-Velázquez, C. L., & Covarrubias-Robles, A. A. (2011). Las proteínas desordenadas y su función: una nueva forma de ver la estructura de las proteínas y la respuesta de las plantas al estrés. *TIP. Revista especializada en ciencias químico-biológicas*, 14(2), 97-105.
- Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F., & Mark, A. E. (1999). Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2), 236-240.
- Deng, Z., Lehmann, K. C., Li, X., Feng, C., Wang, G., Zhang, Q., ... & Wu, W. (2013). Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. *Nucleic acids research*, 42(5), 3464-3477.
- DiMaio, F., Chen, C. C., Yu, X., Frenz, B., Hsu, Y. H., Lin, N. S., & Egelman, E. H. (2015). The molecular basis for flexibility in the flexible filamentous plant viruses. *Nature structural & molecular biology*, 22(8), 642.
- Dodson, G. G., Lane, D. P., & Verma, C. S. (2008). Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO reports*, 9(2), 144-150.
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic acids research*, 32(suppl\_2), W665-W667.
- Dunker A. K., Brown C. J., Lawson J. D., Iakoucheva L. M., Obradovic Z. (2002). Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582
- Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in enzymology* (Vol. 277, pp. 396-404). Academic Press.
- Epa, V., Winkler, D., & Tran, L. (2012). Computational approaches. In *Adverse Effects of Engineered Nanomaterials*(pp. 85-96).
- Fayruzov, T., Dittmar, G., Spence, N., Teredesai, A., & De Cock, M. (2010). A RapidMiner framework for protein interaction extraction. In *1st RapidMiner Community Meeting and Conference (RCOMM 2010)* (pp. 58-63). Rapid-I.
- Fiser, A., & Do, R. K. G. (2000). Modeling of loops in protein structures. *Protein science*, 9(9), 1753-1773.
- Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A., & Schulten, K. (2006).

- Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3), 437-449.
- García, J. A., Glasa, M., Cambra, M., & Candresse, T. (2014). Plum pox virus and sharka: a model potyvirus and a major disease. *Molecular plant pathology*, 15(3), 226-241.
- Govindaraj, R. G., Manavalan, B., Basith, S., & Choi, S. (2011). Comparative analysis of species-specific ligand recognition in Toll-like receptor 8 signaling: a hypothesis. *PLoS one*, 6(9), e25118.
- Grotthuss, M. V., Pas, J., Wyrwicz, L., Ginalski, K., & Rychlewski, L. (2003). Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins: Structure, Function, and Bioinformatics*, 53(S6), 418-423.
- Grzela, R., Szolajska, E., Ebel, C., Madern, D., Favier, A., Wojtal, I., ... & Chroboczek, J. (2008). Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. *Journal of Biological Chemistry*, 283(1), 213-221.
- Guillén, Estefany, F. (2017). Detección e identificación de virus de los géneros Potexvirus, Tombusvirus y Tobamovirus en aguas de riego de cultivos de babaco (*Vasconcellea × heilbornii*. var. pentagona), en la provincia de Pichincha (Bachelor's thesis, Universidad de las Fuerzas Armada ESPE. Carrera de Ingeniería en Biotecnología.).
- Habchi, J., & Longhi, S. (2012). Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Molecular Biosystems*, 8(1), 69-81.
- Hansen, J. L., Long, A. M., & Schultz, S. C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, 5(8), 1109-1122.
- Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171), 913.
- Hooda, V., babu Gundala, P., & Chinthala, P. (2012). Sequence analysis and homology modeling of peroxidase from *Medicago sativa*. *Bioinformatics*, 8(20), 974.
- Hou, B., Liu, X., Zheng, F., Xu, X., & Zhang, Z. (2014). Molecular cloning, modeling and differential expression of a gene encoding a silent information regulator-like protein from *Sporothrix schenckii*. *International journal of molecular medicine*, 33(6), 1415-1422.

- Jankowsky, E., & Fairman, M. E. (2007). RNA helicases—one fold for many functions. *Current opinion in structural biology*, 17(3), 316-324.
- Ince, I. A., Boeren, S., van Oers, M. M., & Vlak, J. M. (2015). Temporal proteomic analysis and label-free quantification of viral proteins of an invertebrate iridovirus. *Journal of General Virology*, 96(1), 196-205.
- Ishibashi, K., Kezuka, Y., Kobayashi, C., Kato, M., Inoue, T., Nonaka, T., ... & Katoh, E. (2014). Structural basis for the recognition–evasion arms race between Tomato mosaic virus and the resistance gene Tm-1. *Proceedings of the National Academy of Sciences*, 111(33), E3486-E3495.
- Kamisetty, H., Ovchinnikov, S., & Baker, D. (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era. *Proceedings of the National Academy of Sciences*, 201314045.
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural and Molecular Biology*, 9(9), 646.
- Karshikoff, A., Nilsson, L., & Ladenstein, R. (2015). Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *The FEBS journal*, 282(20), 3899-3917.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), 845.
- Lakhlili, W., Chev e, G., Yasri, A., & Ibrahimi, A. (2015). Determination and validation of mTOR kinase-domain 3D structure by homology modeling. *OncoTargets and therapy*, 8, 1923.
- Larsson, D. S., Liljas, L., & van der Spoel, D. (2012). Virus capsid dissolution studied by microsecond molecular dynamics simulations. *PLoS computational biology*, 8(5), e1002502.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26(2), 283-291.
- Leiva Sandoval, A. M. *An alisis de diversidad de aislados latinoamericanos del virus del mosaico com un de la yuca (CsCMV; G enero Potexvirus) caracterizaci n biol gica, diagn stico molecular y obtenci n de genoma completo* (Doctoral dissertation, Universidad Nacional de Colombia-Sede Palmira).
- Levitt, M. (2001). The birth of computational structural biology. *Nature Structural &*

*Molecular Biology*, 8(5), 392.

- Levitt, M., & Warshel, A. (1975). Computer simulation of protein folding. *Nature*, 253(5494), 694.
- Lim, H. S., Vaira, A. M., Reinsel, M. D., Bae, H., Bailey, B. A., Domier, L. L., & Hammond, J. (2010). Pathogenicity of Alternanthera mosaic virus is affected by determinants in RNA-dependent RNA polymerase and by reduced efficacy of silencing suppression in a movement-competent TGB1. *Journal of general virology*, 91(1), 277-287.
- Loebenstein, G., & Lecoq, H. (2012). *Viruses and virus diseases of vegetables in the Mediterranean basin* (Vol. 84). Academic Press.
- Lough, T. J., Lee, R. H., Emerson, S. J., Forster, R. L., & Lucas, W. J. (2006). Functional analysis of the 5' untranslated region of potexvirus RNA reveals a role in viral replication and cell-to-cell movement. *Virology*, 351(2), 455-465.
- Mazo, C. X., & Bedoya, O. F. (2011). PESPAD: una nueva herramienta para la predicción de la estructura secundaria de la proteína basada en árboles de decisión. *Ingeniería y Competitividad*, 12(2), 9-22.
- Menyhárd, D. K., Hudáky, I., Jákli, I., Juhász, G., & Perczel, A. (2017). Predictable Conformational Diversity in Foldamers of Sugar Amino Acids. *Journal of chemical information and modeling*, 57(4), 757-768.
- Morozov, S. Y., & Solovyev, A. G. (2015). Phylogenetic relationship of some “accessory” helicases of plant positive-stranded RNA viruses: toward understanding the evolution of triple gene block. *Frontiers in microbiology*, 6, 508.
- Ode, H., Nakashima, M., Kitamura, S., Sugiura, W., & Sato, H. (2012). Molecular dynamics simulation in virus research. *Frontiers in microbiology*, 3, 258.
- Park, M. R., Jeong, R. D., & Kim, K. H. (2014). Understanding the intracellular trafficking and intercellular transport of potexviruses in their host plants. *Frontiers in plant science*, 5, 60.
- Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., & Joannopoulos, A. J. (1992). Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of modern physics*, 64(4), 1045.
- Pisi, A., & Vicchi, V. (1989). Virus-like particles found in babaco plants (*Carica pentagona* H.) in Italy. *Advances in Horticultural Science*, 3(2), 73-77. Recuperado el 15 de Enero del 2018 from <http://www.jstor.org/stable/42883001>
- Ramachandran GN, Ramakrishnan C, Sasisekharan V. (1963). Stereochemistry of polypeptide chain configurations. *J. mol. Biol*, 7, 95-99.

- H Reingewertz, T., E Shalev, D., & Friedler, A. (2010). Structural disorder in the HIV-1 Vif protein and interaction-dependent gain of structure. *Protein and peptide letters*, 17(8), 988-998.
- Shojania, S., & D O'Neil, J. (2010). Intrinsic disorder and function of the HIV-1 Tat protein. *Protein and peptide letters*, 17(8), 999-1011.
- Sneha, P., & Doss, C. G. P. (2016). Molecular dynamics: New frontier in personalized medicine. In *Advances in protein chemistry and structural biology* (Vol. 102, pp. 181-224). Academic Press.
- Song, H., Nakayama, E. E., Yokoyama, M., Sato, H., Levy, J. A., & Shioda, T. (2007). A single amino acid of the human immunodeficiency virus type 2 capsid affects its replication in the presence of cynomolgus monkey and human TRIM5 $\alpha$ s. *Journal of virology*, 81(13), 7280-7285.
- Soria, N., & Viteri, D. (1999). Guía para el cultivo de babaco en el Ecuador. Quito, Ecuador: Instituto Nacional Autónomo de Investigaciones Agropecuarias.
- Tarasova, E., & Nerukh, D. (2018). All-Atom Molecular Dynamics Simulations of Whole Viruses. *The journal of physical chemistry letters*.
- Tilsner, J., Linnik, O., Christensen, N. M., Bell, K., Roberts, I. M., Lacomme, C., & Oparka, K. J. (2009). Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *The Plant Journal*, 57(4), 758-770.
- Tobar Vaca, Martín Nicolás. Análisis de competitividad de los productores de babaco de San Pablo de Tenta (Saraguro, Loja) con enfoque de agro-cadena. Quito, 2008, 109 p. Tesis (Maestría en Estudios Latinoamericanos. Mención en Estudios Agrarios en América Latina). Universidad Andina Simón Bolívar, Sede Ecuador. Área de Estudios Sociales y Globales.
- Tsai, C. H., Cheng, C. P., Peng, C. W., Lin, B. Y., Lin, N. S., & Hsu, Y. H. (1999). Sufficient length of a poly (A) tail for the formation of a potential pseudoknot is required for efficient replication of bamboo mosaic potexvirus RNA. *Journal of virology*, 73(4), 2703-2709.
- Uberto, R., & Moomaw, E. W. (2013). Protein similarity networks reveal relationships among sequence, structure, and function within the cupin superfamily. *PLoS One*, 8(9), e74477.
- University of Illinois at Urbana-Champaign. (2018). VMD. Recuperado el 10 de Agosto del 2018 desde [www.ks.uiuc.edu](http://www.ks.uiuc.edu)

- Viswanathan, K., & Früh, K. (2007). Viral proteomics: global evaluation of viruses and their interaction with the host. *Expert review of proteomics*, 4(6), 815-829.
- Velankar, S., Alhroub, Y., Alili, A., Best, C., Boutselakis, H. C., Caboche, S., ... & Gore, S. P. (2010). PDBe: protein data bank in Europe. *Nucleic acids research*, 39(suppl\_1), D402-D410.
- Venkataraman, S., Prasad, B., & Selvarajan, R. (2018). RNA dependent RNA polymerases: insights from structure, function and evolution. *Viruses*, 10(2), 76.
- Verchot-Lubicz, J., Ye, C. M., & Bamunusinghe, D. (2007). Molecular biology of potexviruses: recent advances. *Journal of General Virology*, 88(6), 1643-1655.
- Verchot-Lubicz, J., Torrance, L., Solovyev, A. G., Morozov, S. Y., Jackson, A. O., & Gilmer, D. (2010). Varied movement strategies employed by triple gene block–encoding viruses. *Molecular plant-microbe interactions*, 23(10), 1231-1247.
- Voth, G. (2016). *Computational approaches for studying enzyme mechanism* (Vol. 577). Academic Press.
- Wang, L. Y., Lin, S. S., Hung, T. H., Li, T. K., Lin, N. C., & Shen, T. L. (2012). Multiple domains of the Tobacco mosaic virus p126 protein can independently suppress local and systemic RNA silencing. *Molecular plant-microbe interactions*, 25(5), 648-657.
- Wang, I. N., Yeh, W. B., & Lin, N. S. (2017). Phylogeography and Coevolution of Bamboo Mosaic Virus and Its Associated Satellite RNA. *Frontiers in microbiology*, 8, 886.
- Yang, S., Wang, T., Bohon, J., Gagné, M. È. L., Bolduc, M., Leclerc, D., & Li, H. (2012). Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus. *Journal of molecular biology*, 422(2), 263-273.
- Yeh, W. B., Hsu, Y. H., Chen, H. C., & Lin, N. S. (2004). A conserved secondary structure in the hypervariable region at the 5' end of Bamboo mosaic virus satellite RNA is functionally interchangeable. *Virology*, 330(1), 105-115.
- Zhang, Z., Shi, Y., & Liu, H. (2003). Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophysical journal*, 84(6), 3583-3593.
- Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research*, 33(7), 2302-2309.
- Zhou, A. Q., Caballero, D., O'Hern, C. S., & Regan, L. (2013). New insights into the interdependence between amino acid stereochemistry and protein structure. *Biophysical journal*, 105(10), 2403-2411

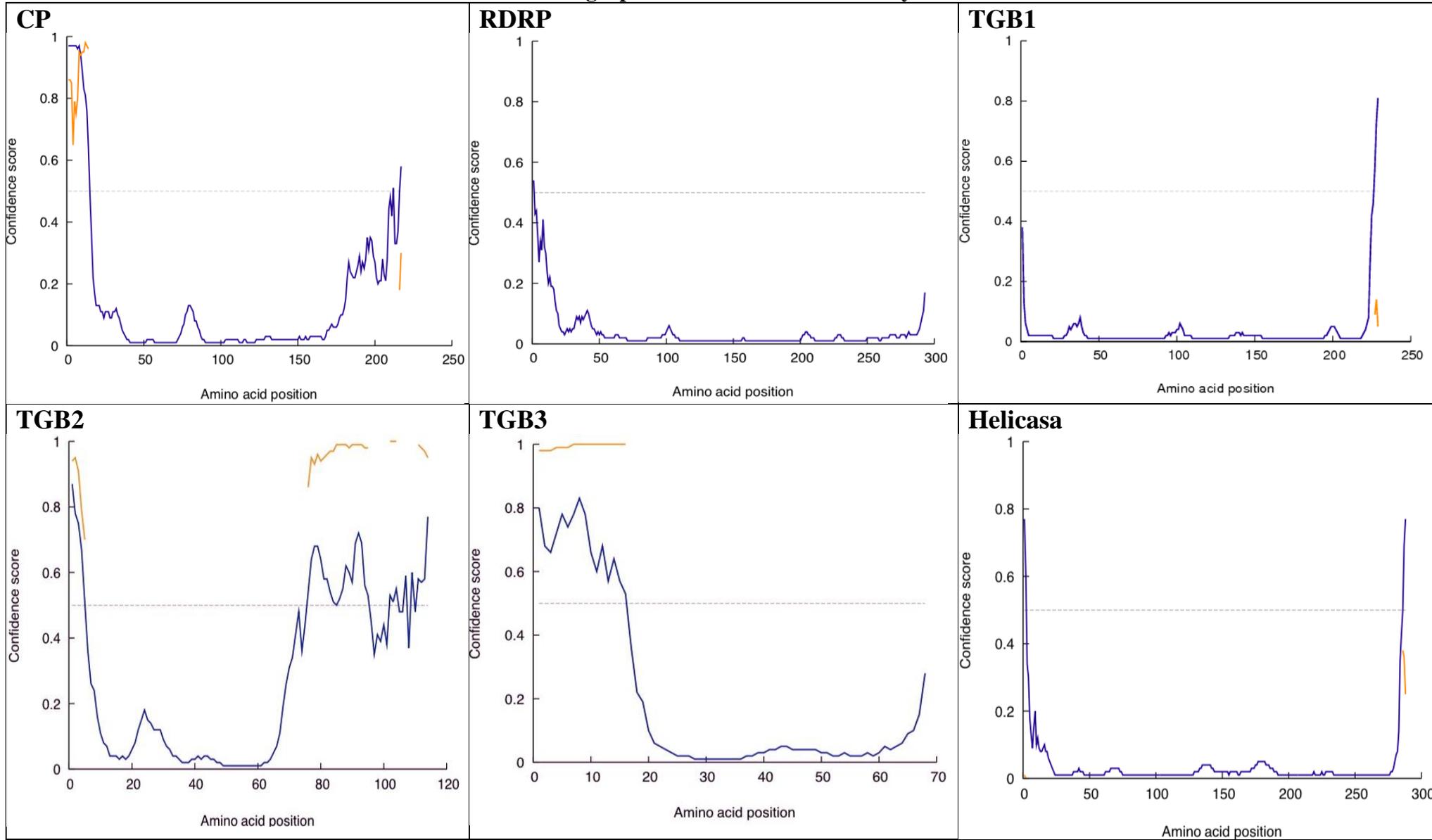
## Lista de Tablas

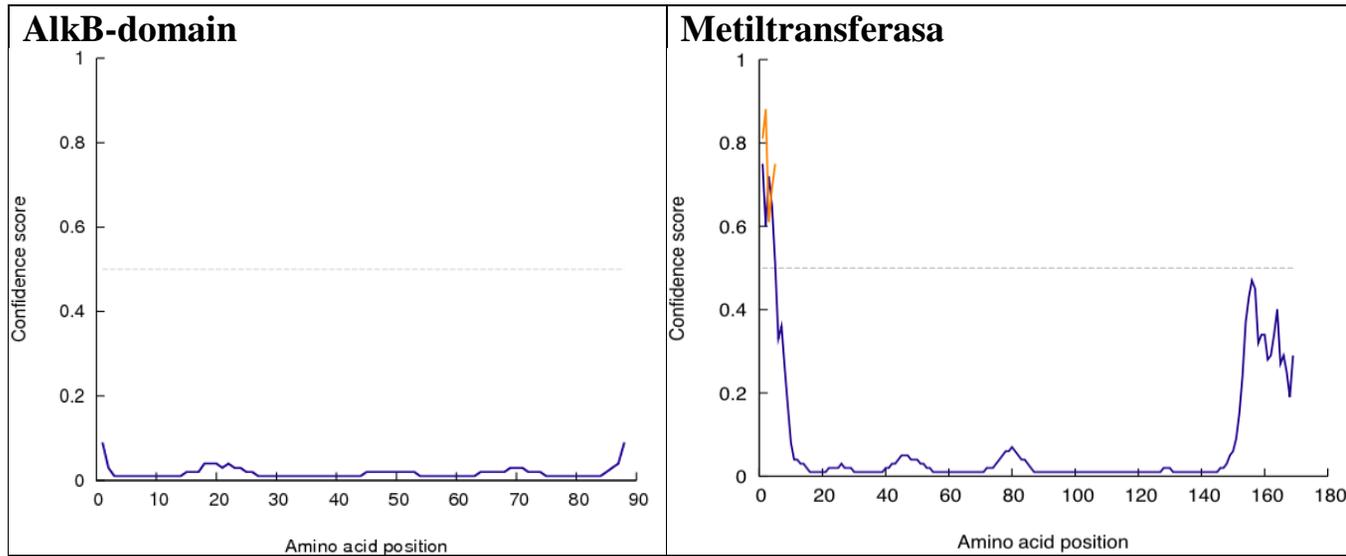
**Tabla No. 1 Valores de la fiabilidad y calidad de los modelos predichos de la estructura tridimensional de cada una de las proteínas del virus de Mosaico del Babaco**

Proteína	Porcentaje de Confianza	C-Score
CP	100%	1.52
TGB1	100%	1.1
ALKB-Domain	100%	0.98
RdRp	100%	-0.26
Helicasa	100%	-0.05
TGB2-helix	51.4%	-3.97
TGB2	27.7%	-3.45
Metiltransferasa-helix	22.4%	-3.36
Metiltransferasa	20.4%	-3.32
TGB3-helix	15.9%	-2.61
TGB3	11.8%	-2.4

Como se puede observar en la Tabla No. 1, no todas las proteínas cumplen con los parámetros de calidad esperados. Para que sean considerados como buenos modelos deben tener un porcentaje de confianza del 100% y un C- score mayor a -1.5. En el caso de este estudio, solo: CP, TGB1, AlkB-domain, RdRp y la Helicasa, cumplieron estos parámetros. TGB2, TGB3 y la metiltransferasa no lograron los puntajes mínimos. Después de realizar la predicción de estructura secundaria y análisis de desorden de TGB2, TGB3 y la metiltransferasa, se puede ver que TGB2-helix, TGB3-helix y Metiltransferasa-helix tampoco obtuvieron valores de calidad buenos.

**Tabla No. 2 Análisis de Desorden de las proteínas del Virus de Mosaico del Babaco empleando la herramienta DISOPRED3 del grupo bioinformático Bloomsbury de la UCL**





Se consideran a las proteínas desordenadas cuando la línea azul esta sobre la línea discontinua gris y por ende su puntaje de confianza fue mayor a 0.5. La línea naranja muestra la confianza de las predicciones de residuos de unión a proteínas desordenadas. Como se muestra en la tabla, TGB2 y TGB3 son las proteínas más desordenadas.

**Tabla No. 3 Análisis de coevolución de las proteínas del virus del babaco**

<b>Proteína</b>	<b>Residuo i</b>	<b>Residuo j</b>	<b>Scaled Score</b>	<b>Probabilidad</b>
<i>CP</i>	93_A	117_C	1.848	0.37
	148_C	220_T	1.640	0.27
	347_A	457_A	1.614	0.26
<i>TGB2</i>	32_C	46_A	1.962	0.73
	34_A	123_A	1.862	0.67
	79_T	144_A	1.729	0.59
<i>AlkB-domain</i>	426_A	504_T	1.622	0.20
	1_C	30_G	1.588	0.19
	196_C	322_C	1.570	0.19

Como se puede observar en la tabla, para cada una de las proteínas se escogieron los tres residuos más representativos y con un nivel de probabilidad alto. La puntuación normalizada (scaled score) debe ser mayor a 1 para decir que un par de residuos en particular son altamente covariables en comparación con el promedio. El valor de la probabilidad corresponde a la posibilidad de que un par de residuos estén en contacto, dado el valor del puntaje normalizado y la cantidad de secuencias por longitud





**Tabla No. 5 RMSF de las diferentes proteínas del virus de mosaico del babaco**

Proteína	RMSF mínima (Å)	RMSF máxima (Å)	RMSF promedio (Å)
<i>CP</i>	0.634	6.328	2.316
<i>RDRP</i>	0.114	1.195	0.327
<i>TGB1</i>	0.505	5.536	2.079
<i>AlkB-domain</i>	0.721	3.692	1.777

Como se puede observar en la tabla, la fluctuación cuadrática media existente entre los residuos presentes en cada una de las proteínas tuvieron un mayor valor en CP y TGB1 es decir tuvieron una mayor flexibilidad de la proteína durante la simulación.

**Tabla No. 6 Cálculo de energía para cada una de las proteínas del virus de mosaico del babaco durante la dinámica molecular**

	CP	RDRP	TGB1	Helicasa	AlkB-domain
<i>Energía Total (kJ/mol)</i>	-942723	-483462	-588833	-4828343	-155232
<i>Temperatura (K)</i>	310.002	296.988	296.991	296.984	297.001
<i>Presión (bar)</i>	0.196994	1.08616	0.714905	0.502795	0.340572
<i>Volumen (nm<sup>3</sup>)</i>	891.836	450.375	541.898	450.678	142.262
<i>Densidad (kg/m<sup>3</sup>)</i>	991.838	1026.06	1012.45	1020.52	1026.54

Como se puede observar en la tabla, cada una de las condiciones a las que se llevó a cabo la dinámica molecular para cada una de las proteínas se encuentra dentro de los parámetros permitidos.

**Tabla No. 7 Análisis de Calidad Errat para las proteínas del BMV**

Proteína	Overall Quality Factor antes de la dinámica molecular	Overall Quality Factor después de la dinámica molecular		
		Cluster 1	Cluster 2	Cluster 3
<i>CP</i>	76.442	95.939	97.98	97.423
<i>RdRp</i>	95.406	85.075	86.397	91.488
<i>TGB1</i>	20	77.251	86.111	86.854
<i>AlkB-domain</i>	87.500	77.94	91.176	73.529
<i>Helicasa</i>	37.652	92.03	91.23	92.94

Como se puede observar en la tabla, el factor de calidad general varía para cada una de las proteínas antes y después de haber realiza la dinámica molecular

**Tabla No. 8 Análisis de calidad de las distintas proteínas de BMV a partir del indicador Verify 3D**

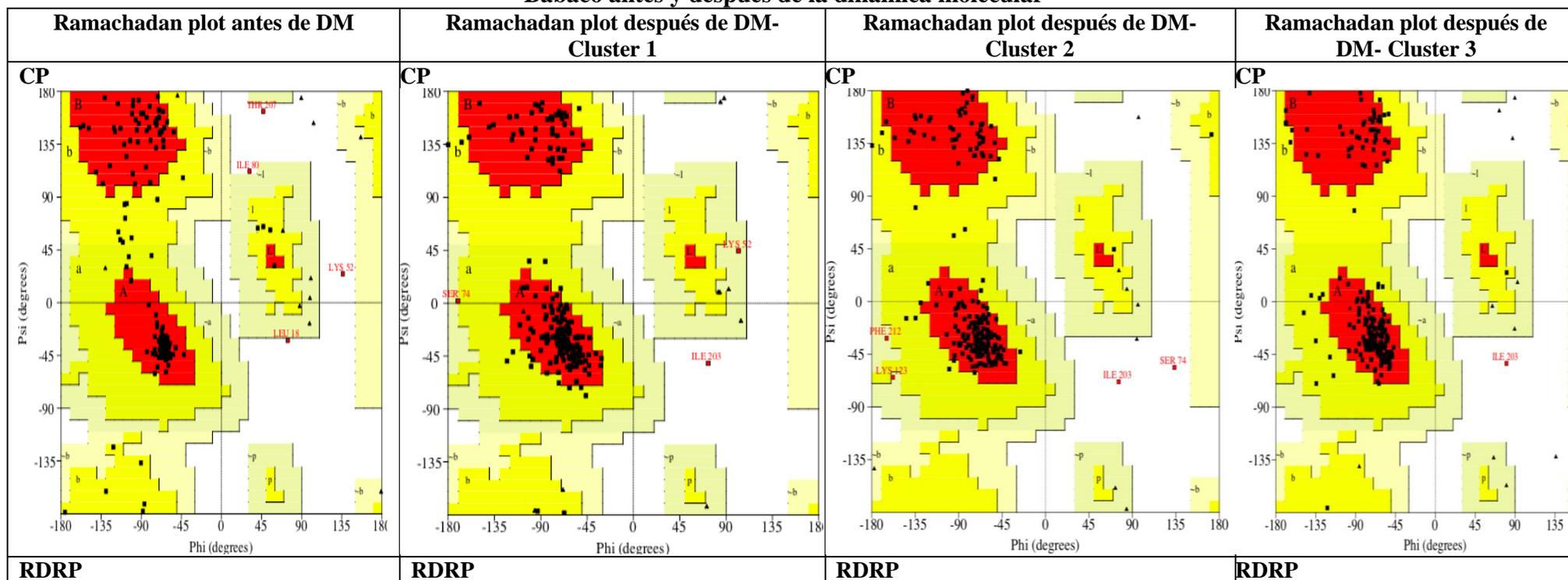
Proteína	3D-1D score $\geq 0.2$ , antes de la dinámica molecular (%)	3D-1D score $\geq 0.2$ , después de la dinámica molecular		
		Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)
<i>CP</i>	55.7	95.939	97.98	97.423
<i>RDRP</i>	55.96	37.54	44.37	45.39
<i>TGB1</i>	57.8	70.31	77.33	68.56
<i>AlkB-domain</i>	64.77	76.14	72.73	81.82
<i>Helicasa</i>	37.652	86.72	85.94	81.23

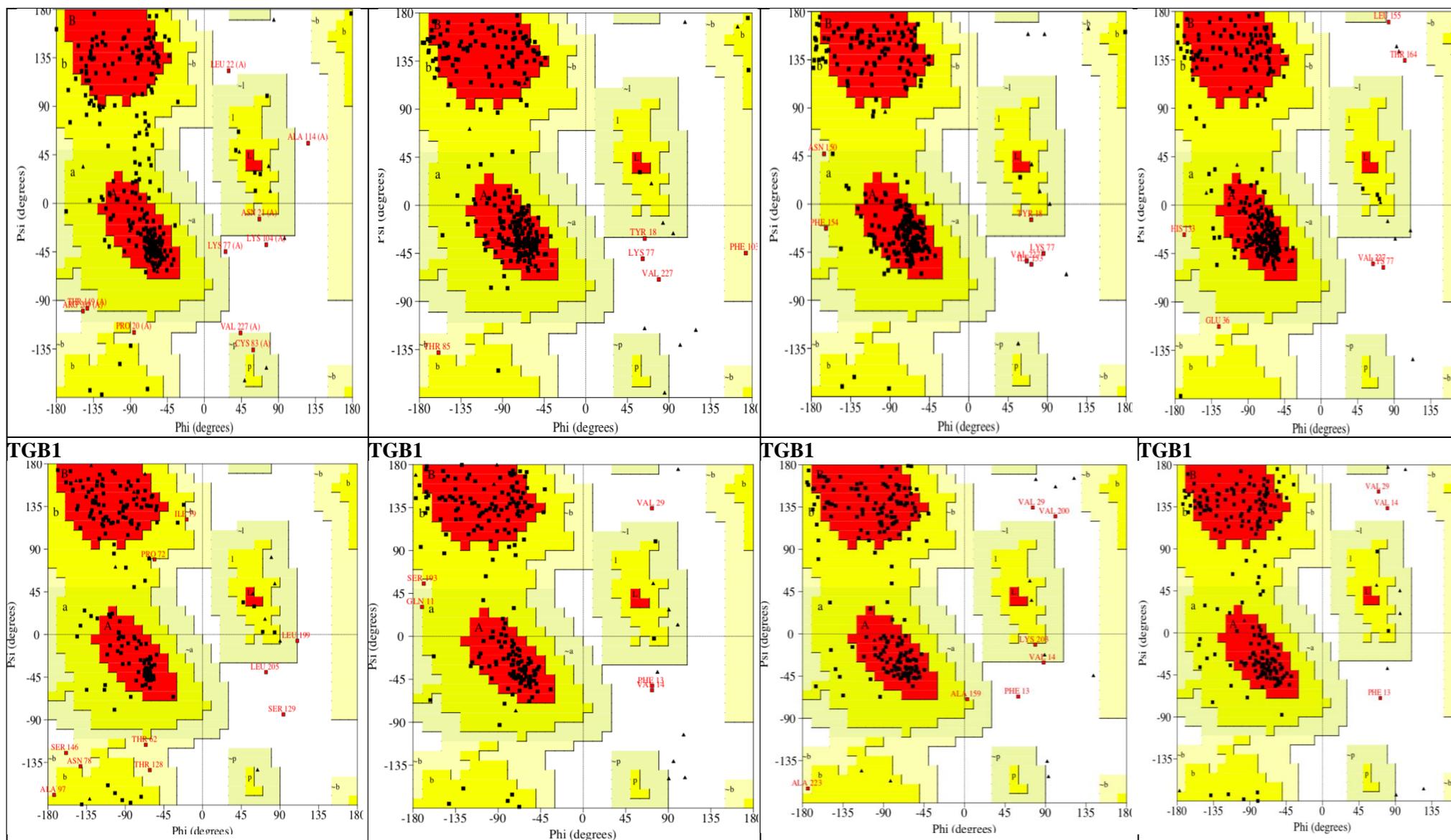
Como se puede observar en la tabla, para este evaluador hay una evidente mejora antes y después de haber llevado a cabo la dinámica molecular.

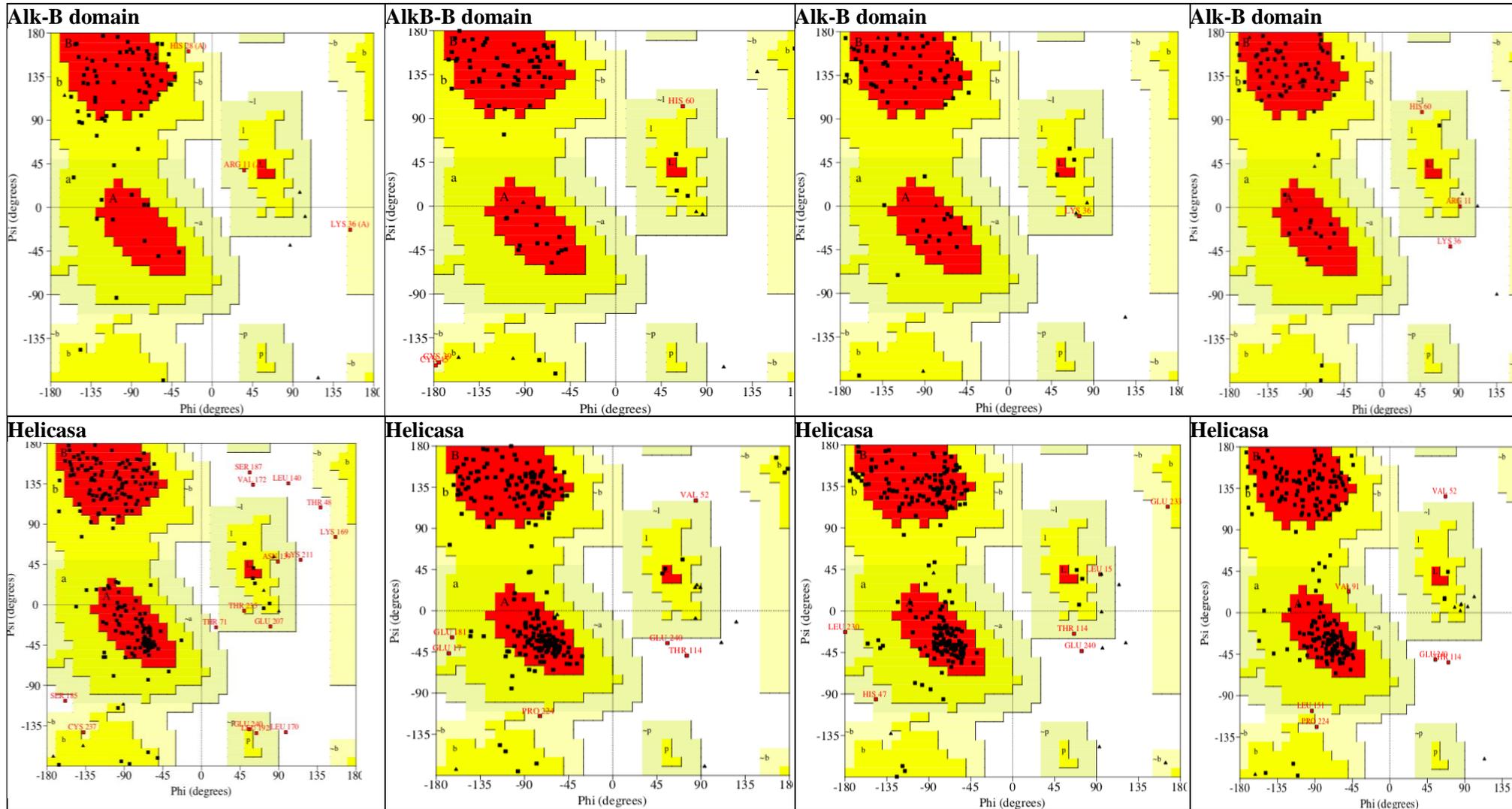
**Tabla No. 9 Análisis de calidad de las distintas proteínas de BMV a partir del indicador Procheck**

Proteína	Valores del diagrama Ramachadan antes de la dinámica				Valores del diagrama Ramachadan después de la dinámica- Cluster 1				Valores del diagrama Ramachadan después de la dinámica- Cluster 2				Valores del diagrama Ramachadan después de la dinámica- Cluster 3			
	Core	Allow	Gener	Disall	Core	Allow	Gener	Disall	Core	Allow	Gener	Disall	Core	Allow	Gener	Disall
CP	86.1%	11.8%	0%	2.1%	88.8%	9.1%	1.1%	1.1%	88.2%	11.2%	0%	0.5%	88.8%	9.1%	1.1%	1.1%
RdRp	73.2%	23.4%	1.9%	1.5%	85.8%	12.3%	0.8%	1.1%	82.8%	14.9%	1.1%	1.1%	84.7%	12.6%	1.1%	1.5%
TGB1	79.9%	15.2%	3.3%	1.6%	81.6%	15.8%	1%	1.5%	82.7%	13.8%	1.5%	2%	79.1%	19.4%	0%	1.5%
AlkB- Domain	66.2%	29.7%	4.1%	0%	85.1%	10.8%	4.1%	0%	79.7%	18.9%	1.4%	0%	81.1%	14.9%	2.7%	1.4%
Helicasa	80.5%	12.8%	3.1%	3.5%	83.6%	14.2%	0.9%	1.3%	82.7%	14.6%	2.2%	0.4%	84.5%	13.3%	0.9%	1.3%

**Tabla No 10. Análisis de calidad Procheck mediante diagramas Ramachadan de las proteínas del Virus en Mosaico del Babaco antes y después de la dinámica molecular**





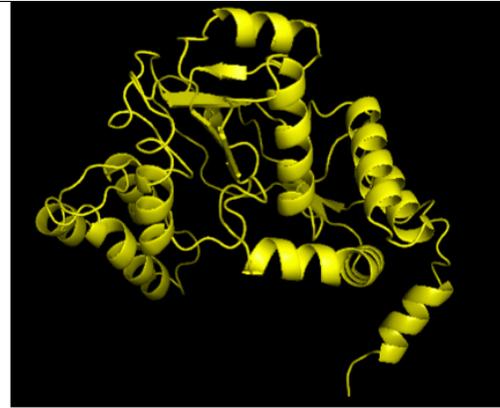


Como se puede observar en la Tabla No. 10, para cada una de las proteínas coincide que la mayor parte se encuentra en regiones favorables y permitidas, y un porcentaje menor en las que no lo son. En los diagramas se puede evidenciar que las zonas favorables son rojas, las amarillas las permitidas, y las beige y blancas las no permitidas.

Para la Tabla No. 9, las regiones denominadas *Core* son favorables, *Allow* permitidas, y *Gener* y *Disall* no permitidas.

**Tabla No. 11 Visualización de los modelos de cada una de las proteínas del BMV antes y después de la dinámica molecular**

Proteína	Modelos proteicos
<p><i>CP</i></p>	<div data-bbox="1026 363 1629 740" data-label="Chemical-Block"> </div> <p data-bbox="743 764 1877 800">a) Proteína CP del virus de mosaico del babaco antes de la dinámica molecular</p> <div data-bbox="569 816 1005 1292" data-label="Chemical-Block"> </div> <p data-bbox="579 1308 995 1425">b) Cluster 1 de la Proteína CP del VMB después de la dinámica molecular</p> <div data-bbox="1039 816 1503 1292" data-label="Chemical-Block"> </div> <p data-bbox="1050 1308 1486 1425">c) Cluster 2 de la Proteína CP del VMB después de la dinámica molecular</p> <div data-bbox="1537 816 1980 1292" data-label="Chemical-Block"> </div> <p data-bbox="1547 1308 1984 1425">d) Cluster 3 de la Proteína CP del VMB después de la dinámica molecular</p>

*RdRp*

a) Proteína RdRp del virus de mosaico del babaco antes de la dinámica molecular



b) Cluster 1 de la Proteína RdRp del VMB después de la dinámica molecular



c) Cluster 2 de la Proteína RdRp del VMB después de la dinámica molecular



d) Cluster 3 de la Proteína RdRp del VMB después de la dinámica molecular

*TGB1*

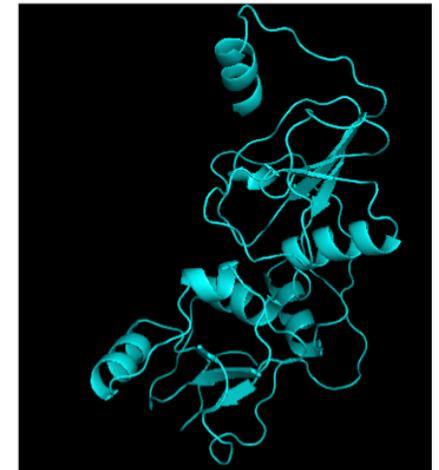
a) Proteína TGB1 del virus de mosaico del babaco antes de la dinámica molecular



b) Cluster 1 de la Proteína TGB1 del VMB después de la dinámica molecular



c) Cluster 2 de la Proteína TGB1 del VMB después de la dinámica molecular



d) Cluster 3 de la Proteína TGB1 del VMB después de la dinámica molecular

*AlkB-domain*



a) Proteína Helicasa del BMV antes de la dinámica molecular



b) Cluster 1 de la Proteína helicasa del VMB después de la dinámica molecular



c) Cluster 2 de la Proteína helicasa del VMB después de la dinámica molecular



d) Cluster 3 de la Proteína helicasa del VMB después de la dinámica molecular

*Helicasa*

a) Proteína Helicasa del BMV antes de la dinámica molecular



b) Cluster 1 de la Proteína helicasa del VMB después de la dinámica molecular



c) Cluster 2 de la Proteína helicasa del VMB después de la dinámica molecular



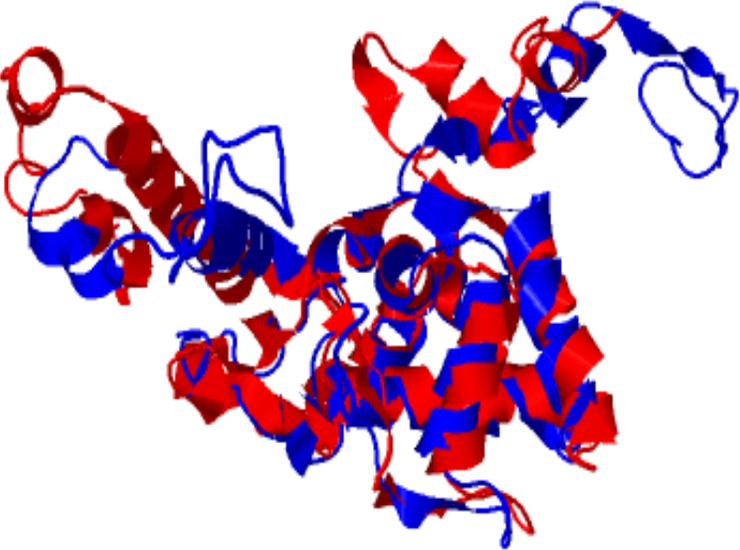
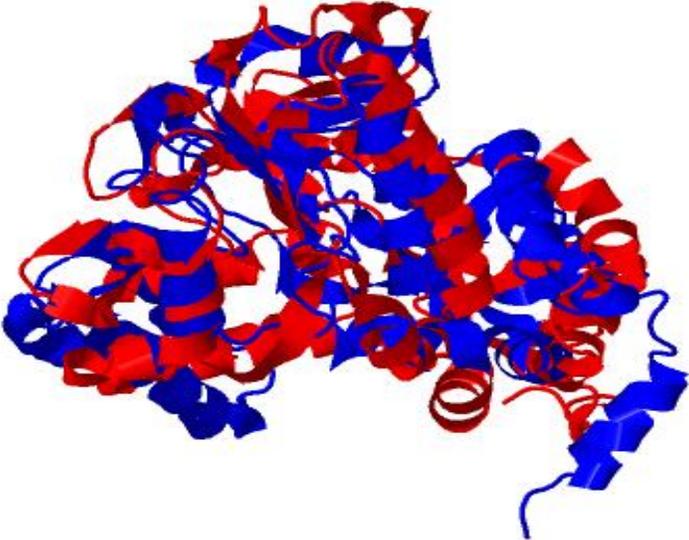
d) Cluster 3 de la Proteína helicasa del VMB después de la dinámica molecular

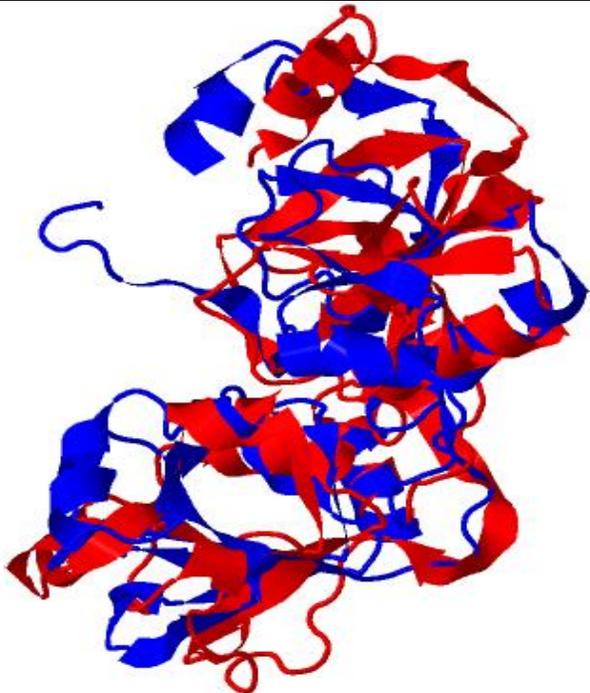
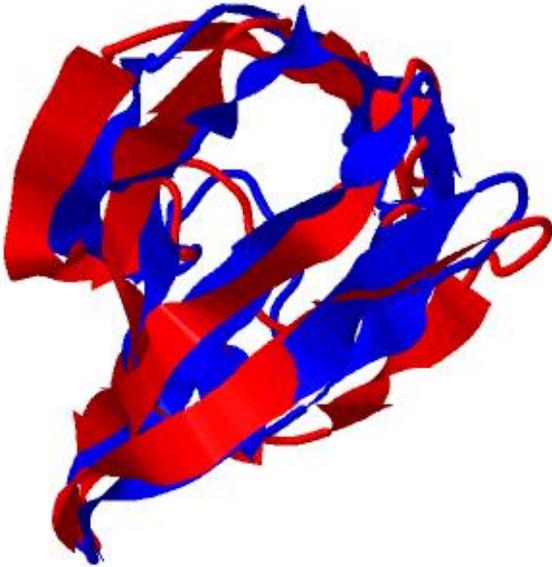
Como se puede observar, cada una de los clusters de las proteínas poseen una estructura más compacta después de la dinámica molecular, a excepción de la proteína RdRp en la cual sucede lo contrario y su modelo es más amplio después de la simulación computacional.

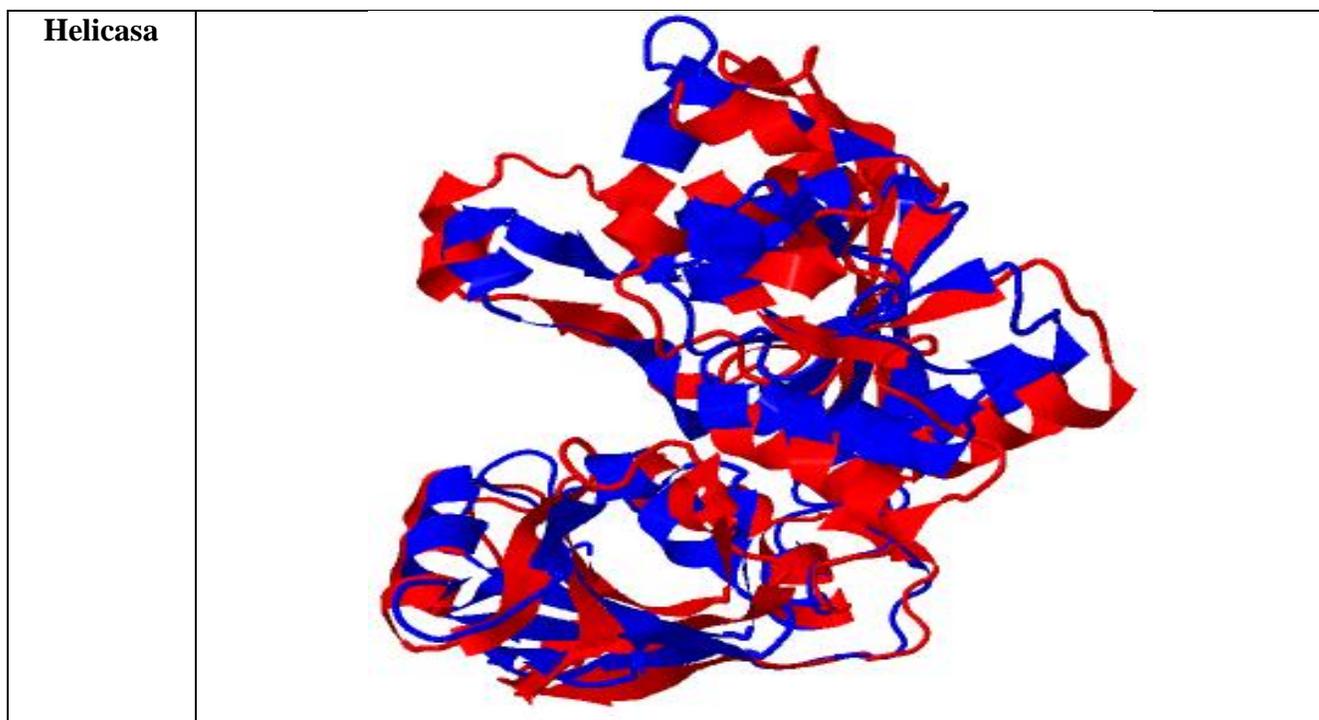
**Tabla No. 12 Alineamiento entre las proteínas del virus de mosaico del babaco antes y después de realizar la dinámica molecular empleando la plataforma TM-align**

Proteína	Longitud alineada (aa)	TM- score
<i>CP</i>	176	0.67830
<i>RdRp</i>	256	0.56268
<i>TGB1</i>	185	0.58019
<i>AlkB- domain</i>	87	0.72811
<i>Helicasa</i>	249	0.72804

**Tabla No. 13 Superposición de las proteínas del virus de babaco del masaco antes y después de realizar la dinámica molecular**

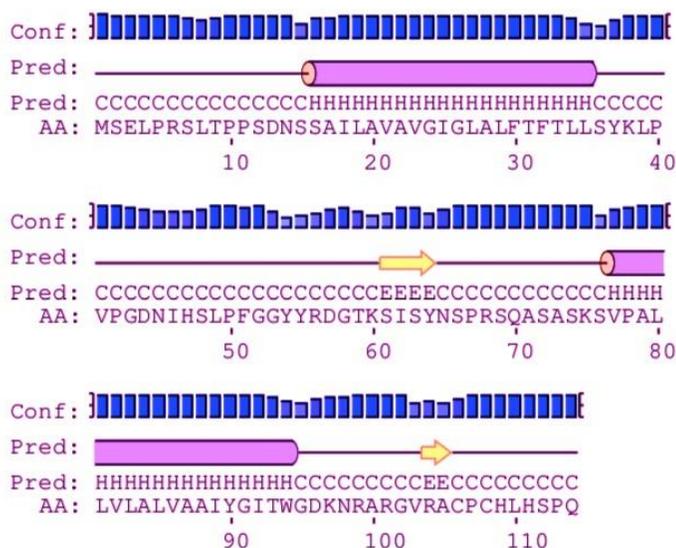
Proteína	Superposición de las proteínas
CP	
RDRP	

<b>TGB1</b>	
<b>AlkB-B domain</b>	



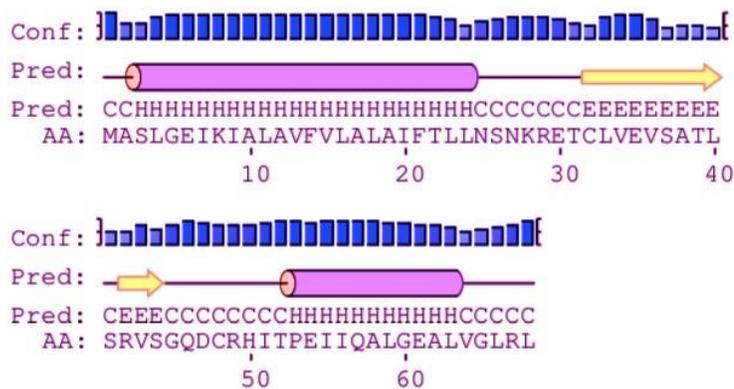
Como se puede observar en la Tabla No 13, cada una de las superposiciones se hace con las proteínas antes (azul) y después (rojo) de haber llevado a cabo la dinámica molecular. Si no se evidenciaran dos colores diferentes para cada modelo estructural de cada proteína se diría que son completamente idénticas, pero como si se observan diferencias a lo largo de las cadenas se dice que tienen solo cierto nivel de similitud.

## Lista de Figuras



**Figura No. 1 Diagrama PSIPRED para la proteína TGB2 donde muestra la predicción de la estructura secundaria.**

Conf: confianza de la predicción; Pred: Estructura secundaria predicha; AA: Secuencia target; Barras fucsias: hélices alfa; Flechas amarillas: laminas beta; Línea negra: coil.

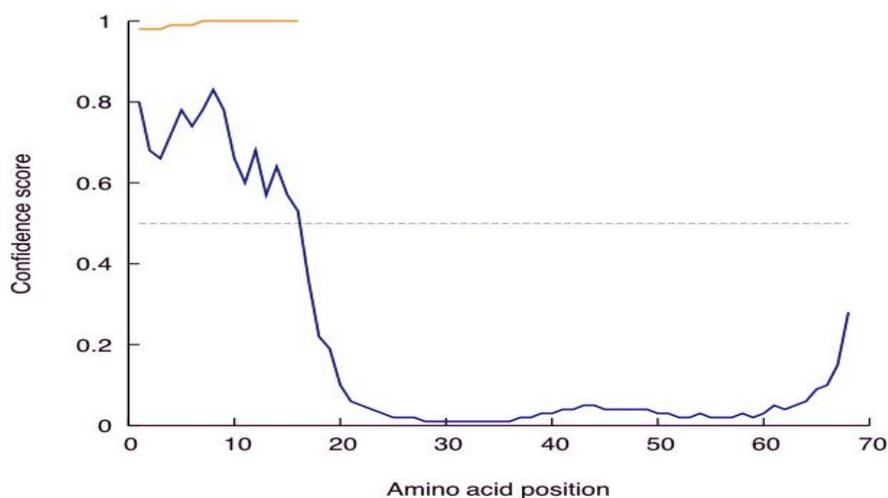


**Figura No. 2 Diagrama PSIPRED para la proteína TGB3 donde muestra la predicción de la estructura secundaria.**

Conf: confianza de la predicción; Pred: Estructura secundaria predicha; AA: Secuencia target; Barras fucsias: hélices; Flechas amarillas: Cadena; Línea negra: coil.

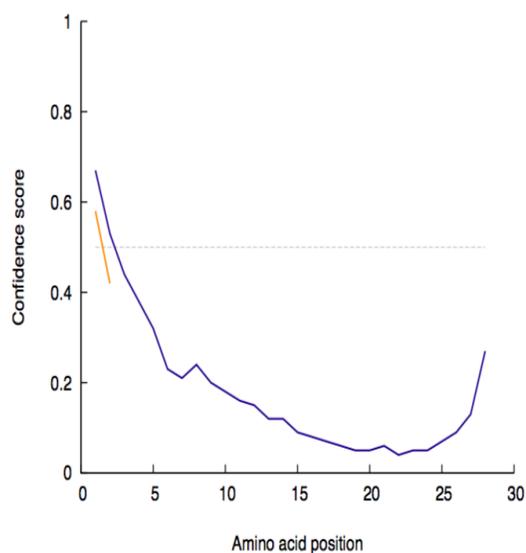






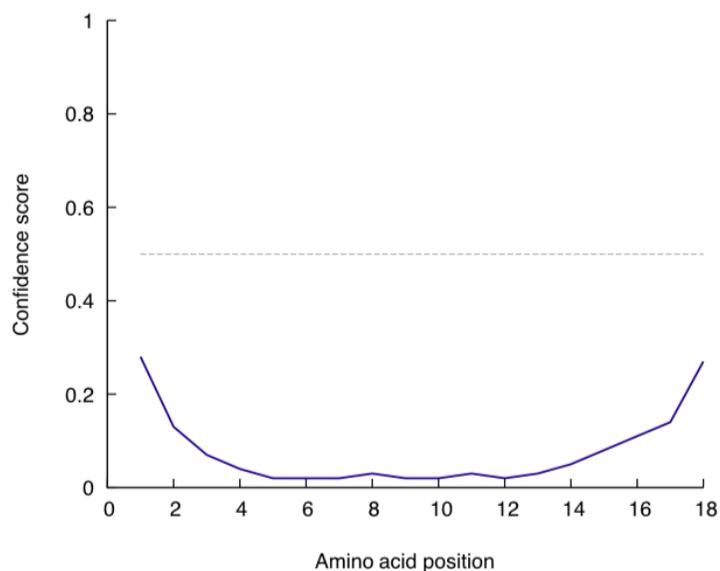
**Figura No. 7 Análisis de desorden intrínseco de la proteína TGB3**

Como se muestra en la figura, los niveles de confianza del modelo en comparación con las posiciones de secuencia como una línea azul continua. La línea gris punteada horizontal marca el límite por encima del cual los aminoácidos se consideran desordenados. La línea naranja muestra la confianza de los residuos desordenados involucrados en las interacciones proteína-proteína.



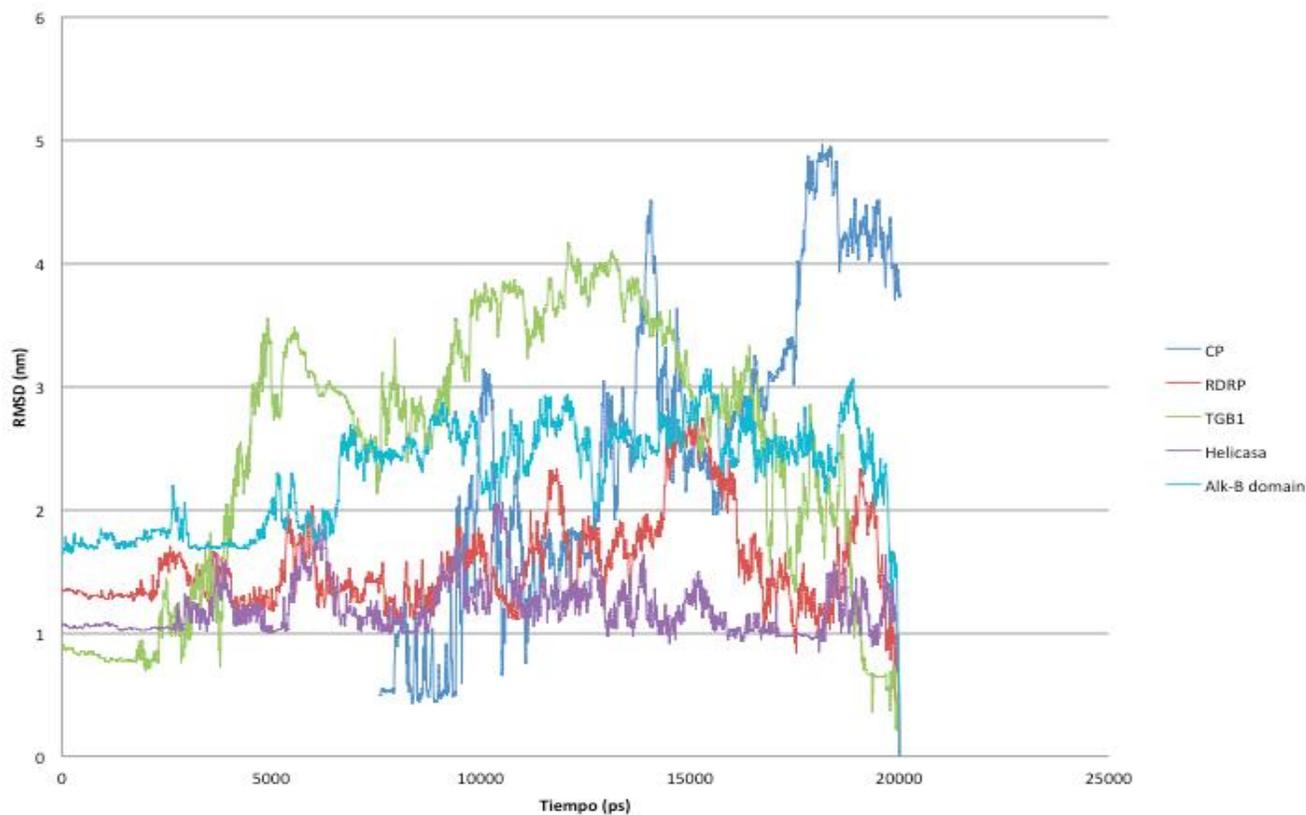
**Figura No. 8 Análisis de desorden intrínseco de la proteína TGB2-helix**

Como se muestra en la figura, los niveles de confianza del modelo en comparación con las posiciones de secuencia como una línea azul continua. La línea gris punteada horizontal marca el límite por encima del cual los aminoácidos se consideran desordenados. La línea naranja muestra la confianza de los residuos desordenados involucrados en las interacciones proteína-proteína.



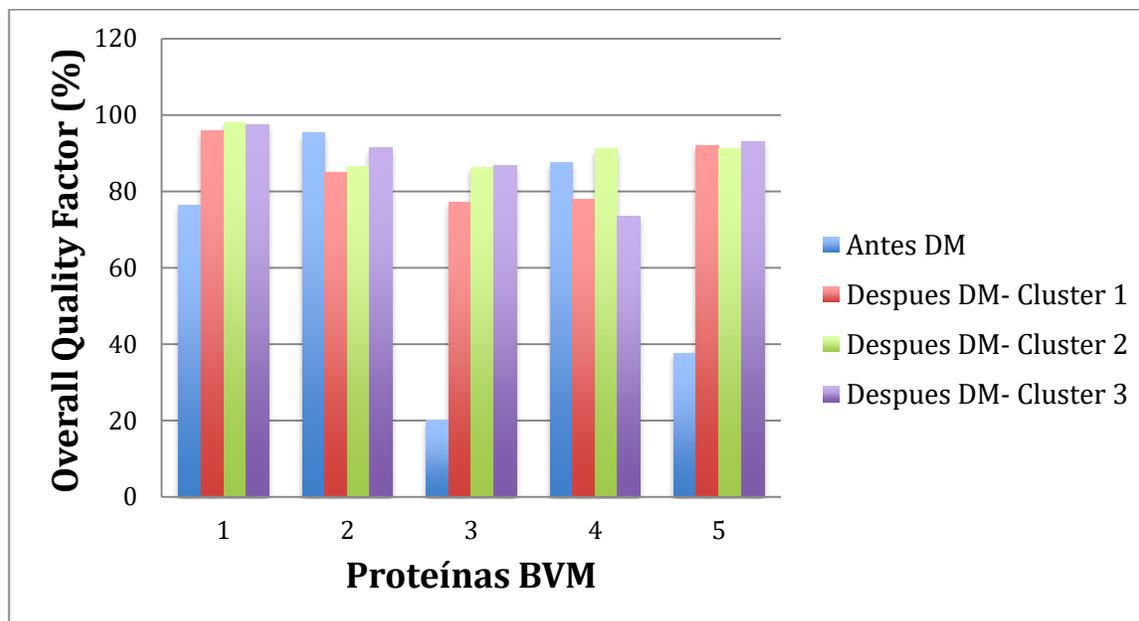
### Figura No. 9 Análisis de desorden intrínseco de la proteína TGB3-helix

Como se muestra en la figura, los niveles de confianza del modelo en comparación con las posiciones de secuencia como una línea azul continua. La línea gris punteada horizontal marca el límite por encima del cual los aminoácidos se consideran desordenados. La línea naranja muestra la confianza de los residuos desordenados involucrados en las interacciones proteína-proteína.



**Figura No. 10 Desviación cuadrática media de los diferentes Frames comparada en diferentes tiempos para las proteínas CP, RDRP, TGB1, Helicasa y AlkB-domain después de la dinámica molecular**

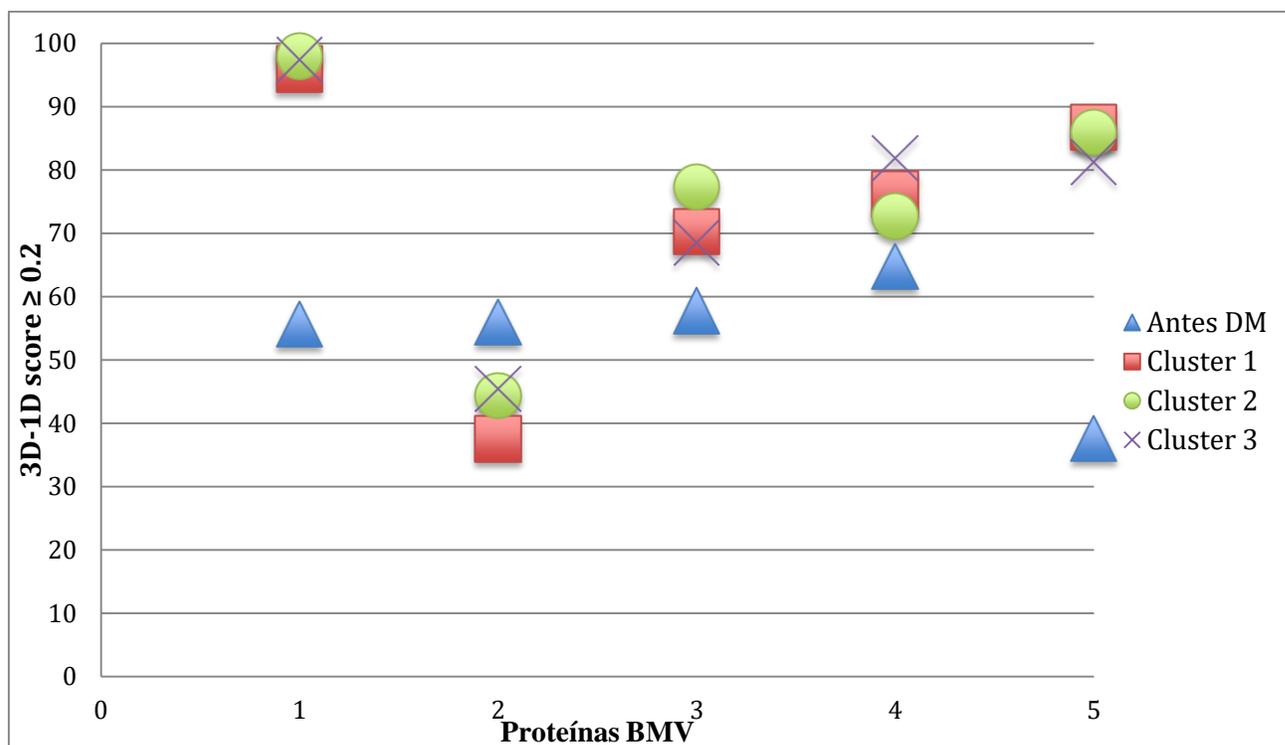
En la Figura se puede analizar la trayectoria de las moléculas mediante la distancia promedio existente entre los átomos de cada una de las proteínas.



**Figura No. 11 Análisis de Calidad ERRAT para cada una de las proteínas del virus de Mosaico del Babaco antes y después de realizar la Dinámica Molecular.**

Como se puede observar en la figura, se esperaba que los modelos mejoren después de haber realizado la dinámica molecular, es decir que el Overall Quality factor sea mayor después de la simulación. Sin embargo, se presentan casos como para la proteína RDRP en donde no se muestra un mejor porcentaje para ninguno de los Clusters.

Respecto a las proteínas del BMV: (1) Cota protein; (2) RDRP; (3) TGB1; (4) AlkB-domain; (5) Helicasa



**Figura No. 12 Análisis de Calidad Verify 3d para cada una de las proteínas del virus de Mosaico del Babaco antes y después de realizar la Dinámica Molecular.**

Como se puede observar en la figura, se esperaba que los modelos mejoren después de haber realizado la dinámica molecular, es decir que el valor del puntaje  $3D-1D > 0.2$  sea cercano a 100% y que sea mayor después de la dinámica molecular. Sin embargo, se presentan casos como para la proteína RDRP en donde no se muestra un mejor porcentaje para ninguno de los Clusters.

Respecto a las proteínas del BMV: (1) Coat protein; (2) RDRP; (3) TGB1; (4) AlkB-domain; (5) Helicasa