Determinación de la estructura terciaria de la proteína citolítica Cyt1Ab de *Bacillus thuringiensis* subespecie *medellín*

MORENO N., Jonathan*

*Departamento de ingeniería química, Universidad de los Andes, Bogotá, Colombia.

RESUMEN: Recientemente se ha encontrado que *Bacillus thuringiensis* también está en la capacidad de sintetizar proteínas citolíticas en fase exponencial contra eritrocitos, es decir, lisan glóbulos rojos. Dado lo anterior, su funcionalidad se puede expandir para aplicaciones en salud, en el tratamiento de células cancerígenas, etc. Sin embargo, la estructura tridimensional de la proteína Cyt1Ab no se ha determinado, luego el objetivo del proyecto era determinar la estructura terciaria de la proteína por homología, el conocimiento de esta podría proporcionar información acerca del mecanismo de inserción de toxinas en la membrana celular. Para la determinación experimental de la estructura terciaria de la proteína citolitica Cyt1Ab producida por *Bacillus thuriengiensis* subesp. *medellín*, se usó el método de modelado por homología, y minimización, obteniendo así una estructura terciaria que no posee restricciones de ninguna clase y que es la conformación de menor energía.

ABSTRACT: Recently has been found that *Bacillus thuringiensis* is also able to synthesize Cytolytic proteins in exponential stage against erythrocytes, that is, kill red globule. Given above, its functionality can be expanded for health applications, in the treatment of cancer cells, etc. However, tridimensional structure of the Cyt1Ab protein has not been determined, so the objective of the project was determine the ternary structure of the protein by homology, the knowledge of this could provide information about the mechanism of membrane-perforating toxins. To determine experimentally the ternary structure of the Cytoytic protein Cyt1Ab produced by *Bacillus thuriengiensis* subsp. *Medellín*, the method of modeling homology and minimization was used, thus obtaining a ternary structure without restrictions of any kind and this is lowest energy conformation.

1. INTRODUCCIÓN

Bacillus thuringiensis es un bacilo esporulado gram-positivo que posee la capacidad de sintetizar proteínas insecticidas usadas ampliamente para el control de pestes en la agricultura o vectores de malaria y fiebre amarilla. Este bacilo se caracteriza por la producción de proteínas parasporales cristalinas (las cuales contienen ∂-endotoxinas), en la fase estacionaria. Además, también sintetiza unas proteínas citolíticas durante la fase exponencial. Las ∂-endotoxinas son altamente tóxicas y especificas para insectos, es decir, estas son activadas por proteasas intestinales, seguidas de una atadura a receptores específicos sobre el intestino medio. Estas interacciones promueven su introducción a la membrana, formando canales iónicos selectivos por oligomerización de monómeros de toxinas, por lo que finalmente, estos insectos mueren y pierden la regulación de la presión osmótica. (Gutierrez, Alzate, & Orduz, 2001) Dicha inserción membranal es provocada por una disminución en el pH que induce un estado de glóbulo líquido a una proteína. (Bravoa, Gillb, & Sobero, 2007)

Como se mencionó anteriormente, las acciones primarias de las proteínas Cry (cristalinas) y las Cyt (citolíticas) son lisar células epiteliales de insectos por inserción dentro de la membrana objetivo y formación de poros. Las toxinas Cyt interactúa directamente con la membrana lipídica y se insertan dentro de esta, a diferencia de las toxinas Cry que interactúan con receptores específicos localizados en la superficie celular del hospedero luego de ser activadas por proteasas, posteriormente las proteínas Cyt se unen a un receptor resultando la formación de estructuras oligoméricas de pre-poros. Las toxinas Cyt son activas contra cepas de dípteros. (Bravoa, Gillb, & Sobero, 2007)

Las proteínas Cyt son proteínas de inclusión parasporal de *Bacillus thuringiensis* que presenta actividad homolítica y tiene similitud de secuencia con una proteína Cyt conocida. Estas toxinas son altamente específicas a un insecto objetivo, e inofensivas para humanos, vertebrados y plantas, además son totalmente biodegradables. (Bravoa, Gillb, & Sobero, 2007)

Las toxinas Cry y Cyt de *Bacillus thuringiensis* pertenecen a la clase de toxinas bacterianas conocidas como toxinas formadoras de poros que son secretadas como proteínas solubles en agua. Allí, estas toxinas experimentan cambios en la conformación con el objetivo de insertarse en las membranas celulares de sus hospederos. Hay dos grupos principales de toxinas formadoras de poros: las toxinas α-helicoidales y las toxinas β-barriles (toxinas Cyt). (Bravoa, Gillb, & Sobero, 2007)

Las identidades de las secuencias primarias entre diferentes secuencias de genes son la base de la nomenclatura de las proteínas Cry y Cyt. Las toxinas Cyt comprenden dos grandes familias de genes relacionadas (Cyt1 y Cyt2). Las toxinas Cyt son

también sintetizadas como protoxinas y una pequeña parte de las terminaciones N y C son removidas para activar las toxinas. Las estructuras terciarias de varias proteínas Cry de tres dominios fueron determinadas por cristalografía de rayos X y por homología. Todas estas estructuras muestran un alto grado de similitud con respecto a la organización con tres dominios, lo cual sugiere un mecanismo de acción similar de la familia de proteínas Cry de tres dominios. (Bravoa, Gillb, & Sobero, 2007)

El mecanismo de acción de las toxinas Cyt esta aun bajo controversia. De acuerdo a una aproximación, las dos capas externas de hélices alfa cerradas oscilan lejos de las laminas beta que se encuentra en contacto con la membrana, a las últimas tres laminas beta se les permite insertarse dentro de la membrana. Luego, ocurre la oligomerización y la formación de poros con barriles beta. Otras aproximaciones dicen que el lado hidrofílico de las hélices interactúa con los residuos de otros monómeros para formar oligómeros. La arquitectura alfa/beta con una lámina beta rodeada de dos capas de hélices alfa que representan un pliegue citolítico. (Cohen, y otros, 2008)

A la fecha, las estructuras de tres proteínas Cry han sido reportadas. Estas están compuestas de tres dominios, además de estructuras superiores similares entre ellas (a pesar de la baja homología de aminoácidos), lo anterior sugiere la conservación de varios rasgos estructurales entre ∂-endotoxinas. Uno de estos dominios por ejemplo, es responsable de la formación de canales iónicos (Gutierrez, Alzate, & Orduz, 2001).

Un análisis de las relaciones filogenéticas de los dominios aislados de los miembros de la familia Cry de tres dominios revela interesantes rasgos similares considerando la creación de diversidad en esta familia de proteínas. Los dominios I y II han coevolucionado. El análisis de las secuencias del dominio III, revela una topología diferente debido al hecho de que diferentes toxinas intercambian dominios III. Algunas toxinas con especificidad dual, son claros ejemplos de este intercambio. La evolución independiente de los tres dominios estructurales y el intercambio del dominio III entre diferentes toxinas genera proteínas con similar mecanismo de acción pero con muy diferentes especificidades. (Bravoa, Gillb, & Sobero, 2007)

1.1 Modelado por homología

Es una técnica para predecir la estructura de una proteína, también llamada modelado comparativo. Esta es una clase de métodos para construir un modelo de resolución atómica de una proteína a partir de su secuencia de aminoácidos objetivo. La técnica de modelado por homología depende de la identificación de una o más estructuras de proteínas conocidas, llamadas patrón, molde o plantilla, que puedan parecerse a la secuencia objetivo, para producir una alineación del mapa de residuos entre la secuencia objetivo y la secuencia patrón. La alineación de secuencias y la estructura patrón son usadas para producir un modelo estructural de la estructura objetivo. Debido a que las estructuras de proteínas son más conservadas que la secuencias de proteínas, los niveles detectables de similitud en las secuencias usualmente implican una similitud estructural significativa. (Wikipedia)

La calidad del modelo por homología es dependiente de la calidad de la alineación de las secuencias y de la estructura patrón. Este enfoque puede ser complicado por la presencia de vacios de alineación que indican la existencia de una región estructural en el objetivo pero no en el patrón, y esto a su vez se deriva de una pobre resolución en el procedimiento experimental (por lo general cristalografía de rayos X) utilizada para resolver la estructura. Hay una disminución de la calidad del modelo con una disminución de identidad entre la secuencia. Sin embargo, los errores son significativamente más altos en las regiones de curvas, donde las secuencias de aminoácidos de proteínas del objetivo y el patrón pueden ser completamente diferentes. (Wikipedia)

Regiones del modelo que se construyeron sin una estructura patrón, por lo general, por el modelado de curvas, generalmente son mucho menos precisos que los del resto del modelo, particularmente si las curvas son largas. Errores en el acoplado de la cadena lateral y su posición también aumenta con la disminución de la identidad, es decir, que las variaciones en las configuraciones de acoplado se han sugerido como una razón importante para la pobre calidad del modelo a baja identidad. En conjunto, estos errores de posicionamiento atómico son importantes e impiden el uso de modelos por homología con fines que requieren datos de resolución atómica, tales como el diseño de fármacos y la predicción de la interacción proteína-proteína e incluso la estructura cuaternaria de una proteína puede ser difícil de predecir a partir de estos modelos estructurales de sus subunidades. Sin embargo, los modelos por homología pueden ser útiles para llegar a conclusiones *cualitativas* sobre la bioquímica de la secuencia objetivo, especialmente en la formulación de hipótesis acerca de por qué determinados residuos se conservan, y que a su vez puede conducir a experimentos para probar estas hipótesis. Por ejemplo, la disposición espacial de los residuos conservados puede sugerir si un residuo particular se conserva para estabilizar el plegamiento al participar en algunas uniones en ciertas moléculas pequeñas, o al fomentar la asociación con otra proteína o ácido nucléico. (Wikipedia)

El modelado por homología puede producir modelos estructurales de alta calidad cuando el objetivo y la estructura patrón están estrechamente relacionadas. Las imprecisiones mayores del modelado

por homología, empeoran con la identidad baja de las secuencias, derivadas de errores en la alineación inicial de la secuencia y de la inadecuada selección de la estructura patrón. (Wikipedia)

El método de modelado por homología es basado en las observaciones de que la estructura terciaria de una proteína es mejor conservada que la secuencia de aminoácidos. De esta manera, todas las proteínas que han divergido en secuencias pero aún comparten similitudes detectables puede que aún compartan propiedades estructurales, especialmente el número total de pliegues. Debido a que es difícil obtener experimentalmente estructuras de métodos tales como cristalografía de rayos X y NMR para todas las proteínas de interés, el método de modelado por homología puede proporcionar unos útiles modelos estructurales para generar hipótesis acerca de las funciones de la proteína y para realizar un mayor trabajo experimental. (Wikipedia)

1.1.1 Pasos en la producción del modelo

En el proceso de determinar un modelo mediante Homología se pueden tomar como base 4 pasos:

i. Selección del molde o patrón con el cual se planea realizar la comparación.
ii. Alineamiento entre el patrón y el objetivo.
iii. Construcción del modelo.
iv. Evaluación del modelo.

Por lo general la selección del patrón con el cual se va a comparar y el alineamiento de este con el objetivo se realizan al mismo tiempo debido a que los métodos utilizados de forma frecuente se apoyan en la producción de alineamiento de la secuencia, sin embargo esos alineamientos puede que no sean cien por ciento confiables ya que al realizar una búsqueda en las bases de datos se busca optimizar tiempo y no calidad en los alineamientos. Estos procedimientos se pueden realizar de forma iterativa para mejorar el modelo final. (Wikipedia)

1.1.2 Selección del patrón y Alineamiento de la Secuencia

Uno de los pasos más importantes para la determinación de un modelo mediante homología es la selección del mejor patrón estructural. El método más simple de identificación se basa en la identificación en serie de parejas de alineamientos secuenciales usando la ayuda de técnicas de búsqueda en bases de datos como lo son el método FASTA el cual utiliza un método de alineamiento de secuencias locales y BLAST el cual utiliza el mismo método pero este solo proporciona patrones con un cierto número de coincidencias entre el patrón y el objetivo. Otra técnica que se puede utilizar para la obtención del patrón es el Trenzado de Proteínas el cual utiliza el reconocimiento de pliegues. Existen situaciones que pueden llevar a que se escoja un patrón que tenga una función similar a la de la secuencia de consulta o objetivo. Un acercamiento mejor se logra al enviar la secuencia primaria a servidores de reconocimiento de pliegues. (Wikipedia)

1.1.3 Generación del Modelo

Dado un modelo y una alineación, se puede usar esta información para generar un modelo estructural tridimensional del objetivo, representado como un conjunto de coordenadas cartesianas para cada parte de la proteína. Existen tres clases de métodos para la generación de modelos. (Wikipedia)

i. Ensamblaje de Fragmentos

El método original para el modelaje de proteínas mediante homología se basa en el ensamblaje del un modelo completo de fragmentos estructurales conservados e identificados que están relacionados de forma cercana con estructuras ya encontradas. Las proteínas sin desarrollar se pueden modelar realizando primero una construcción del núcleo con los fragmentos conocidos y después sustituyendo las regiones variables por la de otras proteínas ya resueltas. (Wikipedia)

ii. Asociación de Segmentos

El método de asociación de segmentos divide el objetivo en una serie de segmentos cortos y cada uno de estos es asociado con su patrón obtenido de la base de datos de proteínas. El único problema que puede tener es que el alineamiento se realiza sobre los segmentos y no sobre la proteína completa. La selección del patrón para cada segmento se basa en: similitud en la secuencia de aminoácidos, comparación de coordenadas de los carbones α y predicciones de conflictos estéricos generados por los radios de Van der Waals de los átomos divergentes entre el objetivo y el patrón. (Wikipedia)

iii. Satisfacción de Restricciones Espaciales

El método actual más común para hallar el modelo mediante homología se inspira en datos de espectroscopia de resonancia magnética nuclear. Los alineamientos entre el patrón y el objetivo se construyen generalmente con un conjunto de criterios geométricos que luego se usan para hallar funciones de densidad de probabilidad para cada restricción. Las restricciones aplicadas a las principales coordenadas internas de la proteína sirven como base para la optimización del procedimiento, el cual también mediante la minimización del gradiente de energía refina las posiciones de los átomos pesados en la proteína. (Wikipedia)

1.1.4 Modelaje de Bucles

Regiones de la secuencia de la proteína la cual se quiere modelar y no tienen un patrón con el cual hacerlo, se modelan mediante el modelaje de bucles; estas regiones son más propensas a errores de modelaje y estos ocurren con mayor frecuencia cuando el objetivo y el patrón tienen poca igualdad en sus secuencias. Las coordenadas de secciones sin acoplar determinadas mediante este método son generalmente menos certeras que esas obtenidas de una estructura ya conocida, especialmente si el bucle tiene más de 10 residuos. Los primeros dos ángulos diédricos consecutivos (ξ_1 y ξ_2) pueden ser estimados entre 30° para un esqueleto estructural acertado. Pequeños errores en ξ_1 pueden causar errores en las posiciones de los átomos al final de la cadena; estos átomos por lo general tienen una importancia funcional especialmente si están ubicados cerca al sitio activo. (Wikipedia)

1.1.5 Evaluación del Modelo

La evaluación de las estructuras halladas mediante homología sin tener la estructura verdadera se realiza generalmente mediante dos métodos: Potenciales Estadísticos y Cálculos Energéticos. Ambos métodos estiman la energía del modelo o modelos a evaluar. Ninguno de estos dos métodos se relaciona excepcionalmente bien con la estructura verdadera de forma certera. (Wikipedia)

El método de Potenciales Estadísticos es un grupo de métodos empíricos basados en la frecuencia de contacto observada entre residuos, dentro de las estructuras de las proteínas encontradas en la Base de Datos de Proteínas (PDB). Estos métodos asignan una probabilidad o un puntaje de energía a cada posible pareja de interacción entre aminoácidos y combinan estos puntajes de interacción en una sola puntuación para todo el modelo. Unos de esos métodos pueden producir también una evaluación de residuo-residuo que identifica las regiones del modelo que tienen una baja puntuación, esto incluso si el modelo tiene una puntuación general aceptable. Estos métodos se centran en los núcleos hidrofóbicos y en los aminoácidos polares expuestos a solventes presentes comúnmente en proteínas globulares. Los métodos de Potenciales Estadísticos son más eficientes computacionalmente que los de cálculos energéticos. Ejemplos de estos métodos son *Prosa* y *DOPE.* (Wikipedia)

Los métodos de Cálculos Energéticos tratan de obtener las interacciones entre átomos que son responsables por la estabilidad de la proteína en solución, especialmente debidas por las interacciones electrostáticas de Van der Waals. Estos cálculos se realizan usando un campo de fuerzas de mecánica molecular; las proteínas comúnmente son tan largas que no es posible realizar cálculos basados en mecánica cuántica. El uso de estos métodos está basado en la hipótesis de panorama energético del plegado de proteínas, el cual predice que el estado originario es el que tiene el mínimo de energía. Estos métodos por lo general usan una solvatación implícita, la cual proporciona una aproximación continua del baño de disolvente para una molécula de proteína sin tener la representación explicita de moléculas individuales del solvente. El campo de fuerzas generado para la evaluación del modelo se conoce como Campo de Fuerza Efectivo (EFF) y este se basa en parámetros atómicos de *CHARM*, el cual genera los EEF basándose en los átomos que contiene la proteína a estudiar. (Wikipedia)

Existen programas los cuales proporcionan un reporte de la validación del modelo como *What Check y What If*, estos programas proporcionan un reporte de aproximadamente 200 aspectos científicos y administrativos del modelo. (Wikipedia)

1.1.6 Métodos de Comparación Estructural

La exactitud de la evaluación de modelos producidos mediante homología es sencilla cuando la estructura experimental es conocida. El método más común para comparar la estructura de dos proteínas es mediante en cuadrático medio o RMSD para medir la distancia media entre los átomos correspondientes entre las dos estructuras después de que se han superpuesto. Sin embargo, el cuadrático medio subestima los modelos en los cuales el núcleo esta modelado correctamente pero algunas regiones de bucles flexibles son inexactas. Un método para la evaluación experimental de modelos *CASP* es conocido como el Test de Distancia Global (GDT) y este mide el total de átomos los cuales a distancia y al superponerlos están bajo un límite. Ambos métodos pueden ser usados para un grupo de átomos en la estructura, pero frecuentemente se aplican únicamente al carbono α o a los átomos del esqueleto estructural de la proteína para minimizar los efectos de ruido. (Wikipedia)

A gran escala varias evaluaciones comparativas se han realizado para evaluar la calidad relativa de los diversos métodos actuales de modelado por homología. CASP es una comunidad experimental de predicción que evalúa los problemas de predicción al presentar los modelos estructurales de una serie de secuencias cuyas estructuras han sido recientemente resueltas experimentalmente, pero aún no han sido publicados. (Wikipedia)

La exactitud de las estructuras generadas por el modelado por homología es altamente dependiente de la identidad de secuencia entre el objetivo y el patrón. Más del 50% de identidad entre secuencias, sugiere que los modelos tienden a ser confiables. En las secuencias de alta identidad, la principal fuente de error en el modelado por homología se deriva de la elección del patrón en que se basa el modelo.

Identidades bajas presentan graves errores en la alineación de secuencias que inhiben la producción de modelos de alta calidad. (Wikipedia)

Graves errores locales en los modelos de homología pueden surgir de una mutación, en consecuencia la estructura debe resolverse aun en una región de la secuencia objetivo para el que no hay patrón correspondiente. Este problema puede minimizar el uso de varios patrones, pero el método es complicado por las diferencias de estructuras locales del patrón y de todo el hueco, y por la probabilidad de que una región que falta en una estructura experimental falta también en otras estructuras de la misma familia de proteínas. Regiones que faltan son más comunes en los bucles locales donde la alta flexibilidad aumenta la dificultad de resolver la región por la estructura de los métodos de determinación. Aunque algunos ofrecen orientación, incluso con un solo patrón, por la posición de los extremos de la región que falta, cuanto mayor sea la diferencia, más difícil es modelo. Bucles de hasta unos 9 residuos pueden ser modelados con una precisión moderada en algunos casos, si la alineación local es correcta. (Wikipedia)

El modelado por homología y el trenzado de proteínas son métodos basados en plantillas o patrones y no hay un límite entre el modelado por homología y el trenzado de proteinas en términos de técnicas de predicción. Pero las estructuras proteínicas objetivo son diferentes. El modelado por homología es para objetivos que tienen proteínas homologas con estructuras conocidas, en cambio, en el trenzado de proteína es para objetivos con solamente se encuentre homología a nivel de pliegues. En otras palabras, el modelado por homología es para estructuras fáciles y el trenzado de proteínas es para estructuras complejas. (Wikipedia)

El modelado por homología trata el patrón en una alineación como una secuencia y solamente la homología en la secuencia es usada para predicción. El trenzado de proteínas trata el patrón en una alineación como una estructura y secuencia, de las cuales se obtiene la información para predecir. Cuando no hay homología significativa, el trenzado de proteínas puede hacer una predicción basada en la información de la estructura, lo que explica por qué este puede ser más eficiente que modelado por homología en muchos casos. En la práctica, cuando la identidad de la secuencia en la alineación de la secuencia es baja, el modelado por homología podría no producir una predicción significativa. En este caso, si hay poca homología sobre el objetivo, el trenzado de proteínas puede generar una buena predicción. (Wikipedia)

El uso de los modelos estructurales incluyen la predicción de la interacción proteína-proteína, los acoplamientos proteína-proteína, acoplamientos moleculares, y la funcionalidad de los genes identificados en el genoma de un organismo. Incluso los modelos de homología de baja precisión puede ser útil para estos fines, porque sus imprecisiones tienden a ubicarse en los bucles en la superficie de la proteína, que normalmente son más variable, incluso entre las proteínas íntimamente relacionadas. Las regiones funcionales de las proteínas, especialmente en su sitio activo, tienden a ser más conservadas y por lo tanto con mayor precisión deben ser modeladas. (Wikipedia)

1.2 Predicción de la estructura de una proteína por el trenzado de proteínas

Mediante el uso de medidas simples se puede determinar la idoneidad de los diferentes tipos de aminoácidos para los entornos estructurales locales definidos en términos de accesibilidad del disolvente y de la estructura secundaria de la proteína, los autores derivan entonces simples pero novedosos enfoques para evaluar si una secuencia de proteínas encaja bien con un pliegue estructural de una proteína dada. El trabajo de Jones, Taylor y Thornton sobre el reconocimiento de pliegue en una proteína permitió el desarrollo de una nueva gama de poderosas herramientas para la predicción de estructuras de proteínas, la cual ahora se denomina *trenzado de proteínas*. Esas herramientas computacionales han jugado un rol clave en la ampliación de la utilidad de todos los solucionadores de estructuras experimentales por cristalografía de rayos X y resonancia magnética nuclear (NMR), probando modelos estructurales y funcionales predichos por muchas de las proteínas codificadas en los cientos de genomas que han sido secuenciados hasta ahora. (Xu, Xu, & Liang, 2007)

Lo que ha hecho atractivo al método de trenzado de proteínas como una herramienta predictiva de estructuras de proteínas es que la observación del número de estructuras plegadas únicas en la naturaleza es pocos órdenes de magnitud inferiores que el número de proteínas en la naturaleza. Cada organismo posee por lo menos miles de diferentes proteínas codificadas en el genoma, el número total de proteínas diferentes en la naturaleza es de por lo menos diez billones, sin considerar las variaciones de proteínas unidas. Esta discrepancia sugiere un paradigma para resolver todas las estructuras de proteínas combinando los enfoques computacionales y experimentales, esto es, resolver estructuras de proteínas con únicas estructuras de pliegue usando técnicas experimentales y modelos computacionales del resto de proteínas usando estructuras experimentales como patrones. (Xu, Xu, & Liang, 2007)

La idea básica del trenzado de proteínas es ubicar los aminoácidos de una secuencia proteínica consultada, siguiendo sus órdenes secuenciales y permitiendo espacios, dentro de posiciones estructurales de una estructura patrón de una manera óptima medida mediante puntajes de idoneidad. Estos alineamientos

estructurales de secuencias son calculadas usando medidas estadísticas de la probabilidad total y energéticas de la proteína objetivo adoptando cada uno de los pliegues estructurales. La mejor alineación estructural de las secuencias proporciona una predicción de la columna vertebral de los átomos de la proteína objetivo, basado en su posicionamiento en la estructura patrón. Actualmente, el trenzado de proteínas está siendo usado en biología molecular y en laboratorios de bioquímica, frecuentemente para estudios iniciales de proteínas objetivos, debido a que este método proporciona la información estructural y funcional, los cuales podrían ser usados para guiar diseños experimentales e investigación. (Xu, Xu, & Liang, 2007)

Como una técnica predictiva, el trenzado de proteínas tiene un número de grandes desafíos computacionales y problemas de modelación (Xu, Xu, & Liang, 2007). Estas incluyen:

- ¿Cómo medir eficientemente y precisamente la idoneidad de una secuencia localizada en una estructura patrón?
- ¿Cómo encontrar eficientemente y precisamente la mejor alineación entre la secuencia objetivo y la estructura patrón basado en un determinado conjunto de medida de idoneidad?
- ¿Cómo calcular cuales alineaciones estructurales de secuencias y diferentes estructuras patrones representan un correcto reconocimiento de pliegue y una exacta predicción de estructuras?
- ¿Cómo identificar cuales partes de una estructura predicha son exactas y cuales partes no?

Por encima de un millón de secuencias proteínicas han sido determinadas, entre las cuales alrededor de treinta mil han obtenido su estructura ternaria resuelta experimentalmente. Dado que podría haber por lo menos diez billones de proteínas diferentes en la naturaleza, ¿Cuántas estructuras únicas de proteínas o estructuras de pliegues estas proteínas podrían haber adoptado?. Los dominios son las unidades estructurales básicas de las proteínas. Un dominio estructural es una unidad estructural clara y compacta que podría plegar independientemente de otros dominios. (Xu, Xu, & Liang, 2007)

Mientras muchas proteínas solo tienen un dominio, hay proteínas con dos, tres e incluso más dominios estructurales. Para organismos eucariotas, fue estimado que por al menos dos terceras partes son proteínas de múltiples dominios. Hay una clara necesidad de automatizar los métodos de partición, para poder resolver eficientemente las estructuras proteínicas. Actualmente programas de computador están siendo usados para la partición de una estructura proteínica en dominios individuales. (Xu, Xu, & Liang, 2007)

Un dominio de una proteína podría ser parte de diferentes estructuras proteínicas, por combinación con otros dominios. Ya que los dominios son las unidades estructurales básicas de las proteínas, se están realizando estudios sobre el número de las estructuras plegadas en la naturaleza, por lo que han sido llevados a cabo sobre los dominios de las proteínas en lugar de toda la proteína. (Xu, Xu, & Liang, 2007)

La estimación del número de pliegues estructurales únicos es obtenida a través de la estimación del número de familias que cada pliegue estructural cubre y estudia la distribución de cobertura por todas las estructuras de pliegue conocidas. El número de pliegues estructurales disminuye como su cobertura de familias de proteínas incrementa. Hay tres clases de pliegues estructurales, *unifolds*, *mesofolds* y *superfolds*. *Unifolds* representan pliegues estructurales que cubren solo una familia de proteínas, *superfolds* representan los pliegues estructurales, los cuales cada uno cubre muchos pliegues estructurales, y *mesofolds* representa los pliegues estructurales en medio. (Xu, Xu, & Liang, 2007)

2. MATERIALES Y METODOS

El modelado por homología, además de práctico, requiere un mínimo de instrumentos y una muy reducida información para su puesta en marcha. Dicha información comprende el conocimiento de la secuencia de aminoácidos objetivo, que en este caso es la de la proteína Cyt1Ab, la cual fue consultada en la base de datos Protein Data Bank (PDB). Asimismo se debe conocer la secuencia de aminoácidos patrón y la estructura terciaria de esta proteína, que para nuestro caso es Cyt2Ba.

Para la alineación de las secuencias y el posterior modelado por homología se utilizó la herramienta de alineamiento estructural del programa SwissPdbViewer® y se corrigió manualmente hasta que un satisfactorio posicionamiento de bloques conservados e identidades de aminoácidos fueran obtenidos. Para la evaluación energética de la estructura, y el análisis de la idoneidad del modelo se usó el servidor SwissModel en *the expasy server* (*http://swissmodel.expasy.org/*) y un modelo preliminar para Cyt1Ab fue devuelto con su respectiva validación a cargo de los programas PROCHECK y WHAT IF. Cabe aclarar que el paso anterior se repetía cada vez que una minimización sobre la estructura fuera hecha. Para la minimización de la estructura se recalculaban las conformaciones curvas y los tamaños de las cadenas con campos de fuerza sin restricciones de distancia con la ayuda de HyperChem®, en el cual se realizó un trabajo iterativo que permitiría obtener la estructura terciaria de menor energía. Finalmente, para

la presentación de las estructuras se hizo uso de Pymol®.

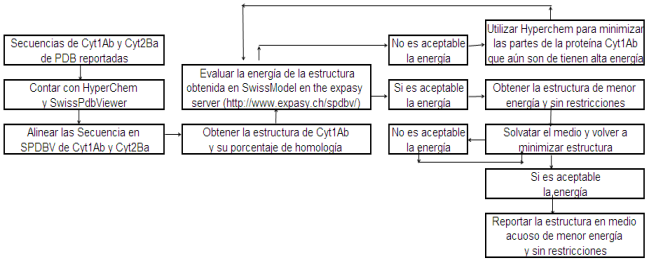

Ilustración 1. Diagrama de Flujo del modelado por Homología y la respectiva minimización de la estructura

3. RESULTADOS

Para la selección de la proteína patrón se tuvo en cuenta que fuera de la misma familia de Cyt1Ab, la proteína que se escogió fue Cyt2Ba. Con lo anterior se trataba de garantizar que se conserve el pliegue que caracteriza a su familia, y por lo tanto se esperaría que tuviera un mecanismo de inserción membranal similar a la de las proteínas citolíticas.

```
Cyt1Ab          1    MENPNHCPLE DIQVNPWKTP QSKARVITLR IDDPNEINNL LSINEIENTN
Cyt2Ba          41                                        TN FNEIFYVEPQ

Cyt1Ab          51   YLLQAIMLAN AFQKALVPTS TEFAEDALQF SMTKGLEVAN TISPPGAVVQ
Cyt2Ba          53   YIAQAIRLTN TFQGAIDPLT -------LNF NFEKALQIAN GL-PNAGVTG

Cyt1Ab          101  YVDQNVSQTN NQVSAMINKV LDVLKSILGV ALGQSVI-EQ LTSAVTNTFT
Cyt2Ba          95   TINQSVIHQT IEVSVMISQI KEIIRSVLGL VINSANFWNS VVSAITNTFT

Cyt1Ab          150  NLNTQKNEAW IFWGRETSTQ TNYTYNVLFA IQNGQTGGVM YCVPVGFEIK
Cyt2Ba          145  NLEPQVDENW IVWRNLSATQ TSYFYKILFS IQNEDTGRFM AILPIAFEIT

Cyt1Ab          200  VSAVKERVLF LTIQDSASYN VNIQSLKFAQ PLVSASEYPI ADLTSAINGT
Cyt2Ba          195  VDVQKQQLLF ITIKDSARYE VKMKALTVVQ ALDS

Cyt1Ab          250  L
Cyt2Ba
```

Ilustración 2. Alineación de la secuencia de aminoácidos de Cyt1Ab y Cyt2Ba

El porcentaje de homología de Cy1Ab con Cyt2Ba es de 40,9%.

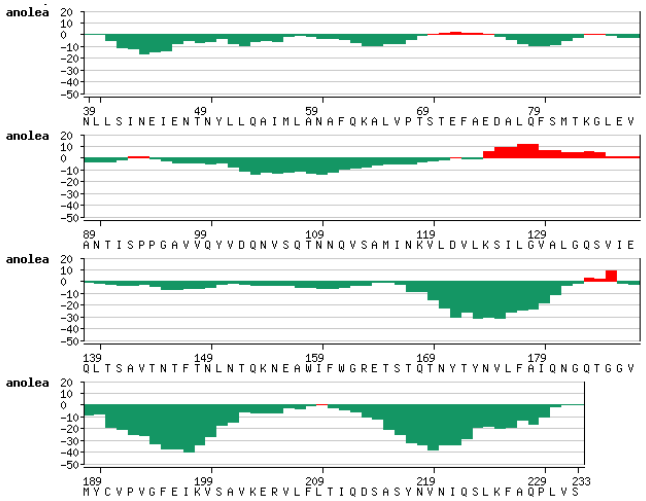

Ilustración 3. Evaluación energética de cada aminoácido de Cyt1Ab sin minimizar

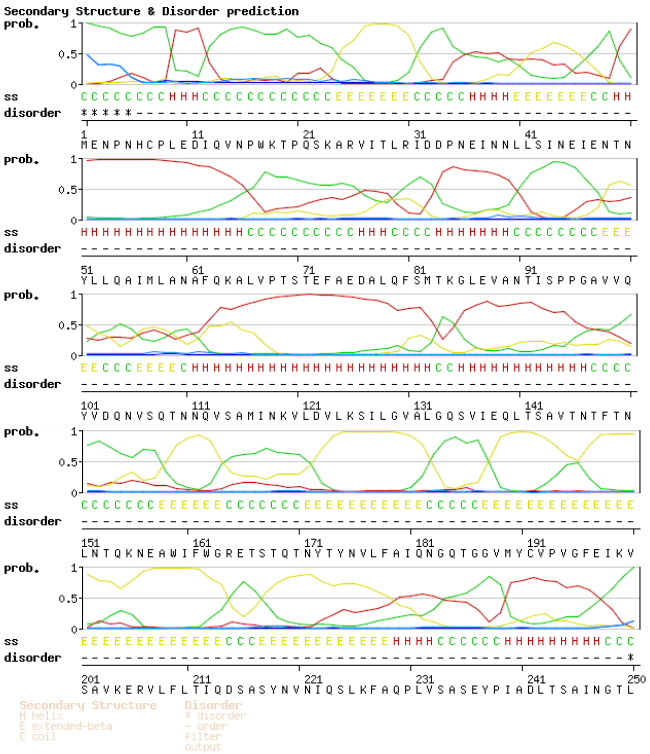

Ilustración 4. Predicción de las estructuras secundarias para la secuencia de aminoácidos

Tras realizar la minimización de todas las zonas que presentaban energía elevada se obtuvieron los siguientes resultados de energía:

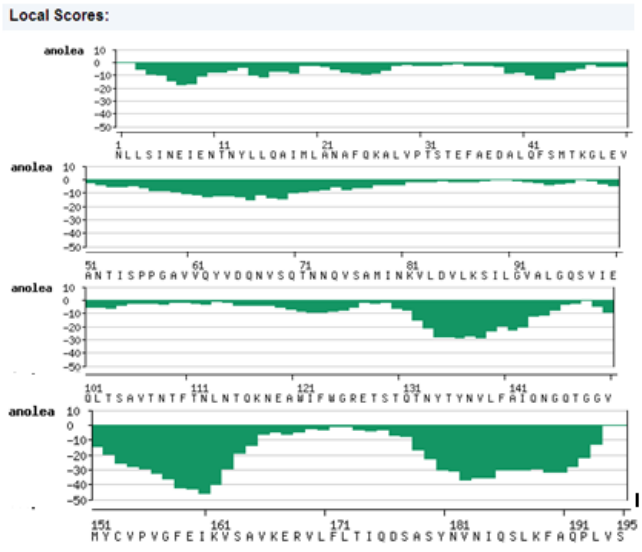

Ilustración 5. Evaluación energética de cada aminoácido de Cyt1Ab minimizada

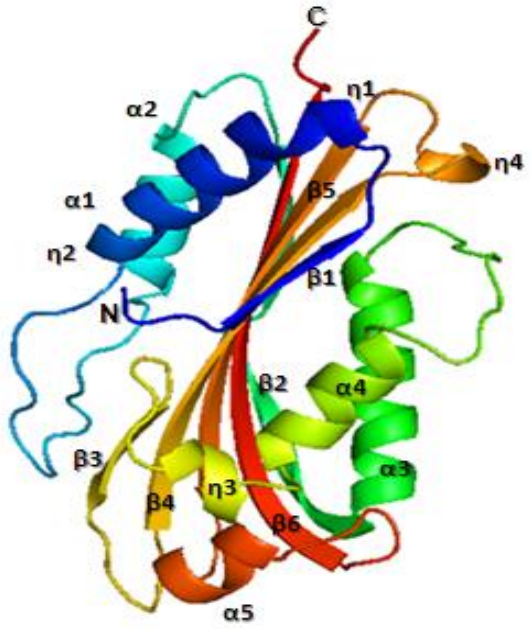

Ilustración 6. Estructura terciaria de la proteína Cyt1Ab minimizada

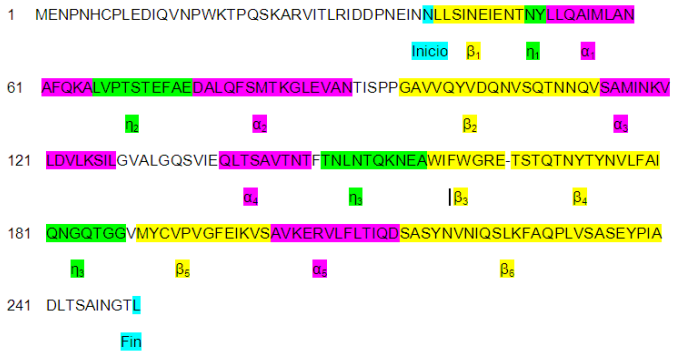

Ilustración 7. Secuencia de aminoácidos con sus respectivos estructuras secundarias

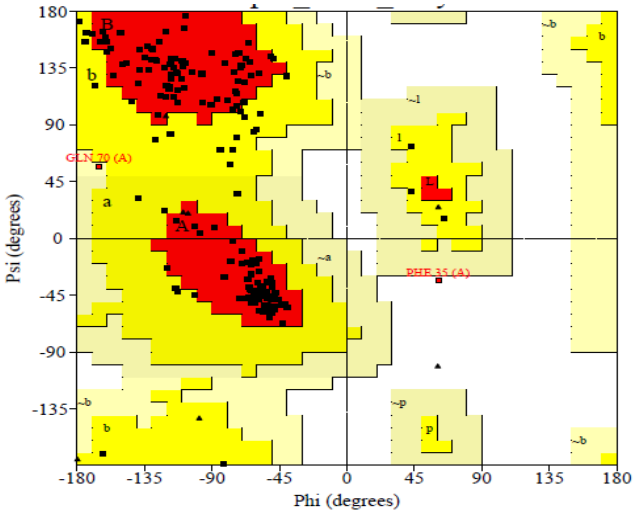

Ilustración 8. Gráfico de Ramachandran de la estructura Cyt1Ab minimizada

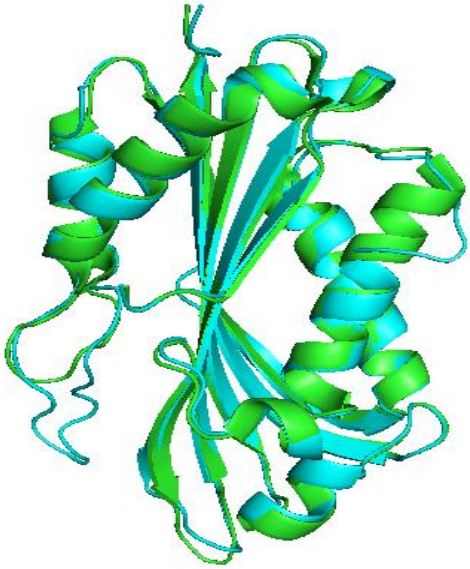

Ilustración 9. Superposición de la Cyt2Ba (Verde) y Cyt1Ab (Azul)

Ilustración 10. Superposición de la Cyt1Ab Minimizada (Roja) y la Cyt1Ab No minimizada (Amarilla)

Ilustración 11. Superposición de la estructura Cy1Ab Solvatada (Roja) y la Cyt1Ab Aislada (Azul)

4. ANALISIS DE RESULTADOS

Dado que Cyt1Ab y Cyt2Ba comparten un 40% de homología, se espera que tengan similares dominios, y por ende un mecanismo de acción similar. Evidencia de lo anterior se puede observar en la *ilustración 8*, en la cual se muestra la superposición de la Cyt2Ba y Cyt1Ab. Allí se ve que la arquitectura α/β con láminas β rodeadas de dos capas de hélices α, el cual representa el pliegue citolítico, se conserva entre estas proteínas de la misma familia.

La *ilustración 3* muestra los resultados de la evaluación de energía de cada aminoácido de Cyt1Ab sin minimizar; dichas gráficas permiten identificar las zonas de la proteína que deben ser minimizadas. Tales zonas son: STEFAE que representa una curva en la estructura terciaria que además presenta muy baja homología con respecto a la Cyt2Ba; otra zona que debe ser minimizada es la curva que une las hélices α_3 y α_4; por otro lado también se debe minimizar la secuencia QTG que representa otra curva anterior a la lámina β_6.

En la *ilustración 4* se presenta un análisis preliminar de las estructuras secundarias que son más probables para cada aminoácido, las cuales se pueden contrastar con las estructuras presentadas por el modelo, localizadas en la *ilustración 7*. Luego de comparar estos dos resultados se llega a la conclusión de que el análisis preliminar fue certero en cuanto a las mayorías de aminoácidos de la estructura encontrada por homología.

En la *ilustración 5* se presenta los resultados de la evaluación energética de la estructura de la proteína Cyt1Ab minimizada. Para observar las implicaciones de la minimización en la estructura de manera global se presenta la superposición de la estructura minimizada y no minimizada localizada en la *ilustración 10*, en esta, se puede evidenciar que las zonas que anteriormente se mencionaron son las que presentan una orientación diferente del espacio, lo cual era de esperarse a raíz de que se debía obtener la organización de menor energía.

En la *ilustración 6* se presenta la estructura de la proteína Cyt1Ab minimizada, con sus respectivas estructuras secundarias que constituyen el único dominio representativo del pliegue citolítico. Esta estructura está compuesta por 6 láminas β antiparalelas rodeadas de una capa de hélices α de α_1 y α_2 y otra capa de α_3-α_4-α_5, las cuales representan el pliegue citolítico. Las láminas β más largas β_4, β_5, β_6 son las encargadas de insertarse en la membrana. Las dos capas externas de hélices α oscilan lejos de láminas β que se encuentran en contacto con la membrana. Por otro lado, las hélices α son las encargadas de la oligomerización, que es la agregación de moléculas para la formación de una estructura más compleja formada por subunidades independientes. En cuanto a la inserción de las láminas β, esta se da entre las hélices α_1 y α_2. Finalmente se puede decir que el aminoácido que representa en C terminal es resistente a la proteólisis en asociación con la membrana.

Del gráfico de ramachandran se puede observar que cerca del 90% de las regiones de la estructura son favorables, por lo que se dice que una calidad buena puede ser esperada del modelo.

En la *ilustración 11* se puede observar las diferencias estructurales entre la estructura minimizada aislada de Cyt1Ab y la estructura minimizada en medio acuosa de Cyt1Ab. Las diferencias más importantes se encuentran en el N terminal y en las curvas que presentaban mayor energía antes del proceso de minimización.

5. CONCLUSIONES

Se obtuvo una estructura terciaria para Cyt1Ab de menor energía y sin restricciones. La estructura de esta proteína esta conservada con respecto a la estructura de las proteínas citolíticas.

Se establecieron las principales similitudes entre la estructura obtenida por el modelado por homología y las predicciones de las estructuras secundarias presentadas por el servidos SwissModel.

Se presentaron las principales características de la estructura terciaria de Cyt1Ab que definen su funcionalidad, actividad y mecanismo de acción.

6. BIBLIOGRAFÍA

1. Gutierrez, P., Alzate, O., & Orduz, S. (2001). A Theorical Model of the Tridimensional Structure of *Bacillus thurigiensis* subsp*. Medellín* Cry 11Bb Toxin Deduced by Homology Modelling. *Mem Inst Oswaldo Cruz* , 357-364.
2. Bravoa, A., Gillb, S. S., & Sobero, M. (2007). Mode of action of Bacillus thuringiensis Cry and Cyt toxins and. *Toxicon, ELSEVIER* , 423–435.
3. Wikipedia, t. f. (n.d.). *Homology modeling*. Retrieved Enero 12, 2010, from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Homology_modeling
4. Xu, Y., Xu, D., & Liang, J. (2007). *Computational methods for protein structure prediction and modeling* (Vol. II). New York: Springer.
5. Cohen S., Dym O., Albeck S., Ben-Dov E., Cahan R., Firer M. and Zaritsky A.(2008) *High-Resolution Crystal Structure of Activated Cyt2Ba Monomer from Bacillus thuringiensis subsp. Israelensis.* J. Mol. Biol. 380, 820–827