

# DISEÑO DE UN GEN SEMISINTÉTICO CRY1Ac Y ANÁLISIS DE LA ESTRUCTURA DE LA PROTEINA TRADUCIDA

## DESIGN OF A *CRY1Ac* SEMISYNTHETIC GENE AND STRUCTURE ANALYSIS OF TRANSLATED PROTEIN

## PROJETO DE UMSEMI-SINTÉTICOGENE *CRY1ACE* ANÁLISE DA ESTRUTURADA PROTEÍNA TRADUZIDA

CRISTINA DIAZ GRANADOS D.<sup>1</sup>, ANA MARÍA SANDOVAL<sup>2</sup>, ALEJANDRO CHAPARRO-GIRALDO<sup>3</sup>

### RESUMEN

*La población mundial está excediendo los 6 mil millones de personas, y es necesario que en los próximos años se incremente la producción de alimentos en un 50 %. Entre los factores limitantes de la producción agrícola están los insectos plaga. Se estima que las pérdidas causadas por diferentes plagas fluctúan entre 20% y 40%. Entre las estrategias que se han implementado para manejar este problema, está la utilización de plantas transgénicas resistentes a insectos, usando genes derivados de *Bacillus thuringiensis* (Bt). La mayoría de estos genes están protegidos por patentes, y por tanto el acceso está restringido. Se hace necesario buscar nuevas formas y herramientas que permitan el uso de estos genes, para la solución de problemas agrícolas. Se diseñó y construyó mediante paquetes bioinformáticos, una versión semisintética del gen *cry1Ac*, optimizando su expresión para arroz (*Oriza sativa*), mediante el criterio del uso codónico de arroz y se realizó un análisis *in silico* de las características de la proteína traducida.*

**Recibido para evaluación:** 30-08-2011. **Aprobado para publicación:** 18-07-2013.

- 1 M.Sc., Grupo de Ingeniería Genética de Plantas. Departamento de Biología & Instituto de Genética. Universidad Nacional de Colombia.
- 2 Bióloga, Grupo de Ingeniería Genética de Plantas. Departamento de Biología & Instituto de Genética. Universidad Nacional de Colombia.
- 3 Ph.D. Grupo de Ingeniería Genética de Plantas, Departamento de Biología & Instituto de Genética. Universidad Nacional de Colombia.

**Correspondencia:** ecdiazg@unal.edu.co ; achaparrog@una.edu.co

## ABSTRACT

*The world's population is exceeding 6 billion people and it's necessary in the coming years to increase food production by 50%. Among the limiting factors to agricultural production are insect pests. It is estimated that losses caused by different pests are ranging between 20% and 40%. Among the strategies that have been implemented to handle this problem is the use of transgenic plants resistant to insects, using genes derived from *Bacillus thuringiensis* (Bt). Most of these genes are protected by patents and therefore access is restricted. It is necessary to find new ways and tools to enable access to these genes, for the solution of agricultural problems. It was designed and built through bioinformatics packages, a semi synthetic version of the *cry1Ac* gene, codon optimized for the use of rice (*Oryza sativa*), and was performed in silico analysis of the characteristics of the translated protein.*

## RESUMO

*A população mundial é superior a 6 bilhões de pessoas, e é necessário nos próximos anos aumentar a produção de alimentos em 50%. Entre os fatores limitantes da produção agrícola são as pragas de insetos. Estima-se que as perdas causadas por pragas variam de 20% a 40%. Diversas estratégias têm sido implementadas para lidar com este problema, um dos quais é o uso de plantas transgênicas resistentes a insetos usando genes derivados de *Bacillus thuringiensis* (Bt). A maioria destes genes estão protegidos por patentes e, portanto, o acesso é restrito. É necessário encontrar novas formas e ferramentas para usar esses genes para a solução de problemas agrícolas. Neste trabalho foi projetado e construído através da bioinformática, uma versão semi-sintético do gene *cry1Ac* para otimizar a expressão para o arroz (*Oryza sativa*), pelo critério de uso códon e foi realizada in silico análise das características da proteína traduzida.*

## INTRODUCCIÓN

La población mundial está excediendo los 6 mil millones de personas, es necesario que en los próximos años se incremente la producción de alimentos en un 50 %, [1,2]. Entonces, se debe combatir eficazmente uno de los factores más limitantes, que son los insectos plaga que atacan los cultivos. Se estima que las pérdidas en la producción agrícola mundial causadas por plagas fluctúan entre 20% y 40%, con costos entre los 30 a 50 mil millones de dólares anuales [3,4].

La forma más usual para el control de insectos plaga de cultivos, es la aplicación de pesticidas químicos. El uso indiscriminado de insecticidas ha traído varias consecuencias: contaminación ambiental, resistencia en los insectos plaga, aumento en los costos de producción y problemas para la salud humana y animal [5, 2]. Una alternativa son las plantas transgénicas resistentes a insectos. La utilización de cultivos genéticamente modificados (GM) puede contribuir de manera significativa al manejo integrado de plagas [6,4].

## PALABRAS CLAVE:

Cultivos Transgénicos, Arroz, *cry1Ac*, Genes Semisintéticos, Bioinformática.

## KEY WORD:

Transgenic Crops, Rice, *cry1Ac*, Gen semi synthetic, Bioinformatics.

## PALAVRAS-CHAVE:

Cultivos Transgênicos, Arroz, *cry1Ac*, Gene semi-sintético, Bioinformática.

Diferentes versiones de genes derivados de *Bacillus thuringiensis* han sido transferidos a algodón, tomate, papa y maíz [7,8]. El gen Bt más usado es el gene *cry1Ac*, que fue transferido a 15 variedades de algodón, una de tomate y 6 de maíz. Se ha reportado que en estos cultivos controla *Heliothis virescens*, *Helicoverpa zea*, *Pectinophora gossypiella*, *Ostrinia nubilalis*, *Diatrea grandiosella* y *Pseudoplusia includens* [7,9,10]. En arroz, se han reportado varios trabajos experimentales usando el gen *cry1Ac* [11,12,13].

*Bacillus thuringiensis* es un bacilo flagelado, esporulado y gram + que produce durante la esporulación, un cristal de proteína tóxica para los insectos, conocido como delta endotoxina, que puede variar de forma y de tamaño [14,15]. Según el espectro de actividad insecticida que presentan, se clasifican en: Cry I tóxicos para lepidópteros, Cry II tóxicos para lepidópteros y dípteros, Cry III tóxicos para coleópteros, y Cry IV tóxicos para dípteros [16,17]. Si la proteína es ingerida por un insecto susceptible en su fase larvaria, llega al intestino medio, se disuelve por la acción de los jugos intestinales de pH alcalino, sufre una proteólisis enzimática, origina una toxina activa, que se une a receptores específicos en las membranas epiteliales del intestino, generando poros que desequilibran el balance osmótico y provocan lisis celular, causando muerte del insecto por inanición [15,18].

Los derechos de propiedad intelectual que se han concedido, en la forma de patentes sobre materia biológica [19], generan redes de patentes que limitan las posibilidades de uso tecnológico del conocimiento. Sobre el gene *cry1Ac* aparecen 16 reportes en la base de datos "Patentlens" usando la palabra clave "cry1Ac" en el "abstract", búsqueda realizada el 10 de agosto de 2011 (www.patentlens.net).

Se debe buscar maximizar la libertad de operación, sobre la base de que las patentes: 1) tienen aplicación en jurisdicciones nacionales; 2) las leyes varían de país a país; y 3) son un monopolio limitado y expiran [20].

Los genes Cry se encuentran cercados por una red de patentes, pero gracias a las herramientas de bioinformática, es posible el diseño de genes semi-sintéticos a partir de secuencias conocidas. Luego se produce una nueva versión optimizada por uso codónico, cuya proteína traducida se analiza in silico. Varias empresas ofrecen la síntesis de secuencias de DNA, mediante un contrato de servicios, en

el que se puede solicitar la exclusión de cualquier restricción de uso [21].

El presente trabajo trata sobre el diseño y construcción mediante paquetes bioinformáticos de un gen sintético Cry1Ac, su optimización por el uso codónico del arroz y el posterior análisis in silico de las características de la proteína traducida. El desarrollo de este gen semi-sintético está orientado a la obtención de plantas transgénicas de variedades colombianas de arroz que presenten resistencia al ataque de plagas de tipo lepidóptero.

## MÉTODO

### Desarrollo de la construcción semi-sintética del gen Cry 1Ac

Se hizo la revisión de literatura correspondiente y la búsqueda de secuencias, para genes que codifiquen la proteína cry1Ac, usando bases de datos de acceso público y/o disponible en el sistema nacional de bibliotecas de la Universidad Nacional de Colombia (SINAB) [22].

Una vez identificadas las secuencias de los genes, se realizó el proceso de diseño y construcción de la versión semi-sintética, empleando los siguientes paquetes: BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>); Lasergene (<http://www.dnastar.com/products-lasergene.aspx>); DNA 2.0 Gene designer (<https://www.dna20.com/genedesigner2/>); DNAsis SmartNote (<http://smartnote.miraibio.com/index.htm>).

Empleando estos software se realizó: 1) alineamientos de las secuencias nucleotídicas de los genes seleccionados; 2) establecimiento de una secuencia consenso; 3) selección y establecimiento de las secuencias de la región promotora y terminadora; 4) construcción del casete de expresión mediante el ensamble de las secuencias; 5) cambio de uso codónico para *Oryza sativa* en la región codificante; 6) ubicación de sitios de restricción dentro de la construcción y elaboración del mapa de restricción [1,22,23,24,25].

### Análisis de la proteína CryIGP

Para evaluar la predicción de las modificaciones post-traduccionales, se utilizó la herramienta SIGNALP (<http://www.cbs.dtu.dk/services/SignalP/>), que predice la presencia y localización del péptido señal y su punto de clivaje. En el análisis de la estructura prima-

ria de la proteína, se utilizaron distintas herramientas de EXPASY, para calcular el peso molecular y el punto isoelectrico COMPUTE pI/Mw (<http://expasy.org/tools/pitool.html>). Para calcular la hidrofobicidad, el porcentaje de residuos accesibles y la tendencia transmembranal, se usó PROTSCALE (<http://expasy.org/tools/protscale.html>). La predicción trans-membranal se realizó con los programas TMHMM (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) y TMPRED (<http://www.ch.embnet.org/software/TMPREDform.html>).

Los motivos y dominios conservados se analizaron con INTER PRO SCAN (<http://www.ebi.ac.uk/Tools/InterProScan/>). Se realizó un Blastx en la base de datos NCBI de Proteínas no redundantes para buscar secuencias con alto nivel de homología.

El análisis de la estructura terciaria, se realizó por "threading" utilizando el servidor I-tasser [26,27,28]. Al mejor modelo obtenido por "threading", se le realizó una evaluación de la estructura 3D, y una búsqueda de posibles errores, utilizando el algoritmo Rampage Ramachandran [29]. Utilizando el servidor DALI se realizó una comparación de estructuras tridimensionales de todas las proteínas contenidas en esta base de datos [30].

## RESULTADOS

### Desarrollo de la construcción semi-sintética del gen *Cry 1Ac*

La revisión bibliográfica se realizó en bases de datos de artículos disponibles vía electrónica de acceso público y en las 62 bases de datos disponibles para el personal de la Universidad Nacional de Colombia. Como resultado de estas búsquedas se encontraron 9 genes *cry1Ac* de interés, cuyas secuencias fueron empleadas para realizar los alineamientos y determinar una secuencia consenso. La obtención de las secuencias de los genes que resultaron de la revisión de literatura, se hizo en las bases de datos: GeneBank del Centro Nacional de Información en Biotecnología de Estados Unidos (NCBI) (<http://www.ncbi.nlm.nih.gov/>); EMBL-Bank del Laboratorio Europeo de Biología Molecular y el Instituto Europeo de bioinformática (EMBL-EBI) (<http://www.ebi.ac.uk/>), y DDBJ el banco de datos de ADN del Japón (<http://www.ddbj.nig.ac.jp/>) [25,22].

Identificadas y definidas las secuencias de los genes *Cry1Ac*, se realizaron los alineamientos de las se-

cuencias nucleotídicas de los genes seleccionados mediante el programa Clustal W del paquete bioinformático Lasergene: MegAlign, que permite comparar las secuencias entre sí, e identificar diferencias entre ellas, determinando el lugar y el nivel de conservación entre las secuencias alineadas [23,24,25,31].

Una vez alineadas y comparadas las secuencias se estableció la secuencia consenso mediante el paquete bioinformático Lasergene: SeqMan. La secuencia consenso porta la información de la región codificante del gen de interés, y sobre ella se realizaron posteriormente comparaciones con otras secuencias reportadas en las bases de datos, mediante la herramienta BLAST del software DNAsis y de la plataforma NCBI. Esto se hizo para asegurar que no existieran diferencias significativas entre la secuencia consenso obtenida y otras secuencias reportadas para el gen *Cry1Ac*. Se empleó un bit score alto y un e-value menor al nivel de significancia empleado que fue de 0,05 [31,32].

Para el desarrollo de la construcción del casete de expresión se decidió emplear el promotor de la Ubiquitina de maíz y el terminador de la Nopalina sintetasa de *Agrobacterium* [33]. Las secuencias nucleotídicas del promotor y el terminador se buscaron en las bases de datos: GeneBank, EMBL-Bank y DDBJ [24,34].

Determinadas las secuencias tanto del gen de interés como del promotor y el terminador se procedió a la construcción del casete de expresión, mediante el ensamble de las secuencias, promotor de la Ubiquitina de maíz: secuencia consenso *cry1Ac*: terminador de la Nopalina sintetasa de *Agrobacterium*, obteniéndose así la secuencia del gen semi-sintético *Cry1Ac/GP* [24,25,32]. Esto se hizo con el programa Lasergene: SeqBuilder.

Ensamblada la secuencia, se procedió con el cambio de uso codónico para *Oryza sativa* en la región codificante, para optimizar la expresión del gen, en el genoma arroz [23,35,36]. Se recomienda la estrategia de optimización del uso codónico en el diseño de los genes semisintéticos [37,38]. Se descargó el uso codónico para *Oryza sativa* de la base de datos CODON USAGE (<http://www.kazusa.or.jp/codon/>) y se transfirió al programa DNA 2.0 Gene designer. Allí se ejecutó el cambio de uso codónico de la secuencia consenso del gen [1,32]. Luego, se comparó la nueva secuencia optimizada con la secuencia consenso tanto a nivel de nucleótidos como de aminoácidos usando la herramienta Clustal W del

programa BioEdit. Esta versión también fue alineada contra secuencias de bases de datos de genes y proteínas, mediante las diferentes formas de BLAST existentes para comparar nucleótidos y aminoácidos, Blastn y Blastx del programa BioEdit [25].

Finalmente para terminar con la construcción del gen semisintético, denominado *Cry1AcIGP* a partir de aquí, se realizó un mapa de restricción sobre la secuencia completa empleando el programa BioEdit. Una vez se conoció este mapa se escogieron los sitios de restricción más apropiados para ser introducidos dentro de la construcción, con el fin de facilitar su manipulación [31,32]. Ello se efectuó mediante el programa DNA 2.0 Gene designer.

Una vez introducida esta modificación final a la secuencia se obtuvo la construcción definitiva del gen semisintético *Cry1AcIGP* (Figura 1).

### Análisis de la proteína CryIGP

Según la predicción para péptido señal, basada en “redes neuronales”, hay valores suficientemente altos de score, como para considerar la presencia de péptido señal. En la predicción hecha con los modelos escondidos de Markov, se determinó que la probabilidad de encontrar péptido señal en la proteína CryIGP es de 0.39, y que el punto de clivaje estaría en el aminoácido número 24. Adicionalmente la herramienta compute pI/Mw permitió el cálculo teórico del punto isoelectrónico y el peso molecular de la proteína CryIGP los cuales son: pI/Mw: 5,47 / 72985,07 [39,40,41]. Con base en estos datos se puede afirmar que la proteína CryIGP, posee péptido señal.

En el perfil de hidrofobicidad los picos más altos muestran las zonas de la proteína que tienden a ser hidrofóbicos. En la grafica resultante (datos no mostrados), no se muestra una tendencia de la pro-

teína a estar en la membrana celular, ya que presenta tanto picos negativos como positivos.

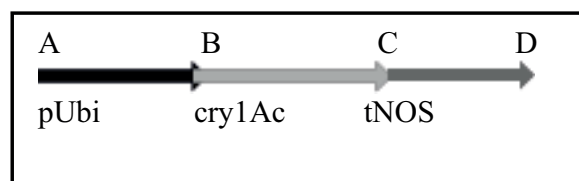
La predicción trans-membranal realizada con TMHMM es más precisa para el análisis, ya que esta no muestra la tendencia de los aminoácidos, sino que arroja la probabilidad de encontrar la proteína, dentro, entre o fuera de la superficie celular. Según la grafica lo más probable es que la proteína sea extracelular, además que la probabilidad de encontrar hélices trans-membranales es de 0,4, es la predicción basada en las regiones de la proteína que abarcan la membrana y su orientación en esta misma [42]. Es posible afirmar de acuerdo con estos resultados que el ánimo terminal se encuentre al exterior de la membrana y que en la secuencia se hallan 3 hélices trans-membranales.

El análisis de la estructura secundaria de la proteína CryIGP, arrojó un porcentaje de 36,09% para un alfa hélice en la posición 262; un porcentaje de 20,39% para una cadena extendida en la posición 148; una lamina beta con un porcentaje de 8,40% en la posición 61 y por ultimo “random coil” en la posición 255 con un porcentaje de 35,12%. Las hélices hidrofóbicas y anfipáticas sugieren que este dominio puede ser responsable para la formación de poros en el epitelio intestinal del organismo susceptible [43]. Adicionalmente al comparar los sitios conservados de la proteína CryIGP, con los reportados para otras proteínas Cry, esta presenta varios dominios para endotoxinas, en especial un dominio de unión a la galactosa.

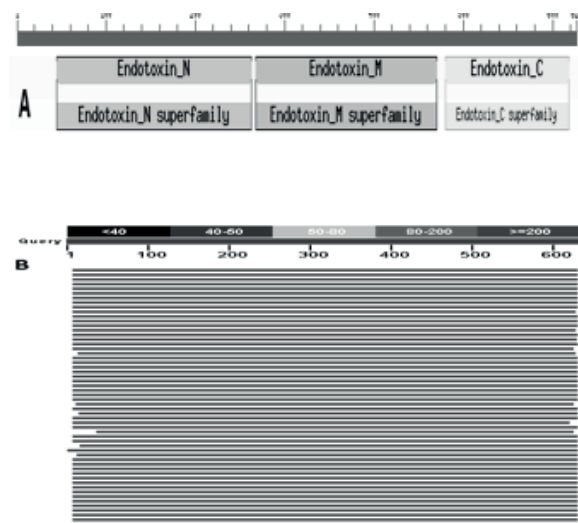
En el alineamiento múltiple de proteínas, se encontraron varias proteínas Cry con un “E-valué” del 0,0 % (Figura 2), lo que indica que no se presenta ninguna diferencia en comparación con la proteína CryIGP.

El cuadro 1 muestra los mejores resultados del alineamiento múltiple de la proteína. Blastx es un algoritmo que permite comparar una secuencia de nucleótidos con la base de datos para secuencias de aminoácidos para buscar homologías [46]. La evaluación de estos alineamientos se realizó con la matriz de sustitución blosum62. Esta matriz evalúa los alineamientos y calcula el bitscore y el e-value de cada alineamiento. Los alineamientos con un valor de bitscore alto y e-value pequeño, representan homologías significativas estadísticamente, confirmando la veracidad del alineamiento. Todos los alineamientos presentes en el cuadro 1 tienen un e-value igual a cero. Esto indica que la probabilidad de encontrar dicho alineamiento al azar es

Figura 1. Esquema del gen sintético *Cry1AcIGP*



Sitios de corte de restrictasas: A (EcoRV); B (NotI); C (AvrII); D (Hind III). Tamaño total 4318 pb.

**Figura 2.** Blast de proteínas. Alineamiento múltiple de secuencias

A. Dominios conservados [44,45,46,47]. B. alineamiento de secuencias de proteínas contenidas en la NCBI.

cero y que estos alineamientos reflejan verdaderas homologías.

En el modelamiento 3-D por “threading”, se obtuvieron cinco resultados posibles, el que presentó mayor score fue de 0,47 (datos no mostrados), mientras que los sitios de unión se establecieron en los aminoácidos GLN:160 ASN:161 TYR:162 GLN:163 VAL:164 ARG:218 TRP:219 ASN:221 THR:222 GLY:223 LEU:224 GLU:225 ARG:226 TRP:228 [26,27,28].

**Cuadro 1.** Alineamiento BLASTX de la proteína.

Accesion <sup>1</sup>	Name <sup>2</sup>	Bitscore <sup>3</sup>	E-value <sup>4</sup>
ADO64600.1	Cry1Ab/c [synthetic construct]	1247	0,0
ADO64599.1	Cry1Ac [synthetic construct]	1247	0,0
CAA70925.1	delta-endotoxin [Bacillus thuringiensis]	1247	0,0
BAH03128.1	insecticidal crystal protein [synthetic construct]	1236	0,0
ADV57667.1	Cry1Ac [Bacillus thuringiensis]	1235	0,0
AAU87037.1	Cry1Ac [Bacillus thuringiensis]	1235	0,0
AAM95162.1	insecticidal protein CryIAc1 [synthetic construct]	1235	0,0
ZP_04118188.1	Pesticidal crystal protein cry1Ae [Bacillus thuringiensis serovar kurstaki str. T03a001]	1235	0,0
AAA73077.1	Insecticidal delta endotoxin [Bacillus thuringiensis serovar kurstaki]	1233	0,0
ACM90319.1	Cry1Ac [Bacillus thuringiensis]	1232	0,0

1. Número de acceso en la base de datos del NCBI. 2. Nombre de la secuencia alineada. 3. Valor del bit score para las diferentes alineaciones. 4. E-value para los diferentes alineamientos. En todos los casos e-value = 0.0 mostrando robustez estadística en los alineamientos.

Para evaluar la calidad de la estructura 3D modelada, se realizó una evaluación de la posición espacial de los residuos que forman la proteína, mediante el servidor Rampage Ramachandran. Este servidor se basa en los ángulos de torsión para cada residuo en la estructura de la proteína. Los enlaces peptídicos tienden a ser planos con dos organizaciones espaciales permitidas, trans = 180° y cis = 0°. Según la gráfica de Ramachandran, la distribución de más del 98% de los aminoácidos se encuentra en regiones favorables, esto es un indicio de la calidad estereoquímica del modelo analizado [29].

Matemáticamente, el gráfico de Ramachandran es la visualización de una función. La glicina tiene un átomo de hidrógeno, con un pequeño radio de van der Waals, en lugar de un grupo metilo en la posición β. Por lo tanto, es menos restringido y esto se puede comprobar en el gráfico de Ramachandran de glicina para que el área permitida sea considerablemente mayor. En contraste, el gráfico de Ramachandran de prolina muestra sólo un número muy limitado de combinaciones posibles de ψ y φ.

A pesar del cambio en la estructura de las proteínas como resultado de la evolución, existen patrones de organización en los residuos de las proteínas, entonces la estructura de las proteínas puede ser identificada mediante la detección de estos patrones conservados. La matriz de alineamiento a distancias DALI, rompe la estructura en hexapeptidos y calcula una ma-

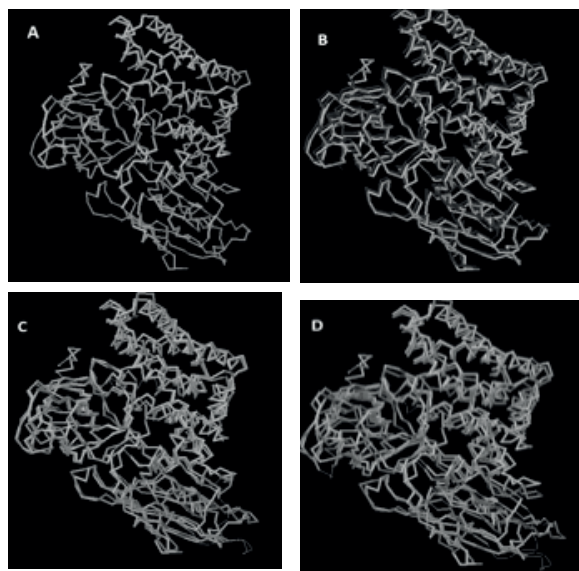
triz de distancia evaluando los patrones de contacto entre fragmentos sucesivos [48,49].

En la figura 2 se muestra 3 alineamientos a distancia usando la matriz DALI, donde la estructura 3D de la proteína CryIGP fue comparada con 3 proteínas Cry (CryIA, CryIIIA y Cry3bb). Se encontró que las proteínas comparten las mismas características o tienen similares características en aproximadamente las mismas posiciones, como también puede decirse que tienen similares plegamientos con bucles de longitud similar conectando sus elementos de estructura secundaria (Figura 3). EscucharLeer fonéticamente

## CONCLUSIONES

Se obtuvo una versión semisintética del gene *Cry1Ac*, con modificaciones en su secuencia, resultantes tanto de la obtención de la secuencia consenso como de la optimización por uso codónico del arroz. Esta versión, aquí denominada *Cry1AcIGP*, difiere mínimo un 12% con respecto a secuencia reportadas anteriormente.

Figura 3. Alineamiento a Distancia DALI



A. proteína CryIGP; B. proteína cryIA; C, cryIIIA delta endotoxina; D, PesticidalCrystalProtein Cry3bb. Usando la base de datos para proteínas PDB, se realizó un alineamiento múltiple estructural, para hallar homología entre la estructura 3D de la proteína CryIGP y las proteínas presentes en la base de datos. Los resultados muestran homología en los sitios de acción y en los motivos como también zonas variables con bucles indefinidos.

El modelamiento molecular y bioinformático permite no solamente predecir la estructura tridimensional de la proteína CryIGP, sino también definir su función. Los dominios y sitios conservados indican su naturaleza de endotoxina con propiedades insecticidas y a su vez su gran parecido con proteínas producidas por la bacteria gram negativa *Bacillus thuringiensis*.

Con el análisis de la estructura primaria, se confirmó que esta es una proteína extracelular, la cual actúa como toxina en diferentes grupos de insectos. El análisis de la estructura secundaria, permitió identificar los dominios conservados, los cuales permiten predecir su función y papel como toxina. Los análisis de alineamiento múltiple, confirmaron la relación de la proteína CryIGP, con otras endotoxinas reportadas, como también una muestra significativa de la diversidad de esta proteína en la naturaleza.

Es clave realizar el estudio detallado, mediante el uso de herramientas bioinformáticas, de la estructura de las proteínas resultantes de genes semisintéticos, lo cual permitirá determinar de una forma presuntiva el buen diseño y expresión del gen.

## AGRADECIMIENTOS

Al Biólogo Alejandro Rodríguez por su valiosa asesoría y colaboración para el desarrollo de este trabajo. Al convenio firmado por FEDEARROZ y la Universidad Nacional de Colombia para el desarrollo de sistema de ingeniería genética para cultivares colombianos de arroz, que financiaron este trabajo.

## REFERENCIAS

- [1] WU, J. and BALDWIN, I. New Insights into Plant Responses to the Attack from Insect Herbivores. *Annu. Rev. Genet.*, 44, 2010, p.1–24.
- [2] KOS, M., LOON, J., DICKE, M. and VET, L. Transgenic plants as vital components of integrated pest management. *Trends in Biotechnology*, 27(11), 2009, p. 621-627.
- [3] STRANGE, R. and SCOTT, P. Plant disease: A Threat to Global Food Security. *Annu. Rev. Phytopathol.*, 43, 2005, p. 83–116.
- [4] XIA, H., CHEN, L., WANG, F. And LUA, B. Yield benefit and underlying cost of insect-resistance transgenic rice: Implication in breeding and

- deploying transgenic crops. *Field Crops Research*, 118, 2010, p. 215–220.
- [5] FERRY, N., EDWARDS, M., GATEHOUSE, J., CAPELL, T., CHRISTOU, P. and GATEHOUSE, A. Transgenic plants for insect pest control: a forward looking scientific perspective. *Transgenic Research*, 15, 2006, p. 13–19.
- [6] MEHLO, L., GAHAKWA, D., TRUNGNGHIA, P., THILOC, N., CAPELL, T., GATEHOUSE, J., GATEHOUSE, A. and CHRISTOU, P. An alternative strategy for sustainable pest resistance in genetically enhanced crops. *PNAS.*, 102 (22), 2005, p. 7812–7816.
- [7] CERA. GM Crop Database. Center for Environmental Risk Assessment [CERA], ILSI Research Foundation, Washington D.C. [online]. Available: [http://cera-gmc.org/index.php?action=gm\\_crop\\_database](http://cera-gmc.org/index.php?action=gm_crop_database) [citado 30 de marzo de 2011].
- [8] GATEHOUSE, J. Biotechnological Prospects for Engineering Insect-Resistant Plants. *Plant Physiology*, 146, 2008, p. 881–887.
- [9] GURR, S. and RUSHTON, P. Engineering plants with increased disease resistance: what are we going to express?. *Trends in Biotechnology*, 23 (6), 2005, p. 275-282.
- [10] KHANNA, H.K., and RAINA, S.K. Elite Indica Transgenic Rice Plants Expressing Modified Cry1Ac Endotoxin of *Bacillus Thuringiensis* Show Enhanced Resistance to Yellow Stem Borer [*Scirpophaga Incertulas*]. *Transgenic Research*, 11, 2002, p. 411-423.
- [11] KIM, H.E., SUH, S.C., PARK, B.S., SHIN, K.S., KWEON, J.K., HAN, E.J., PARK, S-H, KIM, Y.S. and KIM, J-K. Chloroplast-targeted expression of synthetic *cry1Ac* in transgenic rice as an alternative strategy for increased pest protection. *Planta*, 2009, p. 397-405.
- [12] HAN, L., WU, K., PENG, Y., WANG, F. and GUO, Y. Efficacy of transgenic rice expressing Cry1Ac and CpTI against the rice leafhopper, *Cnaphalocrocismedinalis* [Guenée]. *Journal of Invertebrate Pathology*, 96, 2007, p. 71-79.
- [13] AHMAD, A., MAQBOOL, S.B., RIAZUDDIN, S., MARIAM B. and STICKLEN, M.B. Expression of Synthetic Cry1Ab and Cry1Ac Genes in Basmati Rice [*Oryza sativa* L.] Variety 370 via Agrobacterium-Mediated Transformation for the Control of the European Corn Borer [*Ostrinianubilalis*]. *In Vitro Cellular & Developmental Biology*, 38, 2002, p. 213-220.
- [14] SAUKA, D. and BENINTENDE, G. *Bacillus thuringiensis*: generalidades. Un acercamiento a su empleo en el biocontrol de insectos lepidópteros que son plagas agrícolas. *Revista Argentina de Microbiología*, 40, 2008, p. 124-140.
- [15] PACHECO, S., GOMEZ, I., GILL, S., BRAVO, A. and SOBERON, M. Enhancement of insecticidal activity of *Bacillus thuringiensis* Cry1A toxins by fragments of a toxin-binding cadherin correlates with oligomer formation. *Peptides*, 30, 2009, p. 583-588.
- [16] BABU, R., SAJEENAB, A., SEETHARAMANB, K. and REDDY, M. Advances in genetically engineered [transgenic] plants in pest management—an over view. *Crop Protection*, 22, 2003, p. 1071–1086.
- [17] WANG, Y., ZHANG, G., DU, J., LIU, B. and WANG, M. Influence of transgenic hybrid rice expressing a fused gene derived from cry1Ab and cry1Ac on primary insect pests and rice yield. *Crop Protection*, 29, 2010, p. 128–133.
- [18] CHRISTOU, P., CAPELL, T., KOHLI, A., GATEHOUSE, J. and GATEHOUSE, A. Recent developments and future prospects in insect pest control in transgenic crops. *Trends in Plant Science*, 11(6), 2006, p. 302-308.
- [19] LAKSHMIKURAMAN, M. and PHILLIPS, P. Patenting of biotechnological innovations. *Asian Biotechnology and Development Review*, 7 (2), 2005, p. 25-44.
- [20] BIOS. What does “Freedom to Operate” mean? [online]. Available: <http://www.patentlens.net/daisy/patentlens/2768.html> [citado 30 de marzo de 2011].
- [21] RANDHAWA, G., SINGH, M. and GROVER, M. Bioinformatic analysis for allergenicity assessment of *Bacillus thuringiensis* Cry proteins expressed in insect-resistant food crops. *Food and Chemical Toxicology*, 49, 2011, p. 356–362.
- [22] WU, G., ZHENG, Y., QURESHI, I., ZIN, H., BECK, T, BULKA, B. and FREELAND, S. SGDB: a database of synthetic genes re-designed for optimizing protein over-expression. *Nucleic Acids Research*, 35, 2007, p. 76-79.
- [23] RICHARDSON, S., WHEELAN, S., YARRINGTON, R. and BOEKE, J. Gene Design: Rapid, automated design of multikilobase synthetic genes. *Genome Res*, 16, 2006, p. 550-556.
- [24] CZAR, M., ANDERSON, J., BADER, J. and PEC-COUD, J. Gene synthesis demystified. *Trends in Biotechnology*, 27(2), 2008, p. 63-72.



- [25] RICHARDSON, S., NUNLEY, P., YARRINGTON, R., BOEKE, J. and BADER, J. Gene Design 3.0 is an updated synthetic biology toolkit. *Nucleic Acids Research*, 38 (8), 2010, p. 2603–2606.
- [26] AMBRISH, R., ALPER, K. and YANG, Z. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5, 2010, p. 725-738.
- [27] YANG, Z. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins.*, 9, 2009, p. 100-113.
- [28] YANG, Z. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 2010, p. 40.
- [29] LOVELL, S., DAVIS, I., ARENDALL, W., BAKKER, P., WORD, J., PRISANT, M., RICHARDSON, J., and RICHARDSON, D. Structure validation by  $\alpha$  geometry:  $\phi$ ,  $\psi$  and  $\omega$  deviation. *Proteins: Structure, Function & Genetics*, 50, 2002, p. 437-450.
- [30] DIETMANN, S., FERNANDEZ-FUENTES, N. and HOLM, L. Automated detection of remote homology. *Current Opinion in Structural Biology*, 12, 2002, p. 362-367.
- [31] JAYARAJ, S., REID, R. and SANTI, D. GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Research*, 33 (9), 2005, p. 3011–3016.
- [32] VILLALOBOS, A., NESS, J., GUSTAFSSON, C., MINSHULL, J. and GOVINDARAJAN, S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics*, 7 (285), 2006, p. 1-8.
- [33] PEREMARTI, A., TWYMAN, R., GOMEZ-GALERA, S., NAQV, I.S., FARRE, G., SABALZA, M., MIRALPEIX, B., DASHEVSKAYA, S., YUAN, D., RAMESAR, K., CHRISTOU, P., ZHU, C., BASSIE, L. and CAPELL, T. Promoter diversity in multigene transformation. *Plant Mol Biol.*, 73, 2010, p. 363–378.
- [34] BHALERAO, K. Synthetic gene networks: the next wave in biotechnology?. *Trends in Biotechnology*, 27 (6), 2009, p. 368-374.
- [35] CAI, Y., SUN, J., WANG, J., DING, Y., TIAN, N., LIAO, X. and XU, W. Optimizing the codon usage of synthetic gene with QPSO algorithm. *Journal of Theoretical Biology*, 254, 2008, p. 123– 127.
- [36] KUMARA, S., BIRAHB, A., CHAUDHARYA, B., BURMAA, P., GUPTAB, G. and PENTALA, D. Plant codon optimized cry genes of *Bacillus thuringiensis* can be expressed as soluble proteins in *Escherichia coli* BL21 Codon Plus strain as NusA-Cry protein fusions. *Journal of Invertebrate Pathology*, 88, 2005, p. 83–86.
- [37] BURGESS-BROWN, N., SHARMA, S., SOBOTT, F., LOENARZ, C., OPPERMAN, U. and GILEADI, O. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression and Purification*, 59, 2008, p. 94–102.
- [38] RAYMOND, A., LOVELL, S., LORIMER, D., WALCHLI, J., MIXON, M., WALLACE, E., THOMPSON, K., ARCHER, K., BURGIN, A. and STEWART, L. Combined protein construct and synthetic gene engineering for heterologous protein expression and crystallization using Gene Composer. *BMC Biotechnology*, 9 (37), 2009, p. 1-15.
- [39] BJELLQVIST, B., HUGHES, G., PASQUALI, G., PAQUET, C., RAVIER, N., SANCHEZ, F., FRUTIGER, J. and HOCHSTRASSER, D. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14, 1993, p. 1023-1031.
- [40] BJELLQVIST, B., BASSE, B., OLSEN, E. and CELIS, J. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*, 15, 1994, p. 529-539.
- [41] GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M., APPEL, R. and BAIRROCH, A. En: Protein Identification and Analysis Tools on the ExpASY Server. *The Proteomics Protocols Handbook*. Ed. Hertfordshire (England): John M. Walker, Humana Press, 2005, p.571- 607
- [42] HOFMANN, K. and STOFFEL, W. TMbase – A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, 374, 1993, p.166.
- [43] SCHNEPF, E., CRICKMORE, N., VAN RIE, J., LE-RECLUS, D., BAUM, J., FEITELSON, P., ZEIGLER, D. and DEAN, D. *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. *Microbiology and Molecular Biology Reviews*, 62, 1998, p. 775-806.
- [44] MARCHLER-BAUER, A. and BRYANT, S. "CD-Search: protein domain annotations on the fly." *Nucleic Acids Res.*, 32, 2004, p.327-331.
- [45] MARCHLER-BAUER, A. "CDD: specific functional annotation with the Conserved Domain Database." *Nucleic Acids Res.*, 37, 2009, p.205-210.
- [46] STEPHEN, F., ALTSCHUL, C., THOMAS, L., MADDEN, ALEJANDRO, A., SCHÄFFER, ZHANG, J., ZHANG, Z., MILLER, W. and LIPMAN, D. "Gapped BLAST and PSI-BLAST: a new generation

- of protein database search programs". *Nucleic Acids Res.*, 25, 1997, p. 3389-3402.
- [47] STEPHEN, F., ALTSCHUL, C., WOOTTON, E., GERTZ, M., AGARWALA, R., MORGULIS, A., SCHÄFFER, A. and YI-KUO, Y. "Protein database searches using compositionally adjusted substitution matrices". *FEBS J.*, 272, 2005, p. 5101-510
- [48] HOLM, L. and SANDER, C. Mapping the protein universe. *Science*, 273, 1996, p. 595-603.
- [49] WU, G., BASHIR-BELLO, N. and FREELAND, S. The Synthetic Gene Designer: A flexible web platform to explore sequence manipulation for heterologous expression. *Protein Expression and Purification*, 47, 2006, p. 441–445.