

ANÁLISIS BIOINFORMÁTICO DE LAS SECUENCIAS DE AMINOÁCIDOS DE DIVERSAS FOSFOLIPASAS D

Christian Solís Calero¹

RESUMEN

La Fosfolipasa D es una enzima presente en todos los organismos eucariotas, su papel en los mecanismos de señalización intracelular ha sido bastante estudiado en plantas, no obstante el rol que cumplen en mamíferos todavía no ha sido destacado respecto a otros moduladores de la señalización intracelular. Por ello se analizaron insilicó las secuencias de aminoácidos de las isoformas humanas D1 y D2 de la Fosfolipasa D, obteniendo la predicción de sus propiedades químicas responsables de sus diferentes niveles de acción. Usando el programa MACAW realizamos alineamientos múltiples locales de secuencias obtenidas de la base de datos NR del NCBI correspondientes a la "Fosfolipasa D" de varias especies encontrando sólo un segmento conservado entre las secuencias de origen bacteriano, y ocho entre las secuencias de origen vegetal. Se detectó además tres segmentos de secuencia de aminoácidos comunes en secuencias de especies de distante origen filogenético, la visualización de estos segmentos en la estructura 3D de la fosfolipasa D de la Bacteria *Streptomyces Sp* nos reveló que los residuos de aminoácidos altamente conservados de estos segmentos se localizan cercanamente formando una zona en la proteína que probablemente contiene los residuos esenciales para su actividad.

Palabras Clave: Fosfolipasa D, Bioinformática, bases de datos biológicas, alineamiento múltiple.

ABSTRACT

Phospholipases D are enzymes present in all eucariotic organisms, it was first discovered in plants and in recent years, increasing attention has been paid to its roles in signal transduction in mammals. Human isoforms of this enzyme D1 and D2 were analyzed using bioinformatic tools, predicting their chemical properties responsible of their different levels of action. Besides, using local multiply alignment programs like MACAW with amino acid sequences from different organisms, three conserved blocks of sequences in all organisms were detected. Between Plant sequences and not with bacterial, high similarity was detected. When these conserved blocks of sequence was visualized in determined experimentally 3D structure of Phospholipase D from *Streptomyces Sp*, an structural zone having near these blocks of sequence, was observed. This zone is postulated like an important structural feature of these enzymes that possibly contain essential amino acid residues for their biological activity.

Key Words: Phospholipase D, bioinformatics, biological databases, multiply alignment.

INTRODUCCIÓN

La cantidad de datos de origen biológico tales como secuencias, estructuras, expresión de genes es actualmente tan grande que el análisis manual es imposible, la única forma de analizarlos para obtener información útil es mediante el uso de programas computacionales. La Bioinformática es el área multidisciplinaria que organiza y hace posible la utilización de estos datos, con el objetivo de comprender los procesos celulares y predecir los fenómenos biológicos.

La fosfolipasa D es una enzima que cataliza la producción de ácido fosfatídico, el que actúa a su vez como segundo mensajero en una cascada de señalización intracelular que determina la realización de diferentes eventos celulares (1). La investigación del rol Biológico que cumple se ha dado preferentemente en plantas, donde participa durante la germinación de semillas, la senescencia y

bajo condiciones de Stress (2). Es recién en los últimos años que el papel de esta enzima se ha estudiado con mayor detalle en otros organismos eucariotas. De allí nuestro interés en estudiarla, tomándola como modelo para un estudio bioinformático.

MATERIAL Y MÉTODOS

1. Búsquedas en bases de datos de secuencias: Obtuvimos 63 registros en formato FASTA de secuencias de aminoácidos correspondientes a diferentes Fosfolipasas D de la base de datos NR (no redundante) GenBank del NCBI (National Center for Biotechnology Information), localizado en la siguiente dirección: <http://www.ncbi.nlm.nih.gov/>. 32 secuencias corresponden a especies vegetales, 11 animales y 20 bacterianas. El código de acceso de las Fosfolipasas D1 (PLD1) y D2 (PLD2) de humanos es AAB49031 y AAB96655 respectivamente.

2. Análisis de las Fosfolipasa D de origen Humano: Usando sus secuencias de aminoácidos, se predijo algunas características de las PLD1 y PLD2 gracias a herramientas informáticas como NETNGLYC 2.0, NETPHOS 2.0, YINYOANG 1.2, PROTFUN 2.0, Y NETCHOP 2.0, localizadas en el servidor <http://www.cbs.dtu.dk/services/>. Las regiones asociadas a Membrana. Fueron predichas con el programa DAS (<http://www.sbc.su.se/~miklos/DAS/maindas.html>).

3. Alineamiento Múltiple local de Secuencias: Se realizaron los alineamientos usando las secuencias obtenidas de la base de datos NR del NCBI con el programa MACAW de acuerdo al muestreo de Gibbs (Gibbs Sampler), basados en los algoritmos de Lawrence y Col (3).

4. Predicción de segmentos de secuencia conservados: Se usó el programa MACAW, se obtuvo alineamientos múltiples locales usando se-

¹ Universidad Nacional Mayor de San Marcos. Facultad de Farmacia y Bioquímica. biofara@universia.edu.pe Lima, Perú

Tabla 1. Propiedades de las Fosfolipasas D de humanos predecidas a partir de servidores de distintos Programas en Internet

| PREDICCIÓN | PROGRAMA | METODO | RESULTADOS | |
|--|--------------|----------------------|--|---|
| | | | FOSFOLIPASA D 1 | FOSFOLIPASA D 2 |
| Regiones asociadas a Membrana (L>10 aa) | DAS | Métodos estadísticos | 0 | 3 |
| Sitios de fosforilación | NETPHOS 2.0 | Redes Neuronales | Sitios de Ser = 34 Sitios de Thr = 13 Sitios de Tyr = 10 | Sitios de Ser = 28 Sitios de Thr = 18 Sitios de Tyr = 8 |
| Sitios de N glicosilación | NETNGLYC 1.0 | Redes Neuronales | 4 | 1 |
| Aceptores de glicosilación O-(b)-GlcNAc intracelular | YINOYANG 1.2 | Redes Neuronales | 5 | 7 |
| Sitios de Clivaje por el Proteosoma | NETCHOP 2.0 | Redes Neuronales | Más de 30 sitios de Clivaje | Más de 30 sitios de Clivaje |

aa: aminoácidos P: probabilidad, L: longitud

cuencias de especies filogenéticamente distantes, y a partir de esa información se determino segmentos de secuencia conservados que luego se visualizaron en la estructura 3D determinada experimentalmente por Difracción de Rayos X de la fosfolipasa D de la Bacteria *Streptomyces Sp* (4), usando el programa Swiss pdb Viewer V3.7 b2 Obtenido de la siguiente dirección:

<http://www.expasy.ch/spdbv/mainpage.html>.

5. Predicción de las estructuras 3D de las Fosfolipasas D humanas:

Se realizo por el método de reconocimiento de plegamiento "threading" a través del servidor 3D-PSSM (<http://www.sbg.bio.ic.ac.uk/~3dpssm/>).

RESULTADOS Y DISCUSION

1. Análisis de las Fosfolipasas D De origen Humano:

Las diferencias observadas entre las PLD1 y PLD2 de humanos luego del análisis computacional (Tabla 1), pueden tener relación con las diferentes vías de regulación que poseen (1, 10), se ha observado que la PLD1 se expresa preponderantemente en tejidos en desarrollo, en cambio la PLD2 se expresa en casi todo tipo de tejido estudiado (5). La localización de sitios de fosforilación es importante ya que se ha demostrado que ambas isoformas pueden activarse por fosforilación (6). El análisis de la presencia de regiones hidrofóbicas se hacía necesario ya que se trata de un grupo de enzimas que interacciona con la membrana ya que su sustrato un fosfolípido, la fosfatidilcolina es parte de ella. Nosotros encontramos que PLD2 presenta estas regiones y en la PLD1 están ausentes, lo que sugiere un diferente grado de actividad por la diferente accesibilidad al sustrato en común, la fosfatidilcolina. El análisis de sitios de corte por parte del complejo enzimático Proteosoma, es im-

portante por que nos permite inferir el tiempo de vida de la enzima. El complejo proteosoma constantemente hidroliza proteínas intracelulares como parte del recambio molecular normal de la célula. Se encontraron numerosos sitios para clivaje, ello nos permite inferir un tiempo de vida corta para estas proteínas. No obstante, al encontrarse sitios de O-glicosilación y N-glicosilación también podríamos señalar que estas cadenas de oligosacarido podrían servir para proteger a las fosfolipasas D de la acción del complejo proteosómico extendiendo su vida.

2. Alineamiento Múltiple local de Secuencias:

Se encontró sólo un segmento conservado entre las secuencias de origen bacteriano, y ocho entre las secuencias de origen vegetal. En el caso de las Fosfolipasas D de origen Bacteriano este bajo número de segmentos de secuencia conservados, probablemente se deba

porque esta enzima no esta sometida a una alta presión evolutiva, de ello que fácilmente el gen que lo codifica mute, sin perder con ello alguna ventaja selectiva. En las plantas la presión evolutiva es alta debido a que la enzima cumple preponderantes roles en la señalización celular.

3. Predicción de sitios importantes para la Función de la Proteína:

Cuando se realizo el análisis del alineamiento de las secuencias de especies representativas de diferentes reinos se encontró sólo 3 segmentos de secuencia comunes. La visualización de los segmentos conservados en la estructura 3D de la fosfolipasa D de la Bacteria *Streptomyces Sp* nos revela que los residuos de aminoácidos altamente conservados de estos bloques se localizan muy cercanamente formando una zona característica en la proteína (Fig 1-A). Entre los residuos altamente conservados (Fig. 1-B) tenemos aminoácidos aro-

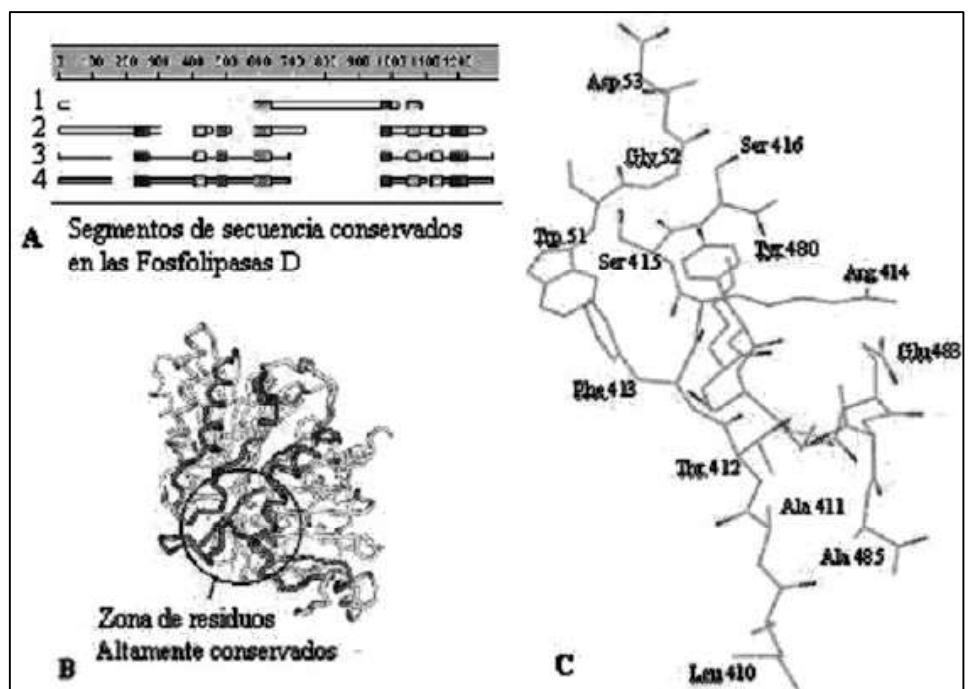


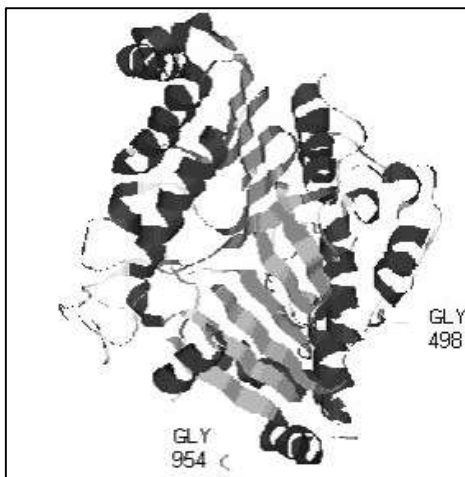
Figura 1. Resultado de Programa Macaw: A. Segmentos conservados (1: *Streptomyces Sp*, 2: *Arabidopsis thaliana*, 3: *Homo Sapiens PD1*, 4: *PD2 Homo sapiens*), B. Estructura 3D con los segmentos de secuencia conservados (cadenas oscuras). C. Aminoácidos altamente conservados

máticos: Trp51, Phe413 y Tyr480, que son característicos del interior o "core" de la proteína. Los aminoácidos Arg414 y Glu483 se observan localizados en forma opuesta, formando un puente salino característico también de los "cores" de las proteínas. En cambio residuos como Ser416 y Ser415 son candidatos para ser parte del sitio activo de la enzima (7).

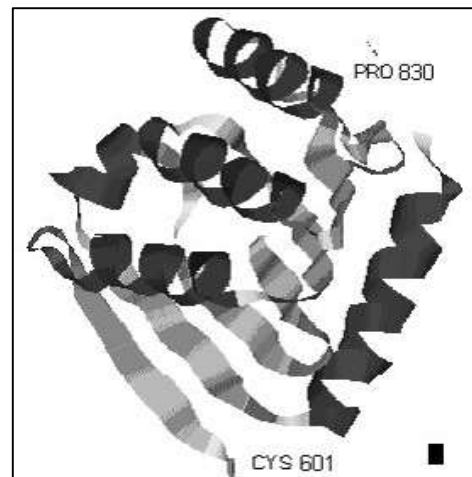
4. Predicción de las estructuras 3D de las Fosfolipasas D humanas:

Las PD1 y PD2 no presentan estructuras 3D determinadas experimentalmente por lo que obtuvimos por el método de "threading" estructuras 3D que solo se ajustaban a una secuencia parcial de la proteína, (Fig 2).

Elegimos este método por que ambas fosfolipasas presentan baja similitud con proteínas cuya estructura 3D se conoce por lo que no se puede aplicar el modelado por homología. Este método se basa en la búsqueda de estructuras 3D conocidas que presenten el mejor modelo de plegamiento compatibles al plegamiento predicho para la secuencia problema (8). De la serie de plegamientos obtenidos y ordenados por su valor E (probabilidad de que el resultado obtenido sea por el azar), nosotros mostramos en la Figura 2, dos de los que se encuentran dentro de los valores de certeza mayores al 95%. La información sobre estas estructuras 3D



PD1: Valor E: 1.61e-05, Longitud: 496 aa
Código de plegamiento: [c1f0ia](#)



PD2: Valor E: 0.00283, Longitud: 130 aa
Código de plegamiento: [c1gd5a](#)

solo es preliminar, ya que se hace necesario combinar los resultados de varios programas de threading y filtrar los modelos resultantes según información adicional que podamos tener de nuestra proteína (posiciones conservadas, mutantes, centro activo, etc.), para obtener una estructura 3D que sea base para estudios computacionales más avanzados (Ej: acoplamiento proteína-ligando: "docking").

Nuestro trabajo ha tenido el intento de brindarles un ejemplo sencillo de Investigación en Bioinformática, partiendo de la inmensa información con la que se cuenta en enormes bases de datos biológicas localizadas vía Internet y usando nuevos programas (software) para la manipulación e interpretación de esta información, algunos de los cuales son de distribu-

ción gratuita y otros se encuentran localizados en diversos servidores. La Bioinformática en una nueva ruta de investigación multidisciplinaria, donde el aporte de Investigadores tanto de las Ciencias Naturales, informática y matemáticas es valiosa.

AGRADECIMIENTO

A la Dra. Luisa Negrón Ballarte por la confianza y facilidades brindadas para la realización del presente trabajo.

REFERENCIAS BIBLIOGRAFICAS

1. Liscovith, M, Malgorzata, Fiucci, G & Tang, X. Phospholipase D: molecular and cell biology of a novel gene family. *Biochem J* 2000; 345:411-415.
2. W. Plant Phospholipases. *Annu Rev Plant Physiol* 2001; 52:211-231
3. Lawrence, C., Altschul, S., Boguski, M., Liu, J., Neuwald, A. & Wootton, J. Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment. *Science* 1993; 262: 208-214 .
4. Leiros, I., Secundo, F., Zambonelli, C., Servi, S. & Hough, E. The First Crystal Structure of a Phospholipase D. *Structure Fold Des* 2000; 8(6):655-67.
5. Shen, Y., Xu, L. & Foster, D . Role for Phospholipase D in Receptor-Mediated Endocytosis. *Mol Cell Biol* 2001; 21(2):595-602.
6. Min, D. S., Ahn, B-H. & Jo. Y-H. Differential Tyrosine Phosphorylation of Phospholipase D Isozymes by Hydrogen Peroxide and the Epidermal Growth Factor in A431 Epidermoid Carcinoma Cells. *Mol Cells* 2001; 11 (3):369-378.
7. Claverie J-M. & Notredame C. . Building a Multiple Sequence Alignment. En: *Bioinformatics for Dummies*, 1a ed, New York: Wiley Publishing Inc; 2003.p. 279-314.
8. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol*, 2000. 299(2):499-520.