

Predicción in silico de la estructura y función de la proteína hipotética Poly(ADP-ribosa) glicohidrolaza de *Caenorhabditis remanei*

Autores:

María Alejandra Rengifo¹, Laura Jiménez², Janneth González³, George Barreto⁴

¹Departamento de Microbiología, Facultad de Ciencias, Pontificia Universidad Javeriana. Colombia. ²Departamento de Microbiología, Facultad de Ciencias, Pontificia Universidad Javeriana. Colombia. ³ Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana. Colombia. ⁴ Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana. Colombia.

RESUMEN

Nuestro estudio se centra en la predicción de la estructura y función de la proteína Poly(ADP-ribosa) glicohidrolaza (PARG) en *Caenorhabditis remanei*. Inicialmente se realizó un alineamiento local por medio de BLAST con el fin de encontrar proteínas homólogas PARG.n Posteriormente se determinaron las características fisicoquímicas por medio del servidor ProtParam, seguidamente se realizó la predicción de la estructura secundaria mediante la herramienta Jpred3 de EXPASY. Finalmente se determinó la estructura terciaria de la proteína hipotética por medio de Phyre2. Con el fin de validar la estructura arrojada por dicho programa se utilizó SwissModel. Al emplear las herramientas anteriormente mencionadas se encontró la presencia del dominio PARG_Cat el cual está asociado a una función catalítica, hidrólisis de compuestos N-glicosil. La estructura 3D al ser validada se encontró por medio del diagrama Ramachandran que únicamente el 1,5% de los residuos se encontraban en zonas no permitidas. Se encontró que PARG está directamente relacionada con la función de una proteína presente en *C. elegans*, debido al 99% de cobertura mostrado al realizar el alineamiento local (BLAST). De tal manera se puede inferir que la proteína hipotética tiene una función similar en ambos nemátodos, la cual es la actividad catalítica de la proteína, particularmente de hidrólisis de compuestos N-glicosil y además la posible relación de PARG con la eliminación por apoptosis de células en exceso que debe eliminar el nemátodo para llevar a cabo su ciclo de vida completo.

PALABRAS CLAVE: PARG, *C. remanei*, actividad catalítica, Hidrólisis de compuestos N-glicosil, apoptosis.

INTRODUCCIÓN

Caenorhabditis remanei es un microorganismo, caracterizado como un gusano redondo de vida libre. Su hábitat por lo general son lugares ricos en nutrientes. Este microorganismo se encuentra en todo tipo de ambientes. Se han caracterizado dentro de la cadena trófica. Estas características hacen que *C. remanei* sea considerado como un indicador de la calidad del suelo y controladores biológicos de fitopatógenos principalmente.¹

C. remanei, posee la proteína Poly(ADP-ribosa) glicohidrolaza (PARG), la cual se ve involucrada en las primeras respuestas celulares ante el daño del ADN, apoptosis, la supervisión de la integridad del genoma, entre otras funciones. PARG trabaja junto con la enzima poly(ADP-ribosa) polimerasa (PARP), pues la función de PARP es básicamente la detección de daños en el ADN los cuales son ocasionados durante el proceso de muerte celular, isquemia, entre otras condiciones. PARP utiliza como sustrato NAD(+) con el fin de llevar a cabo diversos procesos de señalización y así producir polímeros de ADP-ribosa, dichos polímeros serán degradados por PARG. Dicha proteína aun no se ha caracterizado, aunque en nemátodos como *Caenorhabditis elegans* la actividad de la PARG se ve evidenciada en la eliminación por apoptosis de células en exceso, pues es necesario que se eliminen 131 células para llegar a estado adulto, por ende se puede decir que dicho proceso es esencial para que dicho microorganismo lleve a cabo su ciclo de vida completo. Evidenciándose así, la importancia de la actividad PARG en el desarrollo del nemátodo es importante tener conocimiento acerca de la estructura tridimensional y la función de esta proteína hipotética para avanzar en el mecanismo de acción de la misma. Debido a que este estudio se ve obstacuído por dificultades experimentales para la sobreexpresión y cristalización de esta es necesario hacer estudios bioinformáticos para poder obtener más conocimiento sobre esta proteína.² Es debido a esto que nuestro estudio se centra en la predicción de la estructura y función de la proteína Poly(ADP-ribosa) glicohidrolaza en *Caenorhabditis remanei*.

MATERIAL Y MÉTODOS

Para el análisis de la estructura y función de la proteína hipotética, se emplearon diversas herramientas y software. La primera secuencia de la proteína hipotética (Acc. No. >gi|308253153|gb|EFO97105.1) se obtuvo del GenBank en National Centre for Biotechnology Information (NCBI). Posterior a esto, la secuencia fue comparada para hallar secuencias homologas que se encuentran en distintas bases de datos usando Basic Local Alignment Search Tool (BLAST).

Para el cálculo de las propiedades físico-químicas de la proteína se empleó el servidor ProtParam. Los parámetros que se utilizaron fueron el peso molecular, pI teórico, composición de los aminoácidos, composición atómica, coeficiente de extinción (se estima para la vida media), índice de inestabilidad, índice alifático y promedio general de hidropaticidad.³

La estructura secundaria de la proteína se analizó mediante la herramienta Jpred3 de EXPASY, este emplea el algoritmo Jnet para realizar predicciones más precisas. Además de la estructura secundaria de la proteína JPred también hace predicciones sobre la accesibilidad de solventes y regiones en espiral (método Lupas).⁴

Se empleó la herramienta Phyre2 la cual permitió realizar la estructura terciaria de la proteína a partir de la secuencia de aminoácidos obtenida del NCBI. Esta

herramienta predijo los residuos que están probablemente más relacionados a la función de la proteína permitiendo realizar además un plegamiento inverso.⁵

El servidor en línea Swiss Model se empleo para la evaluación de la estructura 3D de la proteína a partir de proteínas homólogas de esta. Esta herramienta identifica plantillas estructurales a partir de secuencias homologas halladas en BLAST.⁶

Para la predicción celular de la proteína se llevo acabó mediante el servidor CELLO v 2,5, esté se emplea para tanto organismos procariotas como eucariotas.⁷

Adicionalmente, se realizó la búsqueda de sitios activos de la proteína hipotética mediante el servidor I-TASSER.

RESULTADOS

Al realizar un alineamiento local por medio de BLAST, se encontró presencia del dominio PARG_cat(Poly(ADP-ribose)glycohydrolase), el cual tiene una expresión ubicua exo y endoglicohidrolaza, y tiene como función mediar procesos de muerte celular. El análisis fisico-químico llevado a cabo mediante Protparam, revela que la proteína hipotética está compuesta por 619 aminoácidos y un punto isoeléctrico teórico de 4,67. El ácido glutámico es el aminoácido presente con mayor frecuencia en la secuencia de la proteína hipotética, con un porcentaje de 13,7; y el aminoácido con menor porcentaje en la secuencia es el triptófano con 0,2%. El número total de residuos negativamente cargados (Asp + Glu) fue de 118, y el número total de residuos positivamente cargados (Arg + Lys) fue de 72. El índice de inestabilidad de la proteína fue de 55,09, lo que la clasifica como una proteína inestable; por último el análisis fisico-químico reveló que el promedio general de hidropaticidad de la proteína fue de -0,762. (Anexo 1-Tabla1)

Para la predicción de la función de la proteína se utilizó InterProScan, mediante dicho servidor fue posible confirmar la presencia de dicho dominio relacionado con la actividad catalítica de la proteína, particularmente de hidrólisis de compuestos N-glicosil (Anexo 2-Figura 1). Adicionalmente mediante la herramienta PANTHER, fue posible determinar la familia de la proteína: Poli (ADP-ribose)glicohidrolasa. La estructura terciaria de la proteína fue realizada mediante el servidor Phyre2, dicha estructura fue validada mediante Swiss-Model, herramienta que arrojó la estructura estereoquímica del modelo, encontrándose que únicamente el 1,5% de los residuos se encuentran en posiciones no permitidas(Figura 1). Por lo tanto la estructura 3D diseñada por Phyre, es una estructura posible (Figura 2).

Al hacer el alineamiento local, empleando BLAST se encontró que estas directamente relaciona con la función que esta tiene en el organismo *C. elegans*, debido al 99% de cobertura que tiene la secuencia de la proteína de este nemátodo respecto a la de *C. remanei*. De tal manera se puede inferir que la proteína hipotética tiene una función similar en ambos nemátodos, la cual es la actividad catalítica de la proteína, particularmente de hidrólisis de compuestos N-glicosil.

Al realizar la búsqueda de sitios activos utilizando I-TASSER, se encontró que la proteína hipotética tiene posiblemente seis sitios activos: en el residuo 140, ácido glutámico(E140); en el residuo 143, alanita (A143); en el residuo 97 ácido glutámico (E97); en el residuo 100, Glutamia (Q100); en el residuo 101, lisina (K101); y por último en el residuo 105, asparagina (N105). Fig.2

DISCUSIÓN

Al momento de ajustar los parámetros para realizar BLAST, se selecciono como base de datos inicialmente Protein Date Bank Proteins (PDB) con el fin de encontrar secuencias homólogas con proteínas experimentales. No se encontró ninguna secuencia con un puntaje(score) alto, confirmándose así que la proteína PARG es hipotética. Posterior a esto se realizó nuevamente un BLAST, pero esta vez utilizando como base de datos secuencias de proteínas no redundantes (Non.redundant protein sequences (nr)). Aquí se encontró que PARG está directamente relacionada con la función que esta tiene en el organismo *C. elegans*, debido al 99% de cobertura que tiene la secuencia de la proteína de este nemátodo respecto a la de *C. remanei*. Esto se puede confirman debido a que ambos comparten varias características, como por ejemplo, ambos pasan por 4 estadios larvales y son muy cercanos filogenéticamente; ambos están en el clado "Eurhabditis", y comparten los mismos hábitats para su desarrollo.⁸ Mediante esta actividad también se descubrió la presencia de un dominio determinante para la predicción de la función de esta proteína hipotética, pues éste está relacionado con la actividad catalítica glicohidrolasa reportada para una proteína procedente de *C. elegans*.

Debido a que *C. elegans* para llegar a su estadio de vida adulta debe eliminar 131 células por apoptosis podríamos inferir que este dominio encontrado puede estar relacionado con el proceso de apoptosis que este nemátodo debe realizar para llevar a cabo su ciclo de vida.

El índice de hidropaticidad (GRAVY) clasifica a la proteína hipotética como una proteína hidrofílica, pues este valor es de -0,762 para PARG. Debido a que generalmente las proteínas de señalización son proteínas hidrofílicas, PARG puede estar involucrada con tales procesos, tal como habíamos nombrado, en procesos de señalización para la reparación del ADN.

Debido a que generalmente las funciones de una proteína están asociadas con la localización subcelular de esta, determinar la ubicación en la célula de una proteína hipotética puede significar un avance para entender y predecir la función de esta. Es por esto que se utilizó CELLO con el fin de predecir la localización subcelular de PARG, la predicción de dicha herramienta clasificó a PARG como una proteína mayoritariamente nuclear. Lo anterior corrobora la función a la que está asociado el dominio PARG_cat encontrado, y así mismo ayuda a avanzar en el funcionamiento que posiblemente pueda tener dicha proteína a nivel de reparación de material genético y de apoptosis que esta tiene en *C. elegans*.

La predicción de la estructura terciaria se realizó por plegamiento inverso (threading) mediante el servidor Phyre2, esto debido a que no se encontró un homólogo en PDB. Debido a que esta técnica se busca una secuencia compatible con un determinado plegamiento, con el fin de formar un esqueleto que mejor se ajuste a la secuencia de la proteína problema, este método tiene varias desventajas, pues, utiliza proteínas homólogas muy lejanas con un porcentaje de identidad menor al 15%. Sin embargo, para el presente estudio no era posible realizar la predicción por medio de modelado por homología, pues al realizar BLAST, y no encontrarse homólogos en PDB, la siguiente opción para la predicción de la estructura terciaria era por medio del plegamiento inverso.⁹

Es por lo anterior, que validar la estructura 3D era más que necesario por lo que se debía asegurar la calidad del modelo en 3D arrojado por Phyre2. Para tal fin se utilizó Swiss-model y, la calidad estereoquímica que mostró que el modelo fue que el 86,4% de los residuos se encuentran en regiones favorables, el 9,5% en regiones favorables permitidas y únicamente el 1,5% se encuentra en regiones no permitidas, por ende se puede decir que el modelo 3D, es un modelo potencialmente confiable y de buena calidad.

Por último, al realizar la búsqueda de sitios activos se encontraron seis proteínas con un sitio activo similar. Una de ellas es una nucleoporina (Nup1p) presente en *Saccharomyces cerevisiae*, la cual tiene como función el transporte de macromoléculas a través de la envoltura nuclear. A pesar de que PARG no es una proteína que este directamente asociada a esta función de transporte, esto ayuda a corroborar su ubicación cerca al núcleo y las función de reparación de ADN reportada en la literatura. La proteína con un sitio activo similar que le sigue a la anteriormente nombrada, es una ligasa ubiquitina presente en *Homo sapiens*, asociada a una función de regulación de diversos sustratos, esta función puede ser análoga a la función de PARG para regular los niveles de los polímeros de ADP-ribosa generados al momento de la detección de daños en el ADN.

CONCLUSIONES

Por medio de las herramientas utilizadas fue posible inferir que PARG es una proteína asociada a una actividad catalítica, la cual está relacionada con el dominio PARG_cat encontrado, que tiene como función mediar procesos de muerte celular. De tal manera que, al encontrar el índice de hidropaticidad (GRAVY), el cual indica que la proteína hipotética es hidrofílica y que su localización subcelular es a nivel nuclear (predicción de la herramienta CELLO) se logró un avance significativo en el entendimiento de la función de dicha proteína, pues generalmente las proteínas hidrofílicas son proteínas involucradas en procesos de señalización celular, lo cual sugiere en este caso que PARG es una proteína que interviene en cascadas de señalización para reparación del ADN.

La búsqueda de sitios activos en la proteína hipotética sirvió para avanzar en el entendimiento de la función, la estructura y la posible localización a nivel subcelular de esta. Pues, no sólo se encontró una función similar o asociada a las funciones de PARG, sino que también se confirmó la posible localización a nivel nuclear de la proteína debido a las funciones asociadas a estos sitios activos similares presentes en otras proteínas.



BIBLIOGRAFÍA

1. De las Heras J, Fabeiro C, Meco R. Fundamentos de agricultura ecológica. Ediciones de la Universidad de Castilla-La Mancha. 2003. 139p.
- 2 Cherezov V, Abola E, Stevens RC. Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods in Molecular Biology* 2010; 654, 141-168.
3. Idrees S, Nadeem S, Kanwal S, Ehsan B, Yousaf A, Rajoka M. In silico Sequence Analysis, Homology Modeling and Function Annotation of *Ocimum basilicum* Hypothetical Protein G1CT28_OCIBA. *INT.J. BIOAUTOMATION*. 2012;16(2),111-118
4. <http://www.compbio.dundee.ac.uk/www-jpred/about.html>. Consultada el 1 de octubre de 2012.
5. Kelley L, Sternberg M. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Publishing Group*. Volumen 4. N° 3. 2009.
6. <http://swissmodel.expasy.org/repository/?pid=smd03>. Consultada el 1 octubre de 2012.
7. Idrees S, Nadeem S, Kanwal S, Ehsan B, Yousaf A, Rajoka M. In silico Sequence Analysis, Homology Modeling and Function Annotation of *Ocimum basilicum* Hypothetical Protein G1CT28_OCIBA. *INT.J. BIOAUTOMATION*. 2012;16(2),111-118.
8. KIONTE K., FITCH D. The phylogenetic relationships of *Caenorhabditis* and other nematodes. *Worm Book*, The online *C. elegans* online biology.
9. Flores O., Rendón J., Martínez F., Guerra G., Erick Sierra E., Pardo J. Las herramientas del modelado molecular. (ISSN-0188-137X)

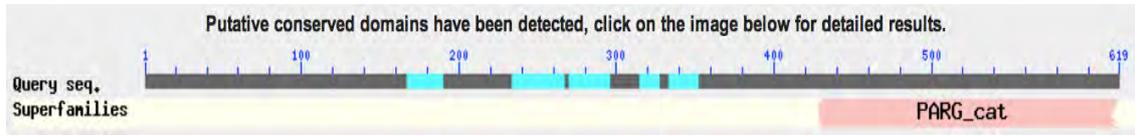
ANEXOS

Anexo 1.

Tabla 1: Propiedades físico-químicas de la proteína (composición de aminoácidos)

Aminoácido	Número de residuos	Porcentaje del residuo
Ala (A)	33	5,3
Arg (R)	37	6,0
Asn (N)	32	5,2
Asp (D)	33	5,3
Cys (C)	8	1,3
Gln (Q)	30	4,8
Glu (E)	85	13,7
Gly (G)	32	5,2
His (H)	8	1,3
Ile (I)	29	4,7
Leu (L)	40	6,5
Lys (K)	35	5,7
Met (M)	24	3,9
Phe (F)	22	3,6
Pro (P)	23	3,7
Ser (S)	55	8,9
Thr (T)	29	4,7
Trp (W)	1	0,2
Tyr (Y)	24	3,9
Val (V)	39	6,3
Pyl (O)	0	0,0
Sec (U)	0	0,0

Anexo 2



Detección de dominios conservados.

Anexo 3

CELLO RESULTS

SeqID: gil308253153|gb|EFO97105.1| hypothetical protein CRE_30484 [Caenorhabditis remanei]

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Nuclear	0.842
N-peptide Comp.	Nuclear	0.390
Partitioned seq. Comp.	Cytoplasmic	0.451
Physico-chemical Comp.	Nuclear	0.878
Neighboring seq. Comp.	Nuclear	0.647

CELLO Prediction:

Nuclear	3.097 *
Cytoplasmic	0.794
PlasmaMembrane	0.599
Mitochondrial	0.118
Golgi	0.098
Extracellular	0.067
Chloroplast	0.065
Cytoskeletal	0.052
ER	0.051
Vacuole	0.034
Peroxisomal	0.016
Lysosomal	0.008

[[Home](#) | [Documentation](#)]

Resultados CELLO

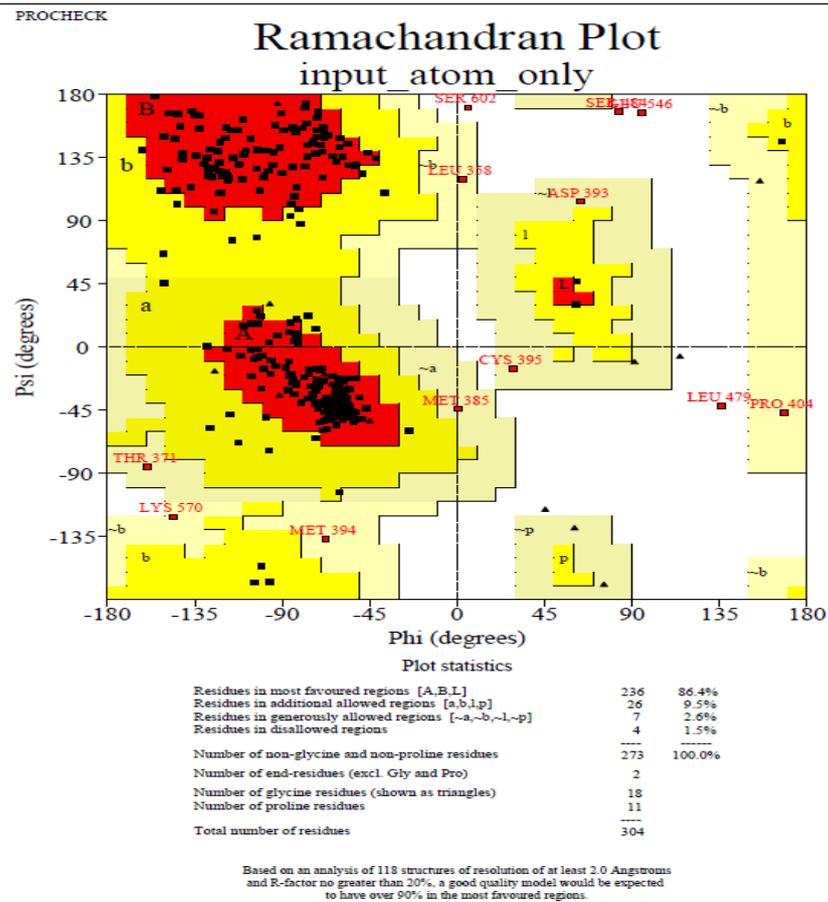


Figura 1. Ramachandran Plot.



Figura 2. Estructura tridimensional de la proteína hipotética. Servidor Phyre2.