

# **Análisis bioinformático de una cisteín proteasa presente en plantas de *Nicotiana benthamiana*, la proteína Metacaspasa 4.**

Ana Marchesini<sup>1,2</sup>

<sup>1</sup> Cátedra de Bioinformática, Área de Biotecnología y Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina.

<sup>2</sup> Instituto de Biotecnología y Biología Molecular. Universidad Nacional de La Plata. Facultad de Ciencias Exactas - CONICET; Argentina.

## **RESUMEN**

Los virus se encuentran dentro de los principales patógenos de los cultivos, causando pérdidas devastadoras, por lo que entender sus mecanismos de infección e interacción con la planta es fundamental para lograr su erradicación. Una de las formas es identificar las proteínas del hospedador partícipes en los mecanismos de defensa ante estos patógenos, como lo es la metacaspasa-4 (MCA4), una peptidasa que se activa ante estrés iniciando una reacción en cadena con el propósito de atacar al organismo invasor y evitar su propagación. En este trabajo se aborda un análisis bioinformático de la MCA4 presente en plantas *N. benthamiana* (NbMCA4) a fin de predecir su posible estructura e identificar segmentos fundamentales en su función, para poder inferir sobre su rol durante la infección viral.

PALABRAS CLAVE: Metacaspasa; *Nicotiana benthamiana*.

## **INTRODUCCIÓN**

El crecimiento de la demanda de productos vegetales a nivel mundial, junto a la limitación de tierras cultivables, hacen cada vez más urgente la necesidad de mejorar las capacidades de producción en un marco de preservación del suelo y de sustentabilidad ambiental. Los virus están entre los principales patógenos de los cultivos, son responsables de pérdidas devastadoras en agricultura y dada la capacidad de adaptación a las estrategias de protección de los cultivos, su erradicación del ambiente parece imposible. Por esto, entender sus mecanismos de infección e interacción con la planta es fundamental para la producción de productos vegetales. Una de las características resaltantes de los virus es su capacidad de evadir los mecanismos de defensa de la planta, los cuales pueden ser receptores virales codificados por genes de resistencia (R) de la misma o mecanismos especializados, como PTGS y, para lograr un análisis profundo al respecto, primero se debe abordar el estudio de estos mecanismos.

Un grupo de proteínas, las metacaspasas (MCAs), funcionan como mecanismo primario de defensa de la planta ante el ataque de patógenos. Son cisteín proteasas representadas en todos los dominios de la vida (Uren et al., 2000), fueron descubiertas años más tarde que las caspasas a partir de una búsqueda de similitud estructural. De acuerdo a las preferencias de corte y estructura, las MCAs se han clasificado en 3 diferentes tipos, I, II y III, encontrándose sólo las MCAs tipo I y II en las plantas verdes (Viridiplantae). Ambos poseen el dominio catalítico conservado en las caspasas compuesto por las subunidades de 20 kDa (p20) y 10 kDa (p10); en la subunidad p20 se encuentran los residuos catalíticos Histidina-86 y Cisteína-139. Las MCAs tipo II poseen además una región de enlace de 160-180 aminoácidos entre p10 y p20. Estas proteínas se expresan como zimógenos y se activan por autoclivaje, esta actividad es regulada por modificaciones postraduccionales, como fosforilaciones, ubiquitinación, nitrosilación, cambios de pH y

concentración de iones, así como por la interacción con otras proteínas (Minina et al., 2017). Están involucradas en la muerte celular programada (PCD, *Programmed Cell Death*), programas de desarrollo y morfogénesis y en la contención de estreses bióticos y abióticos (He et al., 2008, Watanabe & Lam et al., 2011).

Nos propusimos identificar el rol que cumple la metacaspasa-4 presente en plantas de *Nicotiana benthamiana* (NbMCA4) durante la infección viral. La NbMCA4 es una metacaspasa tipo II, homóloga de la MCA4 de *A. thaliana*, la cual ha sido asociada a la activación de un programa de PCD bajo estrés oxidativo o ante el ataque de patógenos (Watanabe and Lam, 2011) a partir del clivaje en *trans* del PROPEP1 dando como producto PEP1, péptido que inicia una gran cadena de defensa en la planta (Hander et al., 2019). Con tal fin iniciamos un análisis de la biología de esta proteína, el cual se encuentra en actual estudio y muestra indicios de la importancia que podría tener en cuanto a la defensa de la planta. Nos planteamos entonces como objetivo de este trabajo realizar un estudio bioinformático de la NbMCA4 para poder recopilar mayor información en base a su secuencia, predecir su estructura y analizar su función en base a homologías, además de reconocer sitios potencialmente importantes para la misma, que permitirían el diseño de mutantes para fortalecer el análisis biológico.

## MÉTODOS Y RESULTADOS

### Análisis de secuencia

Para acceder a la secuencia de nucleótidos completa de NbMCA4 se utilizó la base de datos SolGenomics, ya que tanto en GenBank como en European Nucleotide Archive sólo se puede encontrar la anotación parcial (DQ084024), mientras que la secuencia primaria de la proteína no se encuentra anotada en forma completa en ninguna plataforma, la misma contiene una longitud de 418 aminoácidos, dando una proteína de 45,8 kDa de tamaño, y se proporciona a continuación

>NbMCA4

```
MAKKAVLIGINYPGTKAELKGCINDVKRMYSLIKRFGFSEEDITVLIDTDDSYTQPTGRNIRKVLSDLVGSAAEGDSL FVHYS  
GHGTRLPAETGEEDDTGYDECIVPCDMNLITDDDFRELVDKVPEGCRITIVSDSCHSGGLIDKAKEQIGESHKQGDDENEGH  
GSGFGFKKFLRRSVEDAFESRGIHIPRRHDRREEEESFAESSVIETEDGDQVHVKNKSLPLSTLIEILKQKTGKDDIDVGKLRP  
TLFDVFGEDASPKVKKFMKVIFNKLQHGKGESEGGFLGMVGNLAQEFLKQKLDENDES YAKPAMETHVEGKQEVYAGSG  
SRGLPDSGILVSGCQTDQTSADATPAGGDSYGALSNAIQEILAESDGPITNEEVTKARKKLQKQGYTQRPGLYCSDDHHVDA  
PFVC
```

A fin de analizar la estructura primaria de la NbMCA4, se procedió a una búsqueda de dominios en Pfam y en CDD (base de datos de dominios conservados, perteneciente a NCBI), encontrando en ambas plataformas un dominio conservado en su secuencia perteneciente a la superfamilia CASc en la región N-terminal de la proteína. Este dominio es típico de las caspasas, descrito en Pfam con el código de acceso 00656, llamado peptidase\_C14. Se encuentran dentro de este los dos residuos conservados pertenecientes al sitio activo de las MCAs, H86 y C139 (subrayados en la secuencia proporcionada arriba).



**Figura 1. Análisis en base a la secuencia primaria NbMCA4.** Se observa la detección del dominio Peptidase\_C14 perteneciente a la superfamilia de las CASc predicho por CDD.

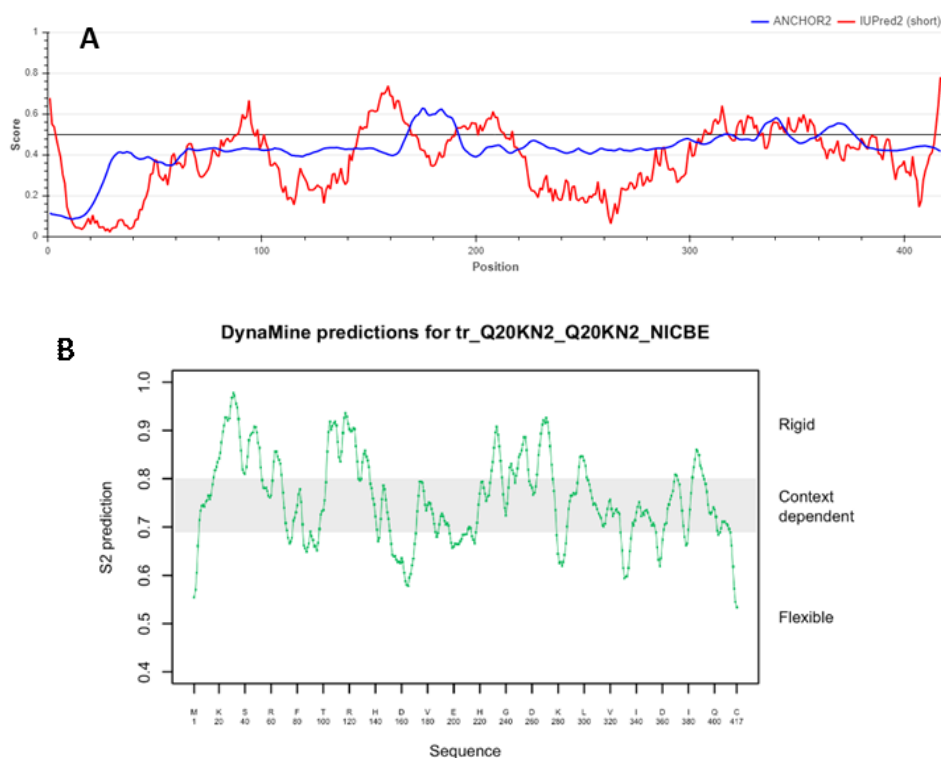
Con el objetivo de inferir estructuras secundarias a partir de su secuencia, se utilizó el programa PSIPred, el cual asigna posibles plegamientos a las regiones e informa la confianza en la predicción. Se puede ver en la

**Figura 2** que el arreglo aparentemente dominante serían las alfa-hélices, predichas con máxima confianza, unidas por una gran cantidad de loops. La misma búsqueda se realizó a partir del servidor Quick2D, el cual muestra predicciones de varios programas, incluido el PSIPred, sin indicar la confianza. Los resultados se muestran en el material suplementario (**Figura S1**) donde se puede ver que mayormente hay coincidencias entre ellos. En cuanto a la búsqueda de motivos o regiones coiled-coils a partir de la secuencia primaria, se corrieron los servidores Prosite y Coils, respectivamente, sin encontrarse nada en ninguno de los dos (datos no mostrados). Así mismo, se buscaron posibles segmentos transmembrana, para lo cual corrimos el servidor Phobius, donde se indicó la ausencia de un dominio de anclaje a lo largo de toda la cadena aminoacídica (dato no mostrado).



**Figura 2. Predicción de estructuras secundarias a partir de la secuencia primaria utilizando el programa PSIPred. Las referencias por colores se observan en el extremo inferior.**

Tanto la plataforma IUPred como DynaMine son muy usadas para la predicción de regiones desordenadas dentro de la secuencia proteica, se corrieron ambas, los resultados se muestran en la **Figuras 3a y 3b**. La mayor parte de la secuencia queda cerca del cutoff en ambos programas, lo cual podría ser coincidente con la gran cantidad de loops que se observaban en las predicciones de estructura secundaria, y aportan movimiento a la proteína. Además de segmentos cortos en los extremos, IUPred predice una región desordenada con un score levemente mayor al cutoff entre los residuos 150 y 170 aproximadamente, vista también en DynaMine junto con otros picos. Se ve en ambas plataformas segmentos con estructura muy ordenada, uno en la región N-terminal, abarcando aproximadamente del residuo 20 al 40, coincidente con la región donde se predice la presencia de una alfa-hélice de esa longitud, otro entre los aminoácidos 230-280, segmento en el cual se predice la presencia de 4 alfa-hélices unidas por loops muy cortos, y otro en el segmento 110-130.



**Figura 3.** Predicción de regiones ordenadas/desordenadas a partir de la secuencia de NbMCA4 por los programas **a.** IUPred y **b.** DynaMine

### Análisis de homología

Se buscaron secuencias homólogas usando el Blastp de NCBI para analizar la conservación de NbMCA4. Como base de datos se usó "non-redundant protein sequences", se indicó un máximo de 250 secuencias alineadas, un E-value de 0.05, ktuple de 6 aminoácidos y BLOSUM 62 como matriz. De un total de 250 proteínas encontradas todas presentaron más del 60% de similitud, con un *coverage* cercano o igual al máximo y valor de e-value 0.0. Se seleccionaron 81 homólogas pertenecientes al clado Mesangiosperma con cobertura de blast mayor o igual a 98% y se procedió a alinearlas a partir de su secuencia primaria.

En el material suplementario se puede ver el alineamiento resultante en formato MSF (**Figura S2**). Al hacer un análisis sobre éste encontramos que NbMCA4 comparte con todas las secuencias un porcentaje de identidad mayor o igual a 70% con gran cantidad de segmentos conservados y el número de gaps, dejando fuera del análisis a las escasas proteínas con distinta longitud, es muy bajo. La totalidad de las secuencias mantienen los residuos correspondientes al sitio activo H86 y C139 presentes en el dominio peptidase\_C14. Estos datos muestran con seguridad que la MCA4 es una proteína muy conservada evolutivamente.

## Modelado de la proteína: estructura terciaria

Para inferir la estructura terciaria, se realizó un modelado por homología. Como primer paso, la asignación de plegamiento, se utilizaron los programas HHpred, seleccionando como base de datos PDB\_mmCIF70, y la plataforma de NCBI, BLASTp, usando como base de datos la PDB. En ambos se introdujo la secuencia primaria de NbMCA4 y los resultados mostraron como mejor template a la estructura cristalizada de la metacaspasa-4 de *A. thaliana* (AtMCA4), con un porcentaje de identidad del 70%, alto score y e-value cercano al 0 para HHpred y 0 para BLASTp, cubriendo el 100% de la secuencia. No hizo falta realizar un análisis con métodos más sensibles como psiBLAST ya que este porcentaje de similitud es más que suficiente para modelar la estructura terciaria de la NbMCA4 a partir del template, como así tampoco un análisis filogenético para evaluar cercanía evolutiva. En el caso de BLASTp, se encontró tanto la AtMCA4 wild type (wt) (código PDB: 6W8S) como su versión mutante C139A, donde se reemplazó la cisteína presente en el sitio activo por una adenina (código PDB: 6W8R), mientras que en HHpred sólo figura el mutante C139A. Continuamos nuestro análisis con la versión wt, 6W8S, ya que en 6W8R la mutación de un residuo tan importante podría afectar la estructura de la proteína a partir de la cual realizaremos el modelado.

Para corroborar la calidad del alineamiento a modo de asegurarnos un buen modelado se procedió a realizar el mismo en EMBOSS Needle (algoritmo Needleman-Wunsch), usando la matriz BLOSUM62 y analizarlo en Gendoc, obteniéndose un porcentaje de similitud del 81%, y 67% de identidad entre las secuencias, con sólo 23 gaps en todo el alineamiento. En la **Figura S3** del material suplementario se puede ver el mismo.

Una vez que validamos el alineamiento, se procedió a realizar una búsqueda en la base de datos PDB de todas de todas las estructuras cristalizadas para la proteína AtMCA4, encontrándose 6W8S con una resolución de 3.484 Å, y 6W8T con una resolución de 3.2 Å, esta última corresponde a la estructura vista cuando se tratan los microcristales con Ca<sup>+2</sup>; para comprender mejor esto nos remitimos a la publicación donde se reportaron las estructuras (Zhu et al., 2020) y vimos que la AtMCA4 se activa por clivaje en el residuo K225 en presencia de una concentración de Ca<sup>+2</sup> elevada, y que la diferencia en las estructuras es que 6W8S corresponde a la proteína entera, presentando una distancia entre K225 y los residuos catalíticos C139 y H86 apta para el ataque nucleofílico, mientras que en la estructura 6W8T se puede ver que hubo un clivaje en el residuo K225 (las estructuras se muestran en la **Figura S4** del material suplementario). Decidimos proceder al análisis con la estructura del zimógeno, 6W8S.

Se pudo ver que la proteína AtMCA4 cristaliza como un monómero, con iones sulfato co-cristalizando con ella. Se identifica en ella un sólo dominio peptidase\_C14 tal como habíamos predicho para la NbMCA4. Presenta una región desordenada (153-172) coincidente con los resultados arrojados por IUpred para NbMCA4.

Una vez encontrado el template, se descargó su secuencia en PDB, la cual tenía coordenadas desde el residuo 2 hasta el 420, presentando 64 missing residues, de los cuales 57 abarcan la región 152-208. Con estos datos se procedió a modelar a partir del programa Modeller, el cual genera distintas estructuras tratando de violar la menor cantidad de restricciones posibles. Se obtuvieron 25 modelos, siendo NbMCA4.B99990014 el más certero (**Figura 4**). Tal como se había predicho en su estructura secundaria, la NbMCA4 presenta una gran cantidad de segmentos hélices alfa, unas pocas beta plegadas y gran cantidad de loops.

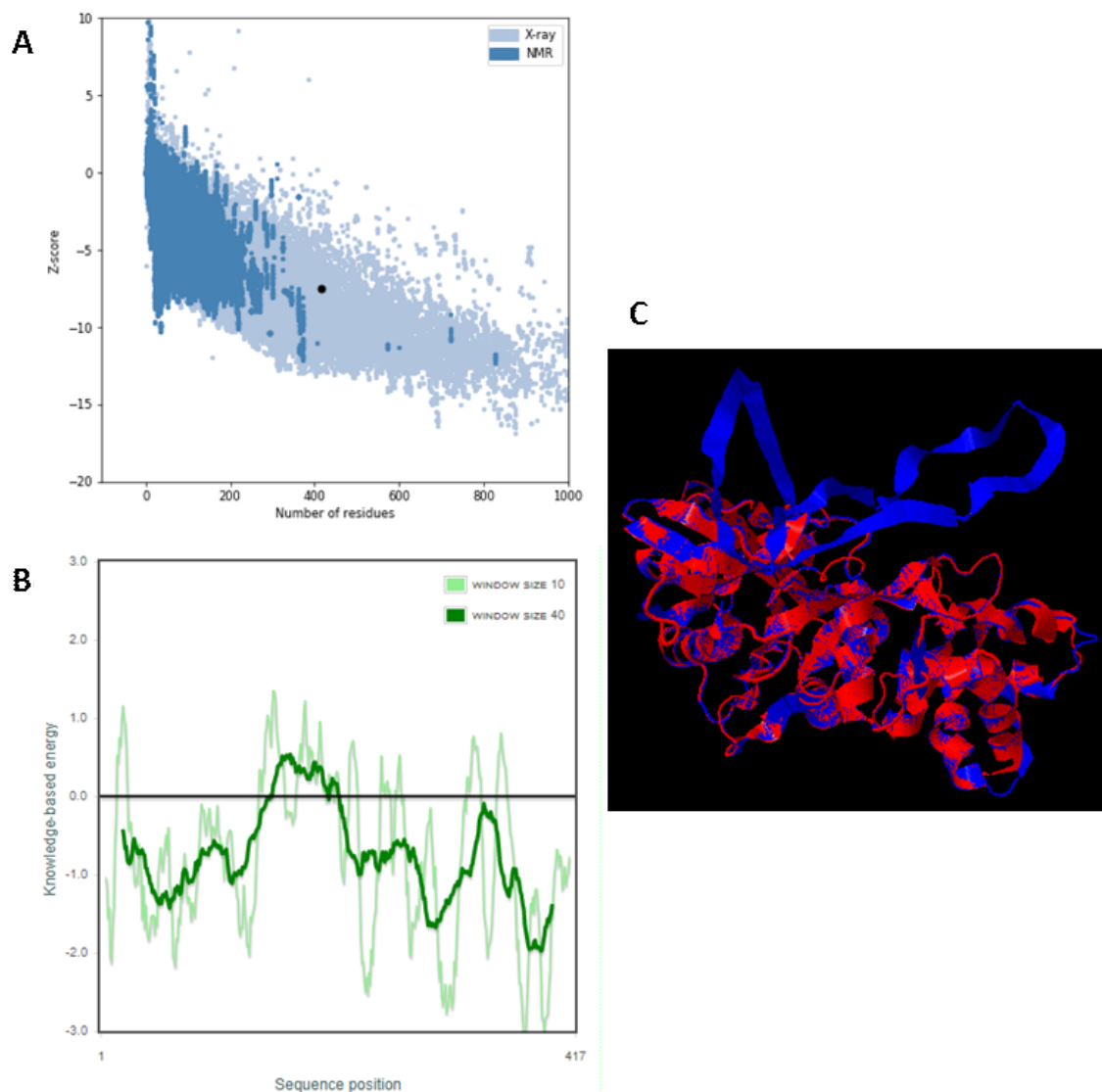


**Figura 4.** Estructura obtenida a partir del modelado por homología de NbMCA4 usando el programa Modeller, con la estructura 6W8S de AtMCA4 como template.

Para evaluar el modelado se usó el programa ProSA, obteniéndose en la evaluación global un Zscore de -7.54, el cual cae dentro de la población de estructuras de referencia (**Figura 5a**). En cuanto a la evaluación local en el mismo programa, pudimos corroborar que el modelo es en su mayoría bueno, excepto por una corta región que presenta elevada energía, la cual podría corresponder al segmento de missing residues donde el programa tuvo que modelar *ab initio* (**Figura 5b**). En segundo lugar se realizó un alineamiento estructural entre el modelo obtenido y 6W8S a partir del programa TM -Align, donde se ve que el primero de estos muestra dos loops que no corresponden a la estructura de su homólogo (**Figura 5c**). El mismo análisis se realizó por el programa SuperPose, el cual muestra un RMSD total de 0.72, el cual es un valor muy aceptable de alineamiento estructural (dato no mostrado).

### **Análisis taxonómico**

Para poder conocer el camino evolutivo de la proteína, se procedió a realizar un análisis taxonómico de esta utilizando el alineamiento múltiple anteriormente mostrado. Se utilizó el método de maximum likelihood realizado por el programa HyPhy para obtener el mejor modelo que explique la evolución de las secuencias, generando como primer árbol filogenético un Neighbor Joining, y asignando 4 categorías de velocidad de evolución para gamma distribution sea cual fuera el modelo de evolución resultante. El resultado arrojó que con un score de 31243.52409783448 y un likelihood de -15441.762, el mejor modelo es JTT+F. Con estos datos nos fuimos al programa PhyML para lograr estimar la topología de las secuencias alineadas a partir del modelo JTT+F+gamma, indicando que realice un análisis de bootstrapping no paramétrico (100 replicantes) para validar la información del alineamiento, el árbol resultante se puede ver en la **Figura 6** y a partir de él podemos inferir la cercanía evolutiva de NbMCA4 a las distintas homólogas, con un gran soporte en la información secuencial para los nodos resultantes.



**Figura 5. Evaluación del modelo estructural obtenido para NbMCA4. a y b.** Evaluación energética global y local, respectivamente, usando el programa ProsaII; **c.** Alineamiento estructural entre el modelo obtenido y la estructura de referencia, 6W8S.

### Predicción de la función

A la hora de predecir la función de una proteína de novo la única herramienta bioinformática que tenemos es la comparación con un homólogo cuya actividad esté reportada, y luego validar biológicamente esta información. En este caso, la proteína que vamos a utilizar es la AtMCA4.

La AtMCA4 está reportada como una cisteín proteasa que cliva específicamente luego de los residuos R o K, sin clivar sustratos específicos de las caspasas. Induce la muerte celular programada (PCD) al sensor estreses bióticos y abióticos, a partir del aumento de la concentración de  $Ca^{+2}$  intracelular. El sitio activo está compuesto por dos residuos, H86 y C139, los cuales están conservados en la NbMCA9, y cliva luego del residuo K225, el cual está presente en una región conservada en NbMCA4 como K224 (**Figura S3**), generando dos subunidades, p20 (1-225) y p10 (226-418). Sus anotaciones en Gene Ontology (GO) muestran que actúa en el citoplasma, citosol, mitocondria, membrana y plasmodesmos, funciona como una cistein-endopeptidasa con unión a proteínas presentando actividad hidrolasa y está involucrada como se había mencionado antes en la regulación de la PCD y de otros procesos biológicos involucrados en la defensa de la planta ante ataques bióticos y abióticos.



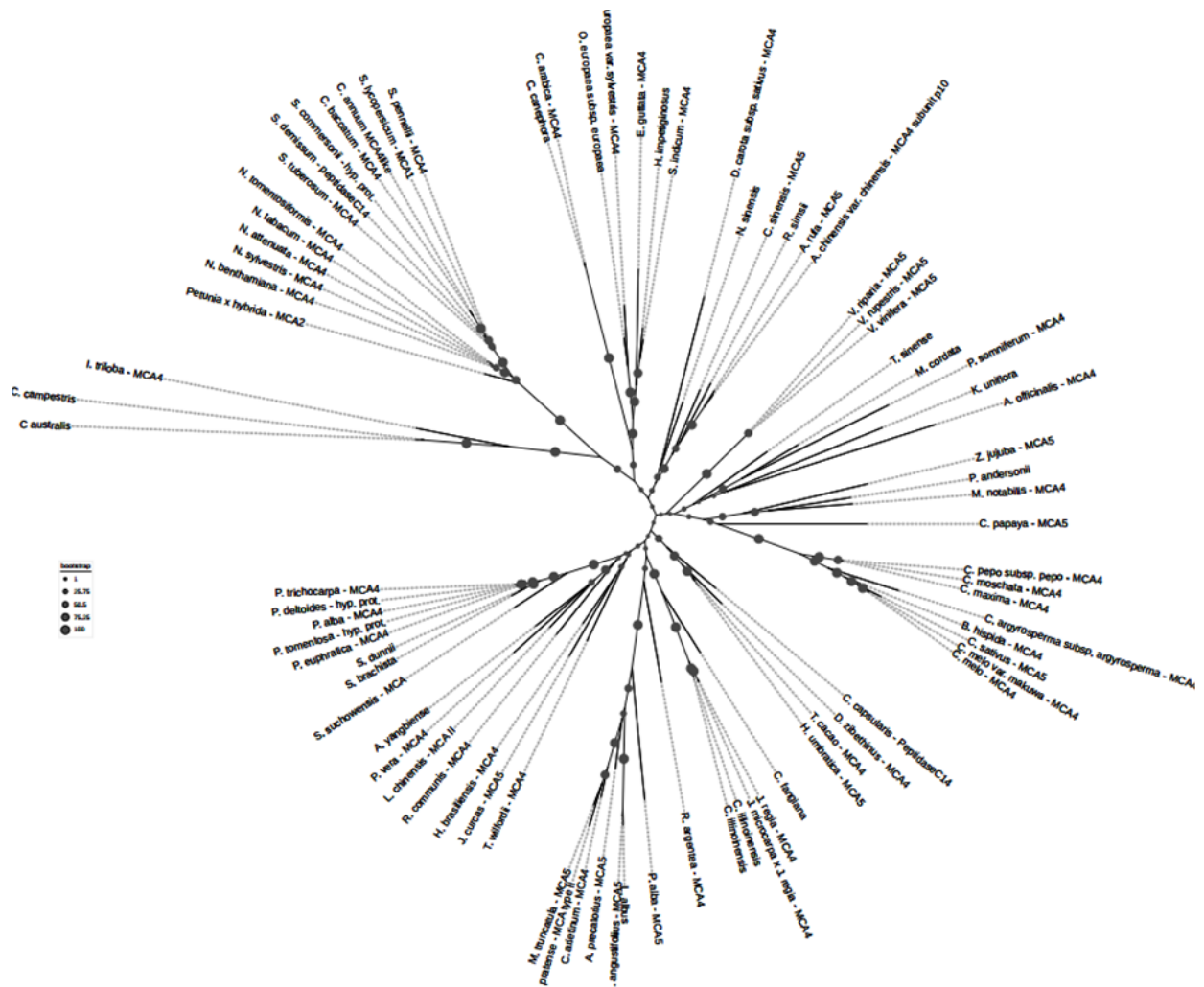


Figura 6. Árbol filogenético obtenido con el programa PhyML para 82 proteínas homólogas evolucionando a partir del modelo JTT+F+Gamma. Se muestra el soporte de cada nodo realizado a partir de un análisis de bootstrapping no paramétrico con 100 replicantes por el mismo programa.

A partir de la secuencia de NbMCA4 se corrió el programa Universal Evolutionary Trace (UET) y se vieron cuáles son los residuos más conservados evolutivamente, obteniéndose los resultados de la **Figura 7**, los cuales muestran una gran conservación a lo largo de toda la cadena, excepto por dos segmentos: 156-220 (región reportada con estructura desordenada en AtMCA4) y 267-338. Con esta misma secuencia pero agregando la estructura de referencia 6W8S cadena A (las 4 cadenas son idénticas) de AtMCA4 se inició el programa ConSurf y los resultados fueron similares: se conserva casi toda la secuencia exceptuando las mismas dos regiones (dato no mostrado).

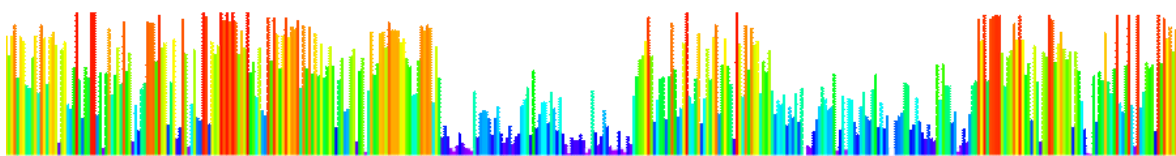


Figura 7. Análisis de residuos conservados evolutivamente en la MCA4 a partir del programa Universal Evolutionary Trace (UET). Los picos más altos y de color rojo corresponden a una mayor conservación, mientras que los de menor altura y azules a una menor conservación.



## CONCLUSIONES Y DISCUSIÓN

La proteína NbMCA4 posee un dominio conservado en su secuencia perteneciente a la superfamilia CASc, el cual contiene los residuos pertenecientes al sitio activo, H86 y C139, también presentes en ella.

La estructura secundaria predicha de la NbMCA4 presenta mayormente arreglos alfa-hélice con algunas beta-plegadas y gran cantidad de loops.

Según las predicciones de IUPred, la NbMCA4 presenta regiones desordenadas en sus colas y en el segmento 150-170, este último reportado en la estructura de su proteína homóloga, AtMCA4.

NbMCA4 es una proteína muy representada a lo largo del clado mesoangiosperma, conservando la mayor parte de los segmentos de su secuencia, con dos regiones variables, una de ellas correspondiente a una posible estructura desordenada.

La proteína homóloga con estructura conocida más cercana es la AtMCA4, la cual presenta un 81% de identidad con NbMCA4 cubriendo el 100% de la proteína y su estructura es un monómero que co-cristaliza con iones sulfato (PDB: 6W8S).

A partir de la estructura de AtMCA4 se modeló la de NbMCA4, obteniendo buenos valores energéticos globales y locales, excepto por una región de elevada energía, y un alineamiento estructural bueno, con un RMSD de 0.72, donde se observa el modelado erróneo de dos loops. Para un mejor modelado habría que optimizar la región correspondiente a esos loops.

La NbMCA4 funcionaría como una cisteín proteasa, y se activaría auto-clivándose a partir de los residuos H86 y C139 luego del aminoácido K224 luego de sentir una concentración de Ca<sup>2+</sup> intracelular elevada por un estrés biótico o abiótico sufrido por la planta hospedadora.

## BIBLIOGRAFÍA

De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Baiocchi A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W362-5. PubMed:16845026 [Full text] [PDF version].

Di Tommaso P, Moretti S, Xenarios I, Orobítz M, Montanyola A, Chang JM, Taly JF, Notredame C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 2011 Jul;39(Web Server issue):W13-7. doi: 10.1093/nar/gkr245. Epub 2011 May 9. PMID: 21558174; PMCID: PMC3125728.

Drozdetskiy A, Cole C, Procter J & Barton GJ. *Nucl. Acids Res.* (first published online April 16, 2015) doi: 10.1093/nar/gkv332

Elisa Cilia, Rita Pancsa, Peter Tompa, Tom Lenaerts, and Wim Vranken. The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic Acid Research* doi: 10.1093/nar/gku270 (2014)

Gábor Erdős, Zsuzsanna Dosztányi. Analyzing Protein Disorder with IUPred2A. *Current Protocols in Bioinformatics* 2020;70(1):e99

GO Enrichment Analysis: Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* Jan 2019;47(D1):D419-D426.

Hander T, Fernández-Fernández ÁD, Kumpf RP, Willems P, Schatowitz H, Rombaut D, Staes A, Nolf J, Pottier R, Yao P, Gonçalves A, Pavié B, Boller T, Gevaert K, Van Breusegem F, Bartels S, Stael S. Damage on plants activates Ca<sup>2+</sup>-dependent metacaspases for release of immunomodulatory peptides. *Science.* 2019 Mar 22;363(6433):eaar7486. doi: 10.1126/science.aar7486. PMID: 30898901.

He R, Drury GE, Rotari VI, Gordon A, Willer M, Farzaneh T, Woltering EJ, Gallois P. Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H<sub>2</sub>O<sub>2</sub> in Arabidopsis. *J Biol Chem*. 2008 Jan 11;283(2):774-83. doi: 10.1074/jbc.M704185200. Epub 2007 Nov 12. PMID: 17998208.

Jon Ison, Matúš Kalaš, Inge Jonassen, Dan Bolser, Mahmut Uludag, Hamish McWilliam, James Malone, Rodrigo Lopez, Steve Pettifer, Peter Rice, EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats, *Bioinformatics*, Volume 29, Issue 10, 15 May 2013, Pages 1325–1332, <https://doi.org/10.1093/bioinformatics/btt113>

Kouza M, Faraggi E, Kolinski A, Kloczkowski A. The GOR Method of Protein Secondary Structure Prediction and Its Application as a Protein Aggregation Prediction Tool. *Methods Mol Biol*. 2017;1484:7-24. doi: 10.1007/978-1-4939-6406-2\_2. PMID: 27787816.

Lu S et al. (2020). "CDD/SPARCLE: the conserved domain database in 2020.", *Nucleic Acids Res*. 48(D1):D265-D268.

Lukas Käll, Anders Krogh and Erik L. L. Sonnhammer. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, 338(5):1027-1036, May 2004

Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268.

Minina EA, Coll NS, Tuominen H, Bozhkov PV. Metacaspases versus caspases in development and cell fate regulation. *Cell Death Differ*. 2017;24(8):1314-1325. doi:10.1038/cdd.2017.18

New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. *Systematic Biology*, 59(3):307-21, 2010.

Pfam: The protein families database in 2021: J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman. *Nucleic Acids Research* (2020) doi: 10.1093/nar/gkaa913

Sergei L Kosakovsky Pond, Art F Y Poon, Ryan Velazquez, Steven Weaver, N Lance Hepler, Ben Murrell, Stephen D Shank, Brittany Rife Magalis, Dave Bouvier, Anton Nekrutenko, Sadie Wisotsky, Stephanie J Spielman, Simon D W Frost, Spencer V Muse, HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies, *Molecular Biology and Evolution*, Volume 37, Issue 1, January 2020, Pages 295–299, <https://doi.org/10.1093/molbev/msz197>

SuperPose v1.0 (2004) Rajarshi Maiti, Gary Van Domselaar, Haiyan Zhang, and David Wishart

Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM. Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell*. 2000 Oct;6(4):961-7. doi: 10.1016/s1097-2765(00)00094-0. PMID: 11090634.

Watanabe N, Lam E. Arabidopsis metacaspase 2d is a positive mediator of cell death induced during biotic and abiotic stresses. *Plant J*. 2011 Jun;66(6):969-82. doi: 10.1111/j.1365-3113X.2011.04554.x. Epub 2011 Apr 28. PMID: 21395887.

Wiederstein & Sippl (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* 35, W407-W410.

Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on TM-score, *Nucleic Acids Research*, 33: 2302-2309 (2005)

Zhu, P., Yu, XH., Wang, C. et al. Structural basis for Ca<sup>2+</sup>-dependent activation of a plant metacaspase. *Nat Commun* 11, 2249 (2020). <https://doi.org/10.1038/s41467-020-15830-8>