

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA

Carrera de Biología

Unidad de Bioinformática y Manejo de la
Información

“Catálogo y relaciones filogenéticas de las sintasas de quitina en animales”

Opción de Titulación por Tesis

Raúl Olivares Rosas

raulolivaresrosas@gmail.com

Alumno

Jorge Antonio Valdivia Anistro

jorge.valdivia@zaragoza.unam.mx

Asesor

Augusto César Poot Hernández

apoot@ifc.unam.mx

Director

Uri Omar García Vázquez

urigarcia@gmail.com

Asesor

M. en C. Guadalupe Bibriesca Escutia

Secretaria Técnica

biologia.stecnica@zaragoza.unam.mx



1. Introducción

En la década de los 80's aconteció un debate entre un grupo de investigadores que estudiaban el gen DG42 en *Xenopus laevis* y su expresión en etapas específicas del desarrollo embrionario. Por un lado sostenían que el producto final de dicho gen era el ácido hialurónico y por el otro sostenían que se trataba de quitina. En el momento del descubrimiento y estudio del gen DG42, no se conocía una función específica del mismo porque no presentaba homología alguna con cualquier otra proteína conocida. En años posteriores se le identificaría como un gen con función parecida al que produce quitina y al gen NodC que participa en la síntesis de oligómeros de quitina en hongos; algo interesante ya que hablamos de un vertebrado que en teoría no es capaz de sintetizarla (Tang, Fernandez, Sohn, & Amemiya, 2015). De hecho, la quitina es el segundo biopolímero más abundante en la biosfera y actúa como el componente estructural principal para diversas formas de vida, principalmente hongos (paredes celulares), crustáceos y artrópodos (como parte fundamental de su exoesqueleto) pero no en vertebrados (Varki, 1996).

A pesar de lo anteriormente mencionado, es posible la acumulación de quitina en algunos tejidos de vertebrados, pero es particularmente tóxica en las neuronas y, por ende, en todo el sistema nervioso. Su acumulación excesiva en el cerebro puede ser la razón del declive cognitivo en pacientes de Alzheimer (Sotgiu, Musumeci, Marconi, Gini, & Bonetti, 2008). Aunque es más común observar esta característica en la forma esporádica del Alzheimer que en su forma familiar (genética). Y aunque la quitina es más un agente neurotóxico que un componente importante de la fisiología de los vertebrados, se conservan en la mayoría de estas proteínas quitinasas que presuntamente participan en la defensa de parásitos. (Stern, 2017).

A pesar de que la quitina es esencialmente tóxica en vertebrados, su componente base (N-acetilglucosamina) es conocido por ser parte del grupo de oligosacarinas solubles, propiedad opuesta a la molécula de quitina que por naturaleza es insoluble. Esto llevó a otros investigadores a descubrir que, en ensayos *in vitro*, DG42 es capaz de sintetizar cadenas cortas de oligómeros además de otros productos más grandes. Fue así que este gen fue propuesto

como la primera sintasa de oligosacaridos de quitosano¹, y además como posible sintasa de quitina. Teniendo esta última una interesante homología con *hasA*: gen de *Streptococcus* que sintetiza polímeros de hialuronano (Varki, 1996). Actualmente se sabe que el producto del gen DG42 durante la embriogénesis de la rana puede sintetizar tanto hialuronano como quitina en periodos específicos, pero la duda de cómo puede el mismo gen sintetizar dos proteínas distintas solo cambiando el sustrato persiste. Debido a que tanto el ácido hialurónico y la quitina comparten muchas similitudes en sus sintasas y componentes base, puede que la respuesta resida en su historia evolutiva.

2. Antecedentes

2.1 Quitina y ácido hialurónico

Descrita por primera vez en 1811 en hongos, la quitina es un carbohidrato insoluble de cadena larga que consiste en unidades repetidas de N-acetilglucosamina en enlaces beta 1-4. Es la segunda molécula más abundante en la biosfera, justo después de la celulosa, y está presente en linajes eucarióticos tan antiguos como las levaduras. Se encuentra como componente esencial de exoesqueletos de invertebrados y paredes celulares en hongos. Es translúcida, maleable, resistente y un material sorprendentemente durable. Durante el proceso de evolución, la quitina “desapareció” del árbol filogenético metazoario (Figura 1) en los Deuterostomados y aún no se ha probado su existencia en Equinodermos (Tsurkan et al., 2021).

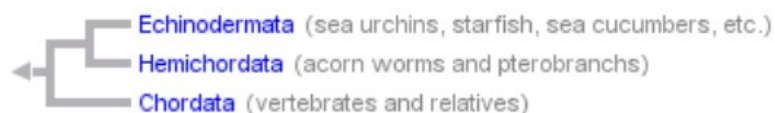


Figura 1. Super filo de los Deuterostomados. Recuperado de Tree Of Life Web Project Nov 2021.

1 Biopolímero de gran interés por su versatilidad y amplias posibilidades de uso.

Por otro lado, el ácido hialurónico (AH) es un carbohidrato y biopolímero que consiste en unidades alternadas de beta 1-4 N-acetilglucosamina y ácido glucurónico unido por un enlace beta 1-3. Comparte similitudes estructurales con la molécula de quitina, pero con propiedades físicas diferentes, además de que parece haber divergido relativamente tarde comparado con otros glicosaminoglicanos (GAG's). La molécula de AH permite el paso de agua de manera libre, es capaz de formar geles en fluidos acuosos como en los ojos de los vertebrados, absorber energía mecánica en forma de golpes en el líquido sinovial de las articulaciones y protege de compresión vascular en el cordón umbilical durante etapas del embarazo. Así mismo, es el principal constituyente de la matriz celular de células somáticas normales y malignas. AH está directamente involucrado en la progenie de células somáticas mientras pasan por distintos estados de diferenciación, más específicamente cuando comienzan el proceso de migración desde las células madre hasta su destino final como parte de algún sistema u órgano. Además confiere movilidad a las células al interactuar con receptores del citoesqueleto y ayuda a abrir espacios entre tejidos para la migración de estas (Stern, 2017).

2.1.1 Síntesis de quitina y ácido hialurónico

La quitina es sintetizada por la quitina sintasa (QS), miembro de la familia 2 de glicosiltransferasas, con regiones transmembranales que varían en número (hasta 17) por ambas terminales y un dominio catalítico central conservado de acuerdo a estudios hechos en hongos e insectos. La formación de quitina incluye tres pasos: la parte citoplasmática del dominio catalítico forma el polímero añadiendo uridina difosfato N-acetilglucosamina (UDP-GlcNac) al creciente oligosacárido; el cual es translocado a través de la membrana y liberado al espacio extracelular; finalmente las unidades de polímeros instantáneamente se re-acomodan en su forma de microfibras cristalinas (Shi et al., 2020).

Los tres miembros de la familia de sintasas de ácido hialurónico HAS1, HAS2 y HAS3 (que ocurren en vertebrados generalmente) son capaces de sintetizar polímeros de AH con una estricta adición alternada de dos azúcares. Estas dos azúcares son UDP-ácido glucurónico y UDP-N-acetilglucosamina, que

son agregadas en el extremo reductor del polímero en formación. Las enzimas tienen un alto nivel de homología funcional y estructural: están incrustadas en la membrana plasmática, y el producto resultante es expulsado al espacio extracelular conforme es sintetizado, lo que permite el crecimiento ininterrumpido del polímero, lo cual sería complicado de hacer en el retículo endoplasmático o el aparato de Golgi (Csoka & Stern, 2013).

Lado a lado, podemos destacar algunas similitudes. Todos los GAG's de los vertebrados son sintetizados via el aparato de Golgi añadiendo una proteína núcleo², a excepción del AH y quitina, que son proteínas sin núcleo y son sintetizadas en la membrana celular. Las sintasas de quitina usan una sola molécula de UDP-azúcar, mientras que las sintasas de AH evolucionaron para utilizar un mecanismo bifuncional que alterna la adición de UDP-azúcar (UDP-ácido glucurónico y UDP-N-acetilglucosamina). Ambos son polímeros monótonos sin modificaciones post-síntesis y, estructuralmente, AH se parece más a la quitina que a cualquier otro GAG (Figura 2)(Csoka & Stern, 2013; Stern, 2017).

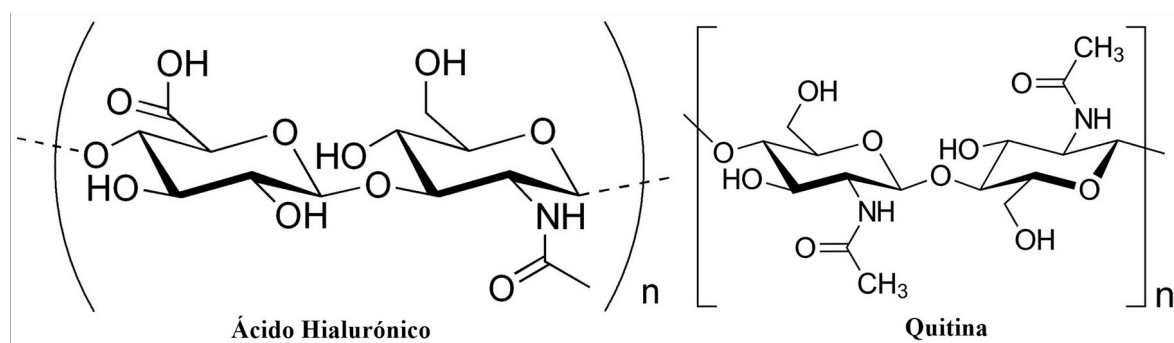


Figura 2. Moléculas de Quitina y Ácido Hialurónico

2.1.2 Evidencias de su parentesco

Es muy probable que la QS sea un precursor ancestral de las sintasas de AH. Ambos AH y quitina son polímeros monótonos sin modificaciones post-síntesis,

² Proteína central que define en buena parte la estructura o función de un complejo proteico.

además de que las enzimas sintasas están ambas imbuidas en la membrana plasmática. Es posible que la membrana, que fue el sitio ancestral de la QS, haya sido el lugar de la subsecuente evolución de la sintasa de AH. Mientras que otros GAG's como condroitina sulfato, queratán sulfato, heparina y heparina sulfato son sintetizados en el aparato de Golgi, y reciben modificaciones post-síntesis (Csoka & Stern, 2013; Stern, 2017).

Las quitinasas degradan y actúan directamente sobre los polímeros de quitina como un regulador de su síntesis. En insectos y crustáceos las quitinasas ayudan en el proceso de muda de exoesqueletos y otras formas de ecdisis. En algunos microorganismos ayudan a la digestión de nutrientes, siendo las bacterias una de las mayores fuentes de quitinasas con enorme potencial industrial. Dependiendo del organismo y función biológica, las propiedades químicas específicas de las quitinasas son las que más cambian han cambiado en el proceso de evolución. La mayoría de las quitinasas en vertebrados son miembros de la familia de glicohidrolasas 18 enzimas que poseen un interesante paralelismo con la familia de hialuronidasas, las cuales degradan ácido hialurónico en vertebrados (Stern, 2017).

En un estudio *in vivo* se logró que un organismo en teoría incapaz de sintetizar HA, como lo es *Drosophila melanogaster*, lo produzca solo con la adición de una proteína de sintasa de ácido hialurónico 2, proteína hallada en mamíferos y otros vertebrados (Takeo et al., 2004).

Actualmente, la mayoría de la información disponible sobre quitina sintasas y su función viene en buena parte de estudios hechos en insectos y hongos. Y si consideramos la gran diversidad de QS's en metazoos, la regulación y funciones de la síntesis de quitina podría ser mucho más compleja. En estudios recientes se reveló que en el cefalocordado *amphioxus* y vertebrados basales existen sintasas de quitina, y son utilizadas para sintetizar quitina endógena. *Amphioxus* podría ser utilizado como modelo base para estudios sobre el entedimiento de la evolución de QS en grupos hermanos a los vertebrados ya que representa el linaje basal entre los tres subfilos de cordados (cefalocordados, urocordados y vertebrados), por lo que ocupa una posición pivotal en la transición de invertebrados a vertebrados. *Amphioxus* es el único cordado que posee tanto genes que codifican para la familia de quitina

sintasa I (la más antigua hallada en poríferos, cnidarios y coanoflagelados) y II (donde hallamos la mayor parte de lofotrocozoos, todos los ecdisozoos, urocordados y vertebrados), con hasta 11 genes que codifican para la familia I y un solo gen para la familia II. Lo que significa que es el único metazoo con una colección de genes de la familia I tan amplia. Y por ello un gran punto de interés evolutivo (Shi et al., 2020).

2.1.3 Datos previos de nuestro laboratorio sugieren que las SH son homólogas a las SQ

En búsquedas anteriores realizadas en BLASTp (proteína-proteína) se pudo constatar que, por ejemplo, al utilizar una secuencia de ácido hialurónico (*Mus musculus*) para buscar coincidencias en *Cryptococcus neoformans* en bases de datos no redundantes, el resultado obtenido son miembros de la familia de quitina sintasas. De manera similar, si utilizamos una secuencia de quitina sintasa (*C. neoformans*) para buscar coincidencias en *M. musculus* se obtienen miembros de la familia de sintasas de ácido hialurónico ("Database resources of the National Center for Biotechnology Information.," 2016).

Pfam, una base de datos de modelos de dominios de proteínas representados por alineamientos múltiples de secuencias y Modelos Ocultos de Markov, puede ser utilizada para buscar a qué familia o dominio proteico pertenece una secuencia de proteína problema. Si hacemos la búsqueda de sintasa de ácido hialurónico de *M. musculus* obtenemos una breve lista de posibles familias a las que pertenezca, dentro de las que encontramos a las familias de Quitina Sintasa y Glicosiltransferasas (Mistry et al., 2020).

RoseTTAFold (<https://robetta.bakerlab.org/>), una herramienta de modelado 3D de proteínas que requiere solo de la secuencia de aminoácidos, es capaz de realizar las posibles estructuras tridimensionales de sintasas de AH y QS. Obtenemos similitudes y diferencias, siendo la más notoria que la sintasa de AH guarda mucho parecido a la de QS, pero más pequeña (Figura 3)(Baek et al., 2021).

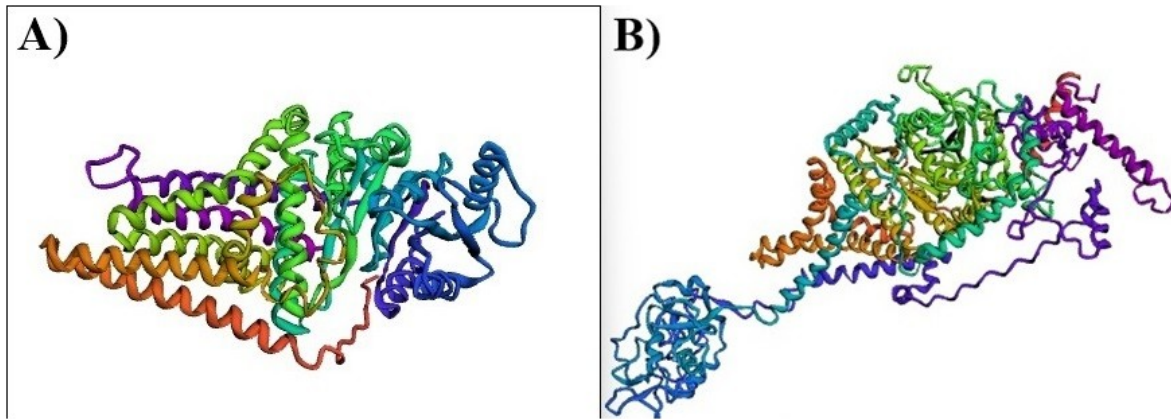


Figura 3: Modelos 3D elaborados con RoseTTAFold A) Sintasa de ácido hialurónico de *M. musculus* (P70312) B) Sintasa de quitina de *S. cerevisiae* (P29465)

2.2 Alineamientos múltiples, familias de proteínas y Modelos Ocultos de Markov

Cuando se trabaja en el análisis de secuencias biológicas podría inferirse que se trata solo de encontrar qué tanta semejanza presenta nuestra secuencia problema con alguna base de datos, pero inevitablemente nos encontramos con problemas del tipo ¿cómo interpretamos estos resultados probabilísticamente? ¿estamos seguros que el mayor puntaje de semejanza es la respuesta correcta? Si quisiéramos también tomar en cuenta posibles eventos de splicing alternativo, adición o eliminación de pares ¿cómo lo hacemos? Aquí es muy útil el uso de los Modelos Ocultos de Markov (HMM por sus siglas en inglés) que son utilizados para problemas de modelación estadística, exploración de bases de datos y alineamientos de familias y dominios de proteínas. Estos mejoran la búsqueda de secuencias distantes emparentadas al convertir un alineamiento múltiple de secuencias en un sistema probabilístico basado en posiciones específicas. De modo que si se tiene en una secuencia problema un aminoácido comúnmente representado en una posición particular, este recibe una puntuación mayor (Eddy, 2004; Krogh, A., Brown, M., Mian, I., Sjolander, K., Hausler, 1994).

Un HMM es un modelo estadístico de estados finitos que permite modelar procesos estocásticos, donde la ocurrencia de los estados está asociada con una distribución de probabilidad y donde las transiciones entre los estados están gobernadas por un conjunto de probabilidades llamadas probabilidades de transición de estados. En un estado particular, una observación se genera también de acuerdo a una distribución de probabilidad. Los estados no son visibles en general y su ocurrencia depende del estado en el instante anterior, de ahí el nombre de Modelo Oculto de Markov (Maldonado, 2012). En biología, proveen de herramientas para construir modelos complejos utilizando descripciones intuitivas. Son el corazón de diversos programas como la búsqueda de homologías lejanas entre genes y sus productos, así como alineamientos múltiples de secuencias e identificación de sitios regulatorios (Krogh, A., Brown, M. , Mian, I. , Sjolander, K., Haussler, 1994).

Usualmente el escenario es una secuencia problema como base sobre la que queremos inferir el rumbo o *state path* que seguirá. El modelo busca inferir qué estado sigue a partir de un estado anterior, pero los posibles caminos son muchos, por lo que para tomar una decisión sobre qué estado es el más realista utilizamos la estadística. El modelo se construye a partir de dos partes esenciales: un rumbo que sigue nuestra secuencia, la transición de estado a estado (*state path*) y; la secuencia observable (*observed path*). Ya que solo contamos con la secuencia observable, el camino a seguir se encuentra oculto, este es el que queremos inferir. El *state path* es una cadena oculta de Markov (Eddy, 2004). El resultado final es que el modelo puede capturar la diversidad observada en un conjunto de proteínas relacionadas evolutivamente.

En el presente trabajo se busca encontrar las posibles relaciones evolutivas entre proteínas sintasa de ácido hialurónico y quitina. Es decir, ver cómo pudo haberse dado el cambio en la QS hacia lo que terminó siendo una sintasa de ácido hialurónico y en qué parte de la secuencia ocurrió.

3. Hipótesis

Se han observado varias similitudes entre las proteínas de sintasas de ácido hialurónico y sintasas de quitina. Se ubican en posiciones similares en la membrana plasmática, su proceso de síntesis es muy parecido, son muy parecidas en su estructura química y se puede inducir la actividad de AH en animales que naturalmente no la producen. Por lo que se espera que las sintasas de ácido hialurónico y de quitina sean homólogas.

4. Objetivos

4.1 Objetivo general

Elaborar un catálogo, identificar y estudiar las relaciones filogenéticas de las proteínas sintasas de quitina y de ácido hialurónico en metazoarios.

4.2 Objetivos particulares

- o Descargar todas las secuencias proteicas de todos los animales de RefSeq (NCBI)
- o Identificar un conjunto de proteínas semilla para un modelo de Markov
- o Crear un modelo oculto de Markov para detectar las sintasas de quitina y ácido hialurónico en animales
- o Usar el modelo para escanear todos los modelos de animales
- o Calcular el árbol filogenético de las sintasas de quitina en animales
- o Estudiar las secuencias proteicas accesorias al dominio de sintasas de quitina

5. Material y Método

5.1 Secuencias proteicas

Las secuencias proteicas serán descargadas directamente de la página del NCBI. Consistirán de Metazoarios cuyo nivel de secuenciación y curación estén a nivel cromosoma y andamio (scaffold) en la base de datos RefSeq.

5.2 Modelos Ocultos de Markov (HMM)

Descargaremos del sitio oficial de Pfam (<http://pfam.xfam.org/>) el HMM de la familia 2 de sintasas de quitina (PF03142). Sin embargo, ya que cuenta con el inconveniente de que las proteínas semilla con las que fue hecho son enteramente de hongos (lo que limita su alcance) utilizaremos un segundo HMM al que denominaremos HMM “específico”, elaborado incluyendo secuencias de animales.

El primer modelo será utilizado como control para obtener resultados de un HMM específico para QS. El segundo consistirá en una modificación del primer modelo, con la adición de 7 proteínas semilla extra. Es decir, constará de las 7 proteínas semilla iniciales, 4 de insectos, atropodos y lanceolados, y 3 de ácido hialurónico de vertebrados (humano, ratón y pez zebra).

5.2.1 Creación del modelo específico

El HMM específico se elaborará con *hmmbuild* (HMMER 3, <http://hmmer.org/>). Estará conformado por secuencias que incluyen hongos, insectos, lanceolados, moluscos, peces y mamíferos (Tabla 1).

Tabla 1. Lista de especies utilizadas para la elaboración del modelo oculto de Markov específico.

Organismo	Clave UNIPROT
<i>Magnaporthe oryzae</i>	O13353 - (CHS4_MAGO7)
<i>Saccharomyces cerevisiae</i>	P29465 - (CHS3_YEAST)
<i>Ustilago maydis</i>	O13394 - (CHS5_USTMA)
<i>Rhizopus delemar</i>	I1BPR1 - (I1BPR1_RHIO9)
<i>Cryptococcus neoformans</i> var. <i>Neoformans</i> serotype D	Q5KDB2 - (Q5KDB2_CRYNJ)
<i>Ustilago maydis</i>	O13395 - (CHS6_USTMA)
<i>Magnaporthe oryzae</i>	G4N0X4 - (G4N0X4_MAGO7)
<i>Drosophila melanogaster</i>	Q8IPN4 - (Q8IPN4_DROME)
<i>Aplysia californica</i>	XP_035825069.1 *
<i>Apis cerana cerana</i>	PBC30924.1 *
<i>Branchiostoma floridae</i>	C3YL08 - (C3YL08_BRAFL)

<i>Homo sapiens</i>	Q92819 - (HYAS2_HUMAN)
<i>Mus musculus</i>	P70312 - (HYAS2_MOUSE)
<i>Danio rerio</i>	Q9DG41 - (Q9DG41_DANRE)

** Especies y secuencias recuperadas de Genbank, dada su ausencia en UniProt.*

5.3 Búsqueda de secuencias

La búsqueda de secuencias con los modelos ocultos de Markov se realizará en la base de datos que contiene a todos los animales disponibles en el RefSeq con la herramienta *hmmsearch* (HMMER 3).

5.4 Armado de árbol filogenético

Una vez identificadas las SQ y SH de todos los animales, se obtendrá la secuencia cubierta por el modelo y se creará un árbol filogenético utilizando el programa PhyML.

5.5 Exploración de las secuencias accesorias de los dominios de sintasa de quitina

Utilizando el modelo ProtENN (Bileschi et al., 2022) de código libre que permite introducir una secuencia problema y obtener resultados de su posible función en tiempo real utilizando redes neuronales que pueden ser utilizadas desde un navegador web obtendremos las posibles funciones de las secuencias accesorias a los dominios catalíticos de las QS.

6. Referencias

- 1) Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- 2) Bileschi, M. L., Belanger, D., Bryant, D. H., Sanderson, T., Carter, B., Sculley, D., ... Colwell, L. J. (2022). Using deep learning to annotate the protein universe. *Nature Biotechnology* 2022, 1–6. <https://doi.org/10.1038/s41587-021-01179-w>
- 3) Csoka, A. B., & Stern, R. (2013). Hypotheses on the evolution of hyaluronan: A highly ironic acid. *Glycobiology*, 23(4), 398–411. <https://doi.org/10.1093/glycob/cws218>
- 4) Database resources of the National Center for Biotechnology Information. (2016). *Nucleic Acids Research*, 44(D1), D7–19. <https://doi.org/10.1093/nar/gkv1290>
- 5) Eddy, S. R. (2004). Primer: What is a hidden Markov model? *Nature Biotechnology*, 22(10), 1315–1316. Retrieved from <http://www.nature.com/naturebiotechnology>
- 6) Krogh, A., Brown, M. , Mian, I. , Sjolander, K., Haussler, D. (1994). Hidden Markov models in computational biology: Applications to protein modeling. *Molecular Biology*, pp. 1501–1531.
- 7) Maldonado, L. (2012). Los modelos ocultos de Markov, MOM. *Telos*, 14(3), 433–438. Retrieved from <https://www.redalyc.org/articulo.oa?id=99324907003>
- 8) Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., ... Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- 9) Shi, Y., Fan, Z., Li, G., Zhang, L., Yue, Z., Yan, X., ... Huang, S. (2020). The family of amphioxus chitin synthases offers insight into the evolution of chitin formation in chordates. *Molecular Phylogenetics and Evolution*, 143(April 2019), 106691. <https://doi.org/10.1016/j.ympev.2019.106691>
- 10) Sotgiu, S., Musumeci, S., Marconi, S., Gini, B., & Bonetti, B. (2008). Different content of chitin-like polysaccharides in multiple sclerosis and

Alzheimer's disease brains. *Journal of Neuroimmunology*, 197(1), 70–73.
<https://doi.org/10.1016/j.jneuroim.2008.03.021>

- 11) Stern, R. (2017). Go fly a chitin: The mystery of chitin and chitinases in vertebrate tissues. *Frontiers in Bioscience - Landmark*, 22(4), 580–595.
<https://doi.org/10.2741/4504>
- 12) Takeo, S., Fujise, M., Akiyama, T., Habuchi, H., Itano, N., Matsuo, T., ... Nakato, H. (2004). In Vivo Hyaluronan Synthesis upon Expression of the Mammalian Hyaluronan Synthase Gene in *Drosophila* - GoodNotes. *Journal of Biological Chemistry*, 279(18), 18920–18925.
<https://doi.org/10.1074/jbc.M314293200>
- 13) Tang, W. J., Fernandez, J. G., Sohn, J. J., & Amemiya, C. T. (2015). Chitin is endogenously produced in vertebrates. *Current Biology*, 25(7), 897–900.
<https://doi.org/10.1016/j.cub.2015.01.058>
- 14) Tsurkan, M. V., Voronkina, A., Khrunyk, Y., Wysokowski, M., Petrenko, I., & Ehrlich, H. (2021). Progress in chitin analytics. *Carbohydrate Polymers*, 252, 117204. <https://doi.org/10.1016/j.carbpol.2020.117204>
- 15) Varki, A. (1996). Does DG42 synthesize hyaluronan or chitin?: A controversy about oligosaccharides in vertebrate development. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), 4523–4525. <https://doi.org/10.1073/pnas.93.10.4523>