

Estudio bioinformático de la proteína Aurora kinasa A de *Bos Taurus*

Josefina Ormaechea

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.

RESUMEN

En la regulación del ciclo celular, las proteínas Aurora juegan un papel fundamental como reguladores de la segregación de cromosomas y la división celular. Debido a esto, su mal funcionamiento ha sido asociado al desarrollo de células malignas y hay interés en su utilización como blancos terapéuticos. En el presente trabajo se aplicaron diferentes herramientas bioinformáticas para obtener información secuencial y estructural de la proteína Aurora kinasa A, de la especie *Bos taurus* (Uniprot ID: Q2TA06). Esta presenta un dominio kinasa y una región N-ter desordenada. Además presenta un 88.8% de identidad con la proteína ortóloga humana, la cual se utilizó para construir el modelo de la proteína por homología. La estructura de la proteína permite observar un loop de activación y el sitio activo.

PALABRAS CLAVE: *Bos taurus*, Aurora kinasa A, bioinformática

INTRODUCCIÓN

La vaca o toro, cuyo nombre científico es *Bos taurus*, es una especie de mamífero artiodáctilo de la familia Bovidae, cuya alimentación es estrictamente herbívora. Estos mamíferos rumiantes fueron domesticados hace unos diez mil años en el Medio Oriente, y su ganadería se desarrolló progresivamente a lo largo y ancho de todo el planeta. Su importancia económica y ganadera deviene del consumo de su carne, leche y cuero. Argentina se posiciona mundialmente como el sexto productor mundial de carne vacuna, con una producción de 3,025 miles de toneladas en 2019, el 5% de la producción mundial¹.

Las proteínas de la familia Aurora kinasas cuentan con tres miembros en mamíferos: Aurora A, Aurora B y Aurora C. La proteína Aurora kinasa A (ARK-1) es una serina/treonina kinasa que participa en la regulación y progresión del ciclo celular. Se encuentra asociada al centrosoma y los microtúbulos del huso durante la mitosis y juega un rol importante en varios eventos mitóticos, incluyendo el establecimiento del huso mitótico, la duplicación del centrosoma, la separación del centrosoma, el alineamiento de los cromosomas, el check-point del ensamblado del huso mitótico y la citoquinesis.

La búsqueda de la proteína ARK-1 de *Bos taurus* en GenBank permite obtener la siguiente secuencia de 402 aminoácidos (N° acceso NP_001033117, Uniprot ID: Q2TA06), derivada del gen localizado en el cromosoma 13 codificado en 10 exones:

>Q2TA06

```
MDRCKENCISGPKTAVPLSDGPKRVPVAQQFPSQNPVSVNSGQAQRVLCPTNSSQRVPSQAQKLVSIQKPVQTLKQKPPQ
AASAPRPVTRPPSNTQKSKQPQPPAGNNPEKEVASKQKNEESKKRQWALEDFEIGRPLGKGFNGVYLAREKQSKFILALK
VLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYGYPHDATRVYLILEYAPLGAVYRELQKLSKFDEQRTATYITELANALSYC
HSKRVIHRDIKPENLLLSGAGELKIADFGWSVHAPSSRRTTLCGTLDYLPPEMIEGRMHDEKVDLWSLGLVLCYEFVGVGKPPFE
ADTYQETYRRISRVEFTFPDCVPEGARDLISRLKHNPSQRPTLKEVLEHPWIIANSKPSSCQKKESTSKQS
```

Las siguientes características de la proteína ARK-1 en *Bos taurus* y su actividad biológica son inferidas de su homólogo humano, por similitud de secuencia. La proteína es necesaria para el correcto ensamblado del huso durante la mitosis y para la localización de NUMA1 y DCTN1 al

córtex celular durante la metafase. ARK-1 es también necesaria para la activación inicial de CDK1 en los centrosomas. Fosforila numerosas proteínas *targets*, como ARHGEF2, BORA, BRCA1, CDC25B, DLGP5, HDAC6, KIF2A, LATS2, NDEL1, PARD3, PLK1, RASSF1, p53/TP53, TPX2. Regula la actividad despolimerizadora de tubulina de KIF2A. Participa en la formación normal de axones, así como en la formación y estabilización de microtúbulos. ARK-1 es un regulador clave en la vía p53/TP53, particularmente en vías críticas en la transformación a células oncogénicas. Fosforila sus propios inhibidores. Regula los niveles de la proteína anti-apoptótica BIRC5 por supresión de la expresión del adaptador FBXL7 de la proteína ubiquitin-ligasa E3 mediante la fosforilación del factor de transcripción FOXP1.

La sobreexpresión de las proteínas Aurora kinasas ha sido descrita en varias células malignas, señalando su participación y rol como oncogenes en la génesis de tumores. Inhibidores de estas proteínas han sido propuestas y estudiadas como posibles terapias anticancerígenas (Yan et al., 2016). Además, su importancia en el desarrollo del ciclo celular ha sido estudiado en la maduración meiótica de oocitos en varias especies, tales como *Xenopus laevis*, *Mus musculus* y *Bos taurus*, brindando información acerca de los procesos bioquímicos involucrados en la formación de oocitos maduros.

El objetivo del trabajo es aplicar diferentes herramientas bioinformáticas para obtener información secuencial y estructural de la proteína Aurora kinasa A, de la especie *Bos taurus*.

MÉTODOS Y RESULTADOS

PRIMERA PARTE: Análisis secuencial

La búsqueda de la proteína ARK-1 de *Bos taurus* en UniProt permite obtener la información anotada para la misma en diferentes bases de datos. Se encuentran anotados los residuos 143, 162 y 274 como sitios de unión a ATP y el residuo 256 como aceptor de protón, perteneciente al sitio activo. También se encuentran anotadas las regiones 210-213 y 260-261 como regiones de unión a nucleótidos fosfato. Las anotaciones de Gene Ontology (GO) molecular function para la proteína ARK-1 son: unión a ATP, actividad serin treonin kinasa, unión a ubiquitina ligasa. Se encuentra regulada por modificaciones post traduccionales, fosforilación y desfosforilación, principalmente en el residuo Thr-288 a lo largo del ciclo celular.

Se buscó la proteína en la base de datos Pfam para evaluar la presencia de dominios. Los parámetros establecidos evalúan secuencias con un valor de corte de E-value de 1.0. Se encontró un dominio significativo, el dominio kinasa, desde el aminoácido 133 al 383, indicado en la figura 1. Se trata de un dominio que posee 12797 arquitecturas anotadas.



Figura 1. Esquema que muestra la predicción de dominios en la proteína ARK-1 de *Bos taurus* en Pfam. Se encontró un solo dominio estadísticamente significativo, el dominio kinasa, en la región 133 a 383.

La búsqueda de la presencia de segmentos transmembrana en el servidor TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) no arrojó ningún segmento, lo que es esperable para la

proteína ya que se encuentra localizada en el centrosoma, el huso mitótico, centriolos y proyecciones neuronales, según las anotaciones en UniProt y de acuerdo a la función biológica de la proteína.

Utilizando el programa Parcoil Scoring Form en la página <http://cb.csail.mit.edu/cb/paircoil/cgi-bin/paircoil.cgi> se buscaron predicciones a partir de la secuencia de regiones con estructura terciaria coiled-coil, obteniendo como resultado que la proteína ARK-1 no presenta regiones con dicha estructura.

Se utilizó el programa IUPred2A (Prediction of Intrinsically Unstructured Proteins) en la página <https://iupred2a.elte.hu/> para buscar la presencia de regiones desestructuradas en la proteína, con los parámetros predeterminados (Long Disorder). El programa arrojó una región con gran probabilidad de carecer estructura en el N-ter de la proteína, y la región estructurada conteniendo el dominio kinaasa anotado en Pfam entre los aminoácidos 133-383. En la figura 2 se muestra el gráfico obtenido del servidor.

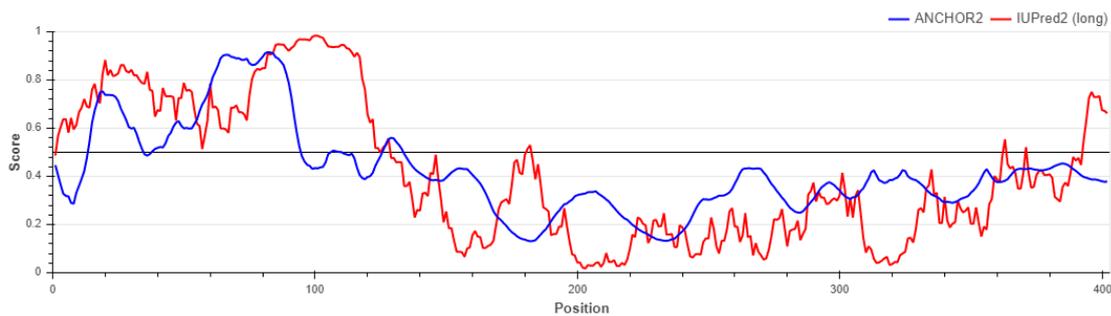


Figura 2. Predicción de segmentos desestructurados de la proteína ARK-1 de *Bos taurus* con el programa IUPred2A. Un score de uno se refiere a un residuo con alta probabilidad de ser desestructurado.

La búsqueda de predicción de la estructura secundaria en el servidor Quick 2D (<https://toolkit.tuebingen.mpg.de/tools/quick2d>) arroja un resultado concordante con la predicción de los segmentos desestructurados, con una región desordenada en el N-ter de la proteína. Luego, la predicción establece segmentos con α -hélice mayoritariamente, y algunas regiones cortas de cadenas β . Al utilizar el servidor PsiPred (<http://bioinf.cs.ucl.ac.uk/psipred/>) para la búsqueda de predicción de la estructura secundaria, el programa arroja un resultado similar en cuanto a las predicciones de α -hélice y cadenas β , pero con la gran diferencia de predecir en la región N-ter largas secuencias de coils. El programa incluye la confianza en la predicción de la estructura secundaria, y esta resulta baja para el N-ter, hasta aproximadamente el aminoácido 75, por lo que podría tratarse de una región desestructurada. El gráfico arrojado por el PsiPred se muestra en la figura 3.

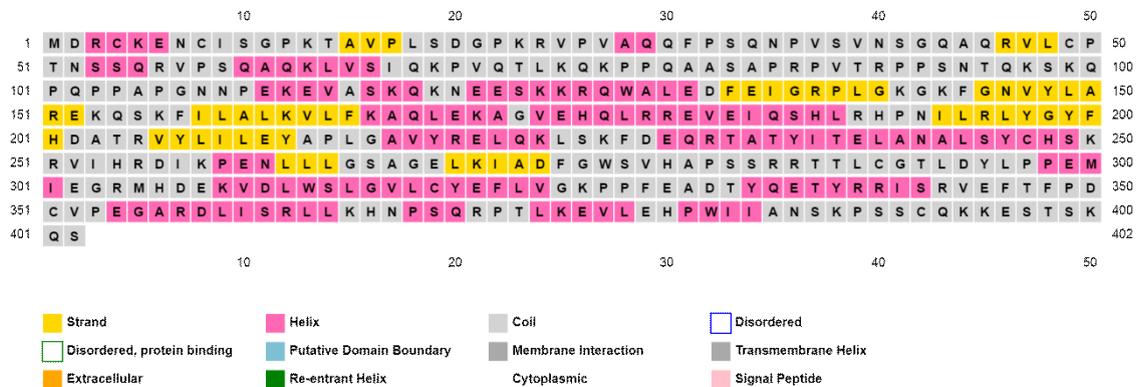


Figura 3. Figura que muestra la predicción de la estructura secundaria a partir del programa PsiPred. En rosa se encuentran los aminoácidos formando parte de una α -hélice y en amarillo aquellos involucrados en cadenas β . En gris, se predicen los residuos participantes en coils. La predicción hasta el aminoácido 75 posee baja confianza (no mostrado).

El análisis de la secuencia en MobiDB predice que el extremo N-ter de la proteína es desordenado, desde el aminoácido 1 al 126 (figura 4). El programa también otorga la predicción consenso del desorden, en la cual esta región no está incluida. Sin embargo, el MobiDB-lite es el único predictor que aparece, por lo que resulta raro que el consenso no incluya la región desordenada predicha por este.

Se realizó una búsqueda de proteínas homólogas utilizando BLAST en NCBI con los parámetros word size: 6, expect threshold: 0.05, la matriz BLOSUM62, gap penalty: 11 y gap extension: 1 y conditional compositional score matrix adjustment. Estos son los parámetros establecidos por el programa, y se utilizó para recuperar 5000 secuencias de una base de datos de proteínas no redundante. El programa recuperó 5000 secuencias de la base de datos. Se aplicó un filtro de coverage de 70% al 100% para obtener homólogos de la proteína completa. Con este filtro, el programa devuelve 2445 secuencias de las 5000 previas.



Figura 4. Imagen de la predicción de segmentos desordenados en el programa MobiDB.

La secuencia recuperada con menor porcentaje de identidad es de 42.35% de identidad. La secuencia con mayor valor de E-value es de $5e-101$. Por lo tanto considero que las secuencias recuperadas corresponden a proteínas homólogas cercanas. Todas pertenecen al dominio eucariota, y se encontraron homólogos en estramenopilos, algas verdes, amebas, algas rojas, hongos y metazoos.

Se realizó luego un PSI-BLAST utilizando los parámetros predeterminados, iguales a los parámetros del BLAST salvo por un word size de 3, para recuperar 1000 secuencias. Se aplicó un filtro de coverage a partir del 70%, y con las 919 proteínas se realizó un Profile y una búsqueda de 1500 secuencias con un valor umbral de E-value de 0.005. El programa devolvió 1342 secuencias. Al realizar otra iteración del nuevo profile construido frente a la base de datos de proteínas no redundantes y solicitar la recuperación de 5000 secuencias, el programa devuelve 3789 secuencias con coverage mayor al 70%. Si comparamos este valor con las 2445 secuencias halladas con el BLAST, evaluamos que se recuperaron más proteínas homólogas. Sin embargo, el porcentaje de identidad menor encontrado es de 48.96%, por lo que no se recuperaron homólogos lejanos. Por otro lado, la realización del primer profile con 919 proteínas puede haber introducido un sesgo hacia las proteínas más similares.

Para evaluar homólogos más lejanos, se realizó un PSI-BLAST contra una base de datos de proteínas clustereadas, lo que permite crear perfiles con mayor diversidad y por lo tanto más representativos de las diferencias entre las proteínas homólogas. Se utilizó el programa en MPI Bioinformatics Toolkit, la base de datos clustereada no redundante nr50_12_Apr predeterminada por el sitio, la matriz BLOSUM62, 3 iteraciones, un umbral de E-value para la inclusión en los perfiles de $1e-8$, un umbral de E-value para el reporte de resultados de $1e-3$ y 5000 como máximo de secuencias recuperadas. El programa recuperó 5000 secuencias de la base de datos. El E-value mayor es de $8.18e-109$ y corresponde a la secuencia titulada "hypothetical protein FGO68_gene235 [*Halteria grandinella*]", la especie corresponde a un protozoo ciliado.

Para evaluar la existencia de homólogos fuera del dominio Eukarya se realizó un BLAST en la página del NCBI con una base de datos de proteínas no redundante y se excluyeron aquellas secuencias en el dominio eucariota. Se utilizaron los parámetros predeterminados detallados previamente, y se realizó la búsqueda para 1000 secuencias. Al filtrarlas para un coverage mayor a 70%, el programa devolvió 69 secuencias. De estas secuencias, solo 15 pertenecen a organismos, mientras que el resto son construcciones sintéticas.

Todas las secuencias derivan de ensayos de ensamblado de metagenoma (MAG: Metagenome assembled genome) a partir de diversas muestras. Las secuencias son entonces proteínas hipotéticas anotadas a partir de bases de datos de proteínas procariotas, y figuran pertenecientes a los siguientes géneros u especies: *Clostridia*, *Waddilaceae*, *Planctomycetes*, *Candidatus Dadabacteria*, *Candidatus Riflebacteria*, *Thermoleophilia*, *Candidatus Latscibacteria*, *Candidatus Brocadiae*, *Gemmatimonadetes*, *Acidobacteria*, *Candidatus Eisenbacteria*.

Con el objetivo de evaluar el alineamiento de proteínas homólogas a ARK-1 a lo largo del árbol de la vida y los residuos conservados entre ellas se eligieron secuencias distribuidas en diferentes ramas del árbol y se preparó un archivo de texto del tipo FASTA para realizar posteriormente un análisis de alineamiento múltiple global. Se escogieron al azar las siguientes secuencias con más del 70% de coverage: 5 secuencias provenientes de bacterias, 2 secuencias de algas verdes y 2 de algas rojas, 6 secuencias del reino Plantae, 6 secuencias del reino Fungi y 10 secuencias pertenecientes a animales. Se realizó un alineamiento múltiple de secuencia en el programa T-Coffee. El alineamiento muestra que entre todas las secuencias, el dominio kinasa se encuentra altamente conservado mientras que las regiones N-ter y C-ter de las proteínas poseen mayores divergencias. Las proteínas alineadas probablemente no posean la misma función biológica, especialmente aquellas obtenidas de bacterias, para cuyas secuencias no está anotada la función biológica. Para las secuencias obtenidas de animales, la similitud es mayor y mejor el alineamiento. El dominio kinasa, presente en organismos tan divergentes como bacterias y mamíferos, implica su temprana aparición en el desarrollo de la vida. Por lo tanto se podrían estar recuperando proteínas que no pertenecen a la subfamilia de Aurora. En la figura 5 se muestra un fragmento del inicio de la región del dominio kinasa alineada por el T-coffee.



Figura 5. Alineamiento múltiple de secuencia en T-coffee, incluyendo secuencias de bacterias, animales, plantas, hongos y algas.

SEGUNDA PARTE: Análisis estructural y filogenético

Se escogió la proteína 1MQ4, código PDB correspondiente a la estructura cristalizada de la proteína Aurora A de *Homo sapiens* como template para el modelado por homología de la proteína Aurora A de *Bos taurus*. La proteína humana de 403 aminoácidos presenta un 88.8% de identidad con la proteína ortóloga bovina (358/403 aminoácidos). La estructura PDB elegida como molde no presenta mutaciones, fue cristalizada por el método de difracción de rayos X y posee una resolución de 1.90 Å (Nowakowski et al., 2002). La estructura cristalizada de la proteína abarca desde el aminoácido 126 al 388, que incluye el dominio kinasa. De los 263 aminoácidos cristalizados, el porcentaje de identidad es de 96,6% (254/263). Se incluye el alineamiento entre ambas secuencias en la figura 6.

```

Bos taurus 126 RQWALEDFEIGRPLGKGFQNVYLAREKQSKFILALKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYG 198
Homo sapiens 126 RQWALEDFEIGRPLGKGFQNVYLAREKQSKFILALKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYG 198

Bos taurus YFHDA TRVYLILEYAPLGAVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLSAGELK 271
Homo sapiens YFHDA TRVYLILEYAPLGVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLSAGELK 271

Bos taurus IADFGWSVHAPSSRRRTTLCQTLDYLPPEMIEGRMHDEKVDLWSLQVLCYEFVLVGGKPPFEADTYQETYRRISR 344
Homo sapiens IADFGWSVHAPSSRRRTTLCQTLDYLPPEMIEGRMHDEKVDLWSLQVLCYEFVLVGGKPPFEANTYQETYKRISR 344

Bos taurus EFTFPDCVPEGARDLISRLLKHNPSQRPTLKEVLEHPWITAN-S 387
Homo sapiens EFTFPDFVTEGARDLISRLLKHNPSQRPMLEVLEHPWITANSS 388

```

Figura 6. Alineamiento entre la secuencia de aminoácidos cristalizada del gen Aurora A de *Bos taurus* y *Homo sapiens*. Se indica con una flecha las posiciones no conservadas.

El modelado por homología se realizó por el programa Modeller a partir del alineamiento entre ambas secuencias proteicas. La calidad del modelo se analizó por el DOPE score, obteniendo el mejor modelo entre los 10 generados el valor de -33556.46094. El análisis del modelo en el servidor PROSA generó los gráficos mostrados en la imagen 7 A) y B), indicativos de la calidad del modelo. El Z-Score del modelo es de -6.8, indicando que se encuentra dentro del rango para modelos de proteínas del mismo tamaño en la base de datos PDB. En el gráfico de energía por residuo se observa un fragmento que posee más energía. Al graficar el valor DOPE por posición del modelo se observa un pico en la misma región, los residuos con valores más positivos que -0,020 corresponden a las posiciones 282-289 (figura 7 C).

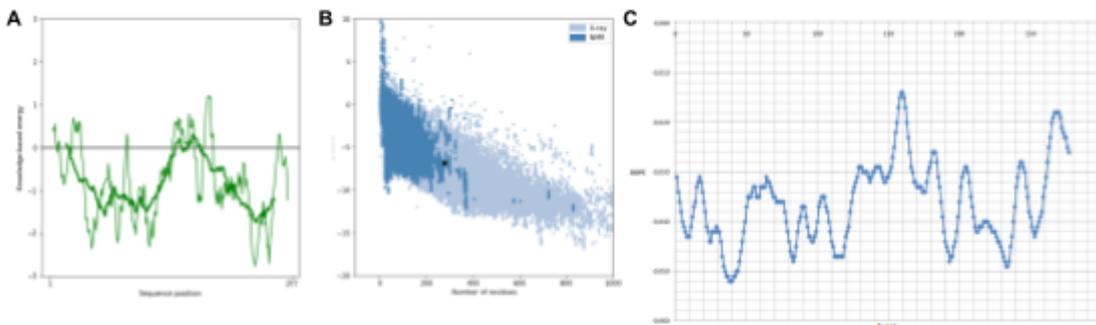


Figura 7. Figuras que indican la calidad del modelo estructural. A) Gráfico generado en PROSA de la energía de cada residuo por posición. Se observa una región levemente positiva. B) Gráfico que grafica el Z-score del modelo en un gráfico conteniendo los Z-scores de los modelos contenidos en PDB. El modelo generado se encuentra dentro de los valores esperados para modelos de la misma longitud. C) Gráfico del DOPE score por posición. Se observa un pico en la misma región del gráfico A, entre las posiciones 282 y 289. Además se observa un pico en el extremo C-ter de la proteína.

Al visualizar el modelo (figura 8) se observa que la región de mayor energía corresponde al loop de activación (Nowakowski et al., 2002).

Se utilizó el programa MEGA X para realizar un árbol filogenético entre varias secuencias de proteínas de diversos organismos recuperadas a partir de un BLAST contra la base de datos del NCBI. Además se eligieron aquellas secuencias anotadas como Aurora A y que representarán una amplia variedad de organismos. El alineamiento múltiple por el algoritmo MUSCLE revela que el dominio kinasa se encuentra bien conservado y el N-ter varía entre las especies, observándose grandes regiones con gaps. Por lo tanto, el N-ter se excluyó para el análisis filogenético, utilizando la región más conservada de la proteína para el análisis posterior, siguiendo el trabajo realizado por Brown et al (2004). Se evaluaron diferentes modelos de evolución de proteínas para el alineamiento realizado por MUSCLE por el método de Maximum Likelihood (ML). El modelo LG+G resultó el mejor modelo, dado por el BIC Score más bajo. Luego se estimó la filogenia utilizando el método ML con el modelo óptimo LG+G con 4 categorías discretas de velocidad de cambio

evolutivo. Se muestra en la figura 9 el árbol inferido a partir de 150 réplicas por el método de Bootstrap no paramétrico.

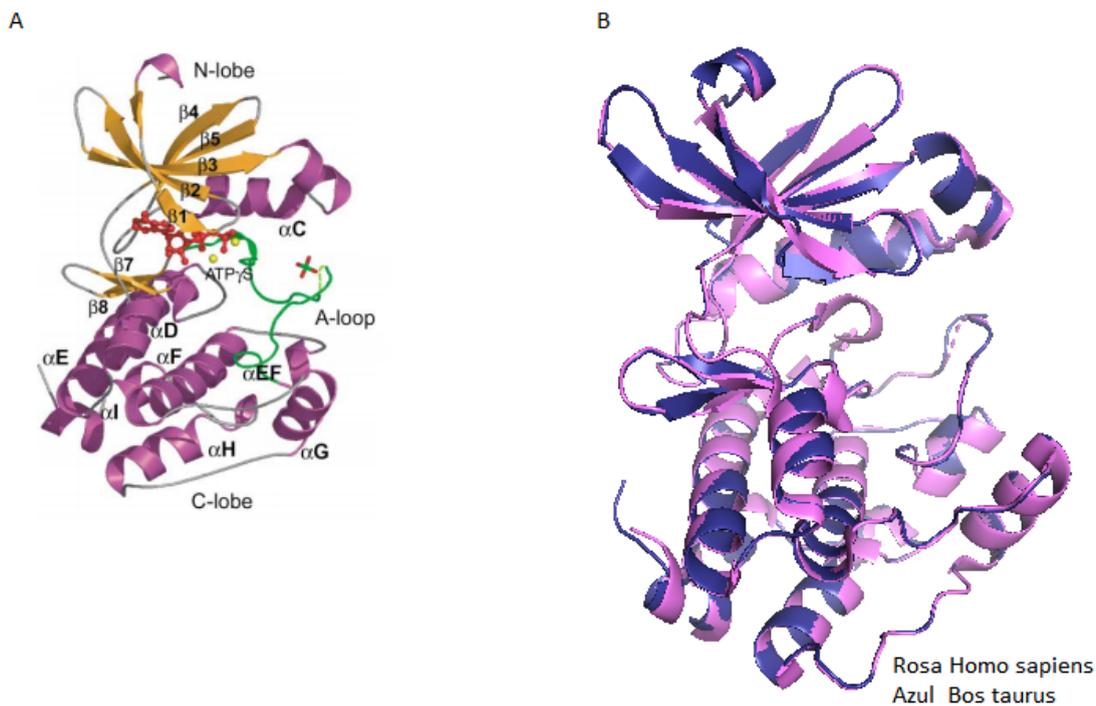


Figura 8. A) Estructura de Aurora kinasa A de *Homo sapiens*. Se indican los α hélices en rosa, las hojas plegadas β en amarillo, el loop de activación en verde, los iones Mg^{2+} como esferas amarillas y el fosfato unido rojo y verde. Imagen obtenida y adaptada de Nowakowski et al. 2002. B) Imagen del modelo generado de la proteína Aurora kinasa A de *Bos taurus* y la estructura molde de la misma proteína de *Homo sapiens* alineados en PyMOL.

Se observa que en el árbol hay dos grupos separados, uno correspondiente a vertebrados, y otro en el que se agrupan secuencias pertenecientes a plantas e invertebrados. Los hongos quedan agrupados, evolutivamente más alejados del resto. Los genes en el cluster de vertebrados son ortólogos. La baja confianza en las ramas del otro cluster no permiten afirmar sus relaciones evolutivas con certeza. Se resalta una divergencia evolutiva entre las secuencias de ambos grupos. Los resultados son concordantes con los descritos por Brown et al (2004). En este trabajo se menciona que los invertebrados poseen dos proteínas homólogas Auroras, los hongos incluidos en el presente alineamiento tienen una única proteína Aurora, y los mamíferos tres. Además concluyen que las proteínas Aurora A de mamíferos y vertebrados de sangre fría son ortólogas, mientras que las proteínas Aurora B y Aurora C de mamíferos evolucionaron más recientemente a partir de un evento de duplicación en vertebrados de sangre fría.

El análisis del modelo pdb de Aurora A de *Bos taurus* por el programa Evolutionary Trace resulta en la coloración de los residuos según la predicción de su importancia biológica tanto en la secuencia como en el modelo 3D de la proteína. Se observa que la región correspondiente al loop de activación y el túnel se encuentran predichos como regiones de mayor importancia. Se incluye en la figura 10 una imagen de la superficie del modelo generado en PyMOL en donde se resaltó en color magenta el loop de activación, el cual dio la energía más alta por DOPE previamente. Por otro lado, se señalan en naranja los 26 aminoácidos indicados por Brown et al. (2004) que se encuentran en el sitio activo de la proteína Aurora A de *Homo sapiens*. De estos residuos, uno solo difiere entre los ortólogos humano y vacuno, indicado en verde, que corresponde a un cambio de T por A en la posición 217 de la secuencia de *Bos taurus*.

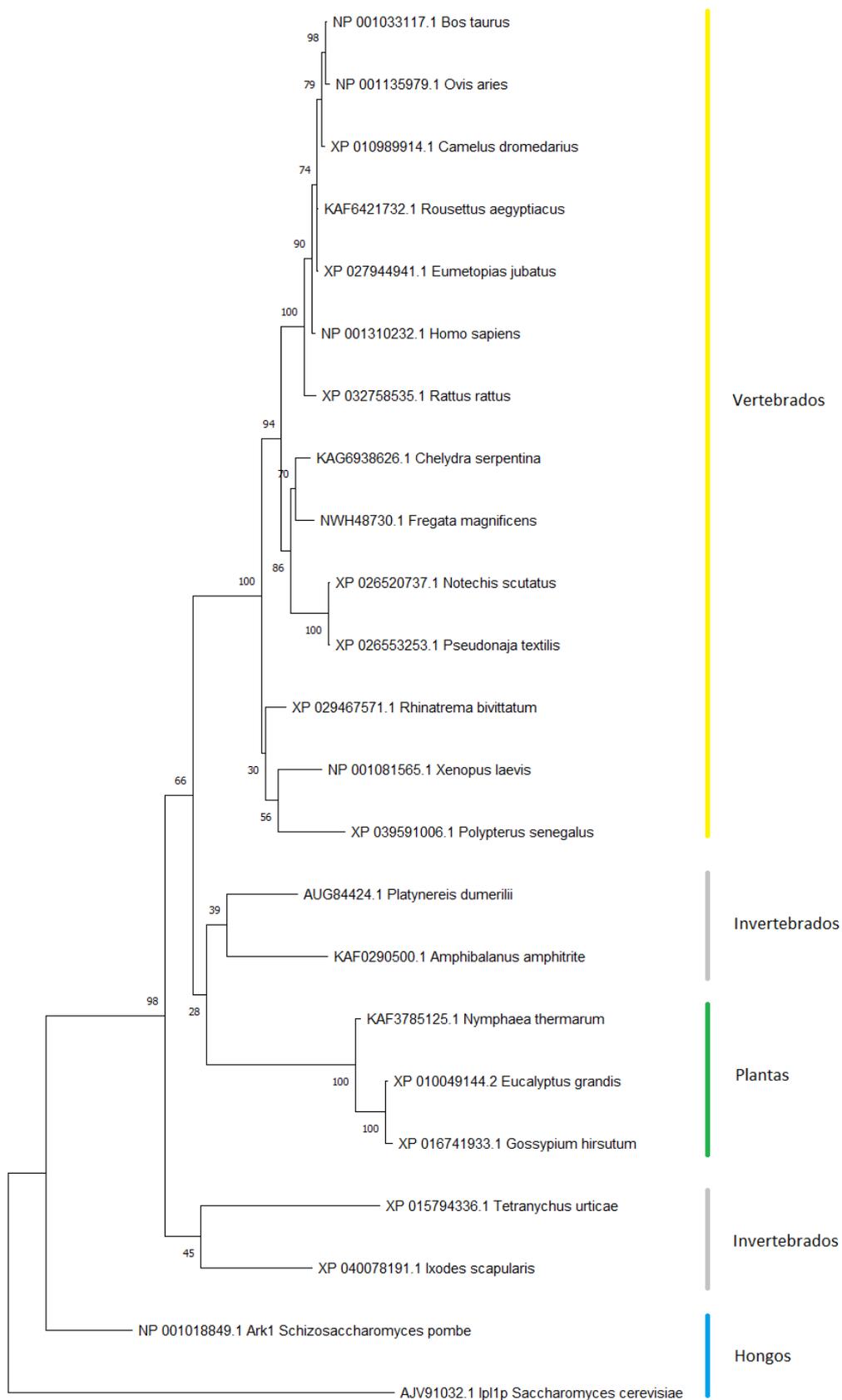


Figura 9. Árbol filogenético inferido a partir del alineamiento múltiple del dominio conservado kinasa de distintas secuencias homólogas de proteínas de Aurora A. Se utilizó el método de Maximum Likelihood con el modelo LG más la distribución gamma con 4 categorías discretas de velocidad de evolución. El árbol inicial fue obtenido automáticamente por los algoritmos Neighbor Joining y BioNJ. Se muestra el árbol a partir de 150 replicados por el método Bootstrap. El análisis involucró 23 secuencias, de las que se observa la especie de la cual derivan y el código de acceso del NCBI.

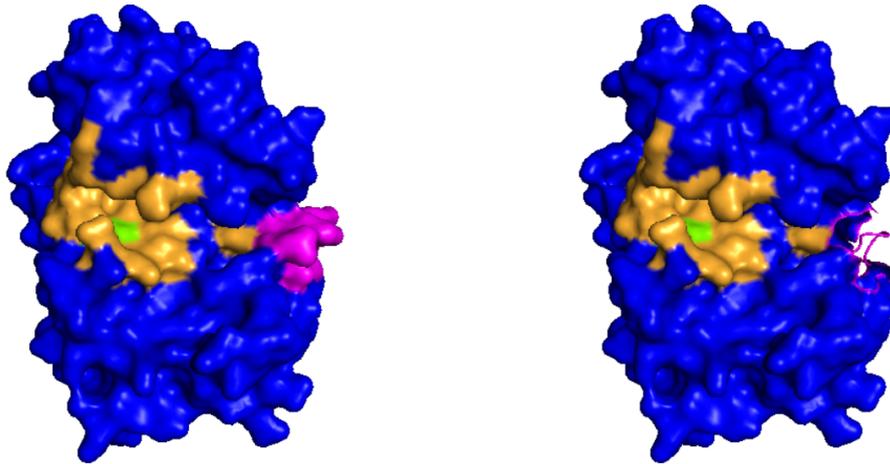


Figura 10. Imágenes de la superficie del modelo generado para el dominio funcional de Aurora A de *Bos taurus* en PyMOL. Se indica en magenta el loop de mayor energía, correspondiente al loop activador. En naranja se resaltan los residuos lindantes al sitio activo para el homólogo humano, según Brown et al (2004). En verde se indica el residuo no conservado entre los homólogos humano y vacuno, que corresponde a un cambio de T por A en la posición 217 de la secuencia para *Bos taurus*.

Al evaluar los mismos 26 residuos lindantes al sitio activo en el MSA utilizado para la construcción del árbol filogenético, se observa que 14 de las 26 posiciones señaladas no se encuentran conservadas para todas las secuencias (figura 11).

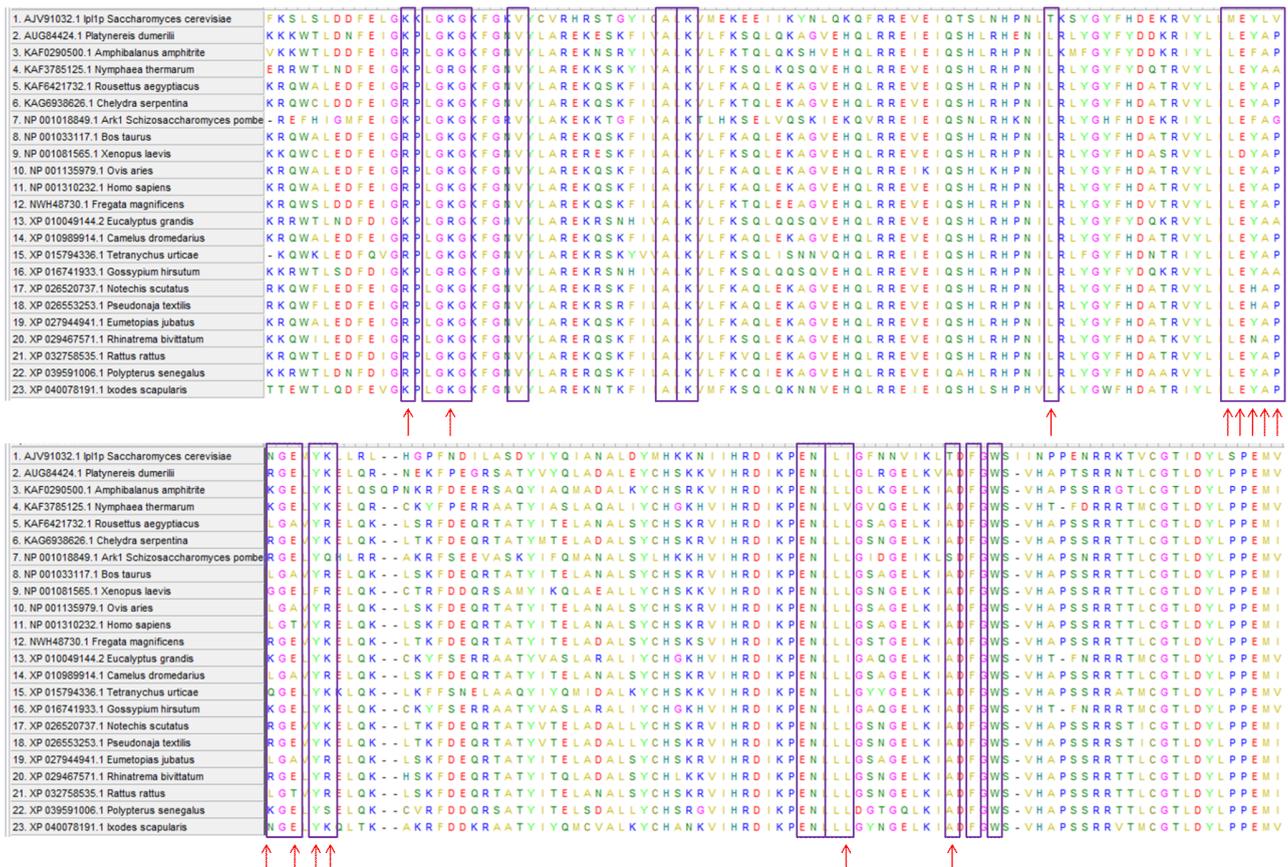


Figura 11. Se muestra una parte del alineamiento múltiple de las secuencias incluidas en el armado del árbol filogenético. Se resaltan con violeta los sitios lindantes al sitio activo según Brown et al 2004, y con una flecha roja aquellos que no están conservados para todas las secuencias. De los 26 residuos, 14 no son idénticos.

CONCLUSIONES Y DISCUSIÓN

La proteína Aurora A de *Bos taurus* presenta un alto porcentaje de identidad (88.8%) con su ortólogo humano. Esta proteína homóloga se encuentra cristalizada y ampliamente estudiada como posible target de drogas anti tumorales.

Aurora A bovina presenta un segmento N-ter desestructurado y un dominio kinasa seguido de una corta región de aminoácidos en el extremo C-ter.

El análisis filogenético reveló que las proteínas Auroras A en vertebrados son ortólogas. Estas presentan diferencias evolutivas con sus homólogas en invertebrados, las cuales se agrupan en el mismo cluster que las secuencias de hongos y plantas.

Con respecto al MSA, resulta llamativo que un alto porcentaje de residuos lindantes al sitio activo no se encuentren conservados entre los distintos organismos (14/26, 54%). Por otro lado, el dominio N-ter de las proteínas es desestructurado y especie específico. Sería interesante realizar un análisis de las regiones N-ter, y un análisis estructural entre el dominio kinasa de proteínas con diferente función biológica versus el dominio kinasa de las proteínas Aurora, de manera de dilucidar relaciones funcionales en estos dominios ampliamente distribuidos.

BIBLIOGRAFÍA

Franco Ramseyer - Emilce Terré (08 de noviembre de 2019) Carne vacuna en el mundo en niveles récord, Argentina aprovecha la mayor demanda global. Bolsa de Comercio de Rosario. <https://www.bcr.com.ar/es/mercados/investigacion-y-desarrollo/informativosemanal/noticias-informativo-semanal/carne-vacuna>

Yan M, Wang C, He B, Yang M, Tong M, Long Z, Liu B, Peng F, Xu L, Zhang Y, Liang D, Lei H, Subrata S, Kelley KW, Lam EW, Jin B, Liu Q. (2016) Aurora-A Kinase: A Potent Oncogene and Target for Cancer Therapy. *Med Res Rev.* 2016 Nov;36(6):1036-1079. doi: 10.1002/med.21399.

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. (2020) *Pfam: The protein families database in 2021*. *Nucleic Acids Research*. doi: 10.1093/nar/gkaa913

Bonnie Berger, David B. Wilson, Ethan Wolf, Theodore Tonchev, Mari Milla, and Peter S. Kim, (1995) "Predicting Coiled Coils by Use of Pairwise Residue Correlations", *Proceedings of the National Academy of Science USA*, vol 92, pp. 8259-8263.

Bálint Mészáros, Gábor Erdős, Zsuzsanna Dosztányi. (2018) [IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding](#) *Nucleic Acids Research*; 46(W1):W329-W337.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. [J Mol Biol. S0022-2836\(17\)30587-9](#).

Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. (2020) Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. [Curr Protoc Bioinformatics. 72\(1\):e108](#). doi: 10.1002/cpb.108.

Notredame, Higgins, Heringa. (2000) T-Coffee: A novel method for multiple sequence alignments. *JMB*, 302 (205-217)

Nowakowski, J., Cronin, C. N., McRee, D. E., Knuth, M. W., Nelson, C. G., Pavletich, N. P., Thompson, D. A. (2002). Structures of the Cancer-Related Aurora-A, FAK, and EphA2 Protein Kinases from Nanovolume Crystallography. *Structure*, 10(12), 1659–1667. doi:10.1016/s0969-2126(02)00907-3

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

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

Brown, J.R., Koretke, K.K., Birkeland, M.L. et al. (2004) Evolutionary relationships of Aurora kinases: Implications for model organism studies and the development of anti-cancer drugs. BMC Evol Biol 4, 39