Implicancias del receptor ionotrópico NMDA subunidad 3A en la esquizofrenia

Iris Quimey López

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

El rápido avance de la bioinformática nos da nuevas formas de analizar enfermedades y trastornos para los que aún no se han podido dilucidar tanto sus causales, como nuevas estrategias de tratamiento más efectivas. Las herramientas bioinformáticas permiten un análisis profundo de las proteínas a nivel secuencial y estructural. Con esta metodología se ha analizado la subunidad proteica 3A del receptor ionotrópico NMDA para encontrar posible información sustancial acerca de su funcionamiento, y variabilidad genética que pueda estar implicada en el desarrollo de enfermedades y trastornos neurodegenerativos, particularmente en la esquizofrenia.

Este informe analiza de forma exhaustiva la secuencia de la proteína, su estructura, su desarrollo evolutivo, y sus posibles funciones biológicas, encontrándose una estructura altamente conservada, con un segmento C-terminal con un alto grado de desorden convirtiéndose en una fuente de variabilidad genética importante, propensa a aumentar la frecuencia de desarrollo de enfermedades neurodegenerativas, y posiblemente implicada en la evolución del cerebro social humano por su papel en el refinamiento sináptico. En cuanto a su función molecular se predice que es un componente intrínseco de la membrana, por ser un receptor de NMDA de canales iónicos activados por glutamato. La proteína consta de al menos cuatro dominios: la familia Lig_chan (receptor ionotrópico de glutamato), la familia Lig_chan-Glu_bd (región con canal iónico ligado L-glutamato, y el sitio de unión a glicina), la familia ANF_receptor (región de unión del ligando de la familia de receptores), y el dominio de proteínas de unión periplásmica tipo 1 y 2. Además se incluye junto al análisis estructural, un modelado de la proteína Q8TCU5 utilizando como template la proteína 7KS0.

PALABRAS CLAVE: GRIN 3A; receptor NMDA; esquizofrenia; primates ; bioinformática.

INTRODUCCIÓN

En la actualidad disponemos de herramientas bioinformáticas que nos permiten averiguar datos de interés sobre una secuencia determinada, predecir su estructura y lograr una aproximación a la predicción de su función biológica. Esto se logra por medio de la aplicación de algoritmos destinados a producir alineamientos por similitud (secuencial y/o estructural) entre las secuencias proteicas acumuladas en la base de datos elegida por el investigador, o bien una combinación de múltiples bases de datos, y analizando estos resultados con criterio biológico. La precisión del alineamiento es resultado directo del algoritmo utilizado (según la matriz en que se base el método, con sus respectivos parámetros), y de la cantidad y calidad de las secuencias de proteínas homólogas utilizadas. En este trabajo utilizaremos diversos métodos bioinformáticos para la caracterización del receptor ionotrópico de glutamato NMDA subunidad 3A de homo sapiens (UniprotID: Q8TCU5). A continuación se muestra un modelo esquemático del complejo proteico NMDA.

Hipótesis sobre la esquizofrenia

La esquizofrenia es un trastorno mental caracterizado por presentar síntomas clasificados como: de primer rango, que incluyen pensamiento sonoro, voces que discuten, experiencia de pasividad somática, influencia, imposición y robo del pensamiento, transmisión de pensamiento, percepciones delirantes, cualquier experiencia que implique voluntad, afectos e impulsos dirigidos; y de segundo rango, que incluyen otros trastornos de la percepción, ideas delirantes súbitas, perplejidad, cambios depresivos o eufóricos, sentimientos de empobrecimiento emocional.

El mecanismo neuroquímico subyacente a la esquizofrenia aún permanece desconocido, pero se considera un trastorno altamente heredable gracias a diversos estudios realizados en familiares cercanos a pacientes que lo padecen. Se han realizado exhaustivas búsquedas de alelos relacionados (preferentemente en genes involucrados en la sinapsis), por ejemplo, se han encontrado polimorfismos de nucleótido único (SNP) en un gen que codifica el receptor de dopamina D2, DRD2 (objetivo de los fármacos antipsicóticos) y muchos genes implicados en las vías de los neurotransmisores de glutamina y la plasticidad sináptica (p. Ej., GRM3, GRIN2A, SRR, GRIA1). En otro estudio se utilizó una matriz personalizada de 1536 SNPs para interrogar a 94 genes candidatos funcionalmente relevantes para la esquizofrenia e identificar asociaciones con 12 endofenotipos neurofisiológicos y neurocognitivos hereditarios, finalmente se observaron asociaciones con endofenotipos para 46 genes de potencial significado funcional, con tres SNPs. Los resultados apoyan colectivamente un papel importante de los genes relacionados con la señalización del glutamato en la mediación de la susceptibilidad a la esquizofrenia (Greenwood Tiffany A. et. al 2011).

La dopamina fue el primer objetivo de los mecanismos de acción de los fármacos antipsicóticos y la capacidad psicoticógena de las sustancias psicoestimulantes desde que se desarrolló la hipótesis dopaminérgica. Actualmente la hipótesis dopaminérgica junto a la hipótesis glutaminérgica son consideradas las más importantes y de mayor base empírica para el estudio de la esquizofrenia y sus espectros (Gurillo Muñoz P. 2016).

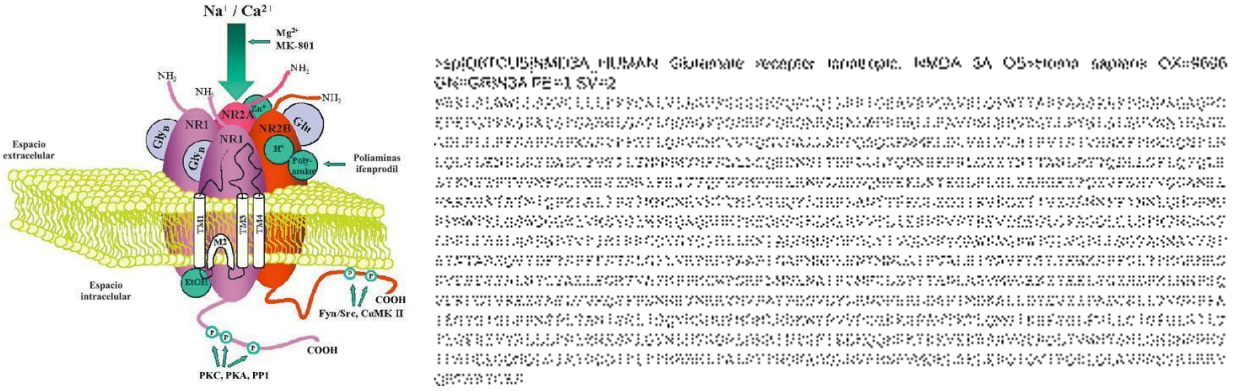


Figura 1 (a). Modelo esquemático del receptor NMDA, y figura 1-B secuencia de la subunidad 3A del receptor NMDA. A la izquierda se representa el ensamblaje heteromérico de 4 subunidades en los receptores NMDA. Cada subunidad tiene cuatro regiones hidrofóbicas, y tres formas de dominio transmembrana (TM1,TM3, TM4). M2 toma la conformación de horquilla dentro de la membrana y forma el poro. Los complejos funcionales del receptor están formados por combinaciones de las subunidades NR1 y NR2, las cuales contienen los sitios de unión de los agonistas, algunos antagonistas y otros moduladores. Fuente: Jozsef Nagy, tipo de subunidades NR2B del receptor NMDA; un blanco posible para el tratamiento de la dependencia del alcohol. Current Drug Targets CNS Neurol Disord., 2004,3(3):169-79. (b) A la derecha se exhibe la secuencia en formato fasta.

En relación a la hipótesis glutaminérgica diversos servidores aportan estudios basados en la búsqueda de genes (OMIM), y estos relacionan el desarrollo de la esquizofrenia con la participación del gen GRIN3A, en

conjunto con los genes señalados anteriormente, en el desarrollo de los trastornos esquizoafectivos (esquizofrenia en combinación con el trastorno bipolar, y la depresión).

Existen diversas hipótesis acerca del desarrollo de este trastorno, entre ellas la de la insuficiente `poda sináptica', o bien el exceso de poda neuronal (De las Matas Martín, y Del Carmen María, 2014).

Entre las hipótesis evolutivas podemos encontrar aquella que relaciona la evolución del cerebro social humano con la esquizofrenia:

La esquizofrenia puede ser una compensación de la evolución de la inteligencia social, la cual conlleva un dramático incremento en la conectividad cortical de los primates. Se sugiere un modelo de dos pasos en el trasfondo evolutivo de la esquizofrenia. El primer paso evolutivo se dio hace unos 5 ó 6 millones de años, hacia una más compleja conectividad inter- e intrahemisférica. Un segundo paso fue hace, aproximadamente, 150.000 años, cuando alguna desconocida mutación incrementó la vulnerabilidad de tales conexiones, lo que podría estar asociado con la evolución de la metacognición y la 'teoría de la mente'. De acuerdo con el autor Burns, la esquizofrenia podría ser un costo de la evolución del cerebro social humano, y, en este sentido, su hipótesis se hallaría entre las que suponen que el trastorno significa una desventaja compensatoria en la evolución del cerebro social. Otro autor propone que un gen que regula la dominancia cerebral está involucrado en el origen de los trastornos psicóticos, y que en este origen podría ser trascendente en la evolución del lenguaje (Altschul, S.F, et. al, 1990).

En este trabajo se realizará una búsqueda de los homólogos tanto cercanos como remotos, con su respectiva distribución taxonómica. Con esta información se pretende predecir la estructura secundaria, los segmentos transmembrana, regiones desordenadas, dominios, además de proporcionar un modelo estructural de la proteína, construir un árbol filogenético, y de ser posible predecir la función biológica de la proteína. Con este análisis se busca enriquecer el conocimiento que se tiene de esta proteína en relación a su función neurotransmisora y los trastornos que están arraigados a su alteración, así como la función biológica de sus homólogos más cercanos, y de ser posible detectar las diferencias secuenciales y/o estructurales que causen la ausencia de estos trastornos en el resto de los primates (en base a las teorías evolutivas de la esquizofrenia y la evolución del cerebro social humano, anteriormente citadas), por lo cual se le considera un trastorno propio de la especie humana.

MÉTODOS Y RESULTADOS

Búsqueda de secuencias homólogas

Para comenzar con la caracterización del receptor ionotrópico de glutamato NMDA subunidad 3A, se realizó la búsqueda de homólogos por Blast (Altschul S.F., 1997). Blast es una herramienta básica de búsqueda de secuencias similares basada en alineación local. El programa compara una secuencia (ya sea nucleotídica o proteica) contra las secuencias contenidas en una base de datos seleccionada (método de comparación secuencia a secuencia), calculando a su vez la significancia estadística (e-value) para cada resultado. Este algoritmo prioriza la velocidad en vez de la sensibilidad de la búsqueda, y se basa en la comparación de K-Tuples. Al realizar la búsqueda con los parámetros por defecto se obtienen 5074 resultados de posibles homólogos. Siendo Blast un algoritmo de alineamiento local, encontrará homólogos tanto de la proteína completa como de sus dominios. Para quedarnos solo con homólogos correspondientes a la proteína completa utilizamos los siguientes filtros y valores para los parámetros: Coverage > 70%, %Identity > 40%, Expected threshold = 100, Matriz: compositional score matrix adjustment, Database: All non-redundant GenBank CDS translations+PDB+SwissProt +PIR+PRF excluding environmental samples from WGS projects.

De esta forma se encontraron 1254 posibles homólogos. Considerando los resultados del alineamiento obtenidos con un query coverage (>70%) y con un cutoff correspondiente a un %ID>30 los alineamientos encontrados son homólogos por tener una distancia evolutiva cercana que puede deducirse con el análisis secuencial, sin necesidad de realizar un análisis estructural.

Adicionalmente, Blast aporta toda la información taxonómica de las proteínas alineadas, el dominio, el género, la especie, orden, etc. Con el análisis taxonómico podemos observar que los hits con mayor significancia estadística, es decir, con mayor score (hasta 2315 score total) pertenecen a la especie *Homo sapiens*, los 47 homólogos más relevantes que le siguen son también pertenecientes a la orden de los primates. Los hits de menor score pertenecen mayormente a diversos vertebrados, especialmente mamíferos. Se obtuvo como homólogo más lejano al receptor ionotrópico del glutamato, NMDA 3A del organismo Crassostrea virginica (%ID de 25,96, e-value=1e-82, y query coverage del 74%).

Con el objetivo de buscar homólogos remotos se realizaron búsquedas utilizando PSI-BLAST (Profile – secuencia) y HMMER (HMM) PSI-BLAST (o Position Specific Iterated BLAST) es un programa muy rápido que realiza un simple BLAST con una secuencia y, a partir de los resultados, construye un perfil o PSSM. Entonces, la siguiente búsqueda la realiza con ese perfil, lo que permitirá encontrar, idealmente, homólogos remotos. Dados los nuevos homólogos se genera un nuevo perfil, que idealmente contendrá mayor cantidad de información y podrá realizar otra búsqueda. Es un proceso iterativo. Con una búsqueda refinada (con parámetro word size=2) se obtuvieron homólogos remotos, donde el resultado con mayor distancia evolutiva encontrado fue el receptor ionotrópico 6 perteneciente al organismo *Diaphorina citri* (%ID de 23,69, e-value=1e-59, y query coverage de 75%).

HMMER proporciona herramientas para crear modelos probabilísticos de familias de dominios de secuencias de proteínas y ADN (HMM) y para usar estos perfiles para anotar nuevas secuencias, buscar en las bases de datos de secuencias de homólogos adicionales y crear alineamientos profundos de múltiples secuencias. Como resultado se encontraron 12.052 posibles homólogos (base de datos: uniprot refprot v.2019_09). Se logró llegar hasta un 25.6% de ID, con un 51.2% de similitud, y con un query coverage en el intervalo de 671-756 (Figura S1).

Predicción de estructura secundaria

Se realizó la predicción de estructura secundaria con diversos programas, incluyendo Quick2D (Gabler F, et. al, 2020), PSIPRED (Buchan DWA, & Jones DT, 2019), y Porter (M.Torrisi, 2019). Por medio de Quick2D (Figura S2) se detectó un péptido señal potencial en su extremo N-terminal, además se predijo que la estructura secundaria estaría compuesta de manera uniforme por segmentos de hélice alfa y hoja beta plegada, y se predijeron cuatro regiones transmembrana (tres de ellas en una región intermedia de la secuencia, y la última hacia el extremo C-terminal). Además se detectaron pequeños segmentos desordenados dentro de la secuencia. Los resultados de PSIPRED, un método de predicción de estructura secundaria que incorpora dos redes neuronales de retroalimentación que realizan un análisis de la salida en función de las matrices de puntuación específicas de posición generadas por PSI-BLAST (Position Specific Iterated – BLAST) se muestran en la figura 4, S3, S4, y S5.

Finalmente, las predicciones obtenidas mediante el programa Porter 5 (M.Torrisi, 2019), compuesto por un conjunto de redes neuronales recurrentes bidireccionales en cascada y redes neuronales convolucionales, concuerdan en gran proporción con la predicción de la estructura secundaria aportada por PSIPRED (Figura S6).

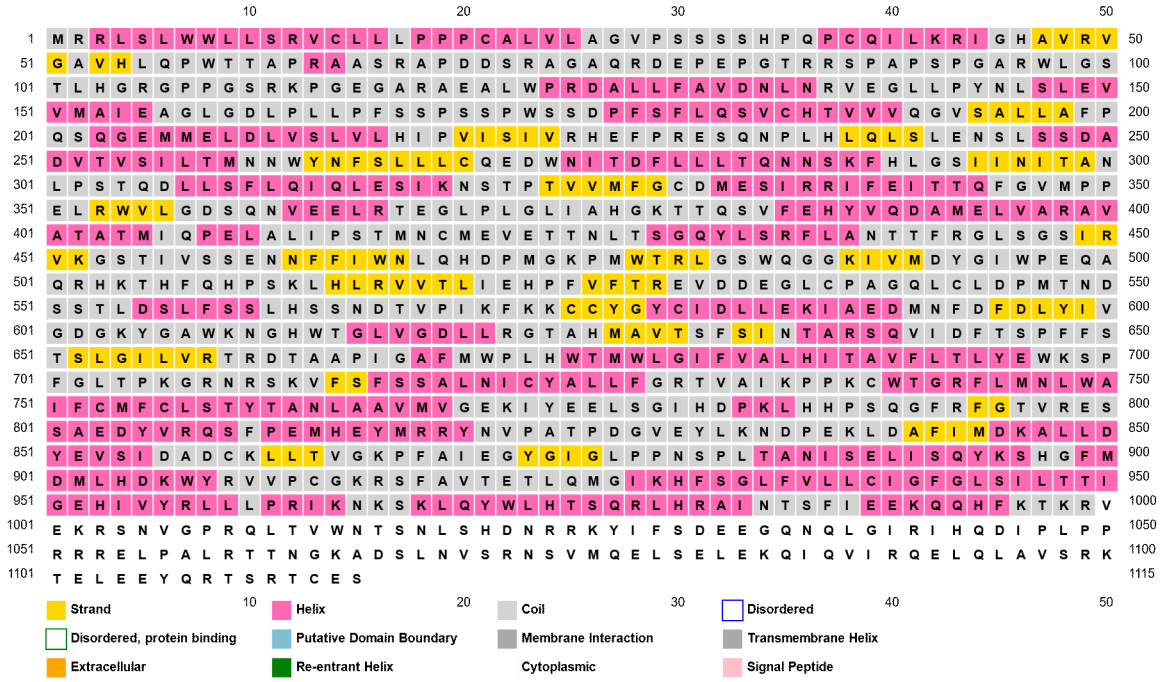


Figura 4. Predicción de estructura secundaria con PSIPRED. La figura representa la estructura secundaria predicha en cada región de la proteína. Se predijo que la proteína presenta una conformación mayormente del tipo coil, con una distribución uniforme de conformaciones tanto hélice alfa como hoja beta plegada, así como también se predice una región citoplasmática hacia el final de la secuencia. No se predijeron péptidos señal.

Regiones de baja complejidad

Las regiones de baja complejidad (Low complexity regions, LCR), son secuencias de proteínas cuya composición de aminoácidos es muy simple. Su expansión descontrolada provoca varias enfermedades humanas, incluyendo la enfermedad de Huntington y otras enfermedades neurodegenerativas y de desarrollo. Sin embargo, son sorprendentemente abundantes en las proteínas, lo que parece paradójico dado su alto potencial patógeno. Por otra parte, los datos experimentales han demostrado que la formación de nuevas LCR, o la modificación de las existentes, puede tener consecuencias funcionales. Pueden ser una importante fuente de variabilidad genética y podrían estar implicadas en los procesos de adaptación, además pueden estar involucradas en la diversificación de la proteína, ya sea proporcionando nuevas secuencias funcionales que modificarán las proteínas existentes o estando involucradas en la formación de nuevas secuencias codificantes en la proteína. Para el estudio de este tipo de secuencias se utilizó SEG: Prediction of Low Complexity Regions (Zhang M, et. al, 2018). Los resultados muestran una gran cantidad de regiones de baja complejidad en el fragmento N-inicial y el C-terminal de la secuencia de la proteína Q8TCU5 (Tabla 1), lo que implica que estas dos regiones son fuentes de alta variabilidad genética, propensas a aumentar la frecuencia de desarrollo de enfermedades neurodegenerativas.

Tabla 1. Resultados obtenidos mediante SEG.

Predicción	**SEG 12 2.2 2.5**	**SEG 25 3.0 3.3**	**SEG 45 3.4 3.75**
Regiones de baja complejidad	1-1; 2-26; 27-26; 61-74; 75-156; 157-181;182-547;1047-1059; 1060-1115	1-2; 3-33; 34-56; 57-128; 129-156; 157-181; 182-547	1-1;2-255;256-1115

Nota: Parámetros por default para la secuencia proteica Q8TCU5, NMD3A_HUMAN Glutamate receptor ionotropic, NMDA 3A, Homo sapiens (GRIN3A).

Predicción de dominios globulares

Las proteínas globulares son proteínas formadas únicamente por aminoácidos, suelen estar compuestas de una sola molécula proteica, o unas pocas moléculas combinadas que se pliegan en forma esférica y forman una estructura más compleja. Se caracterizan por doblar sus cadenas en forma esférica compacta dejando grupos hidrófobos en el core de la proteína y grupos hidrófilos hacia afuera, lo que hace que sean solubles en disolventes polares como el agua. Forman suspensiones coloidales. La mayoría de las enzimas, anticuerpos, algunas hormonas y proteínas de transporte son globulares. Para la predicción de regiones globulares se utilizó GlobPlot (Linding R, et. al, 2003), un servicio web que permite al usuario trazar la tendencia dentro de la proteína para el orden / globularidad y el desorden. Adicionalmente identifica segmentos entre dominios que contienen motivos lineales y también regiones aparentemente ordenadas que no contienen ningún dominio reconocido. Como resultado (Figura S7), se predijeron regiones globulares en los segmentos: 240-596, 626-930, 946-1174.

Predicción de segmentos transmembrana

Para la predicción de segmentos transmembrana se utilizaron los servidores TMHMM (DTU Health Tech, 2017), TMPred (Hoffman & Stoffel, 1993) y Phobius (Lukas Käll, 2007). TMHMM se encarga de predecir las hélices transmembranas en las proteínas, está basado en un modelo oculto de Markov (HMM), donde una de las principales ventajas es que es posible modelar la longitud de la hélice estableciendo límites superiores e inferiores para la longitud de una hélice de membrana. Los HMM son muy adecuados para la predicción de hélices transmembrana porque pueden incorporar hidrofobicidad, sesgo de carga, y longitudes de hélice. Los resultados obtenidos con TMHMM se muestran en la figura 9.

sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	outside	1	674
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	TMhelix	675	694
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	inside	695	713
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	TMhelix	714	733
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	outside	734	747
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	TMhelix	748	770
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	inside	771	933
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	TMhelix	934	956
sp|Q8TCU5|NMD3A_HUMAN	TMHMM2.0	outside	957	1115

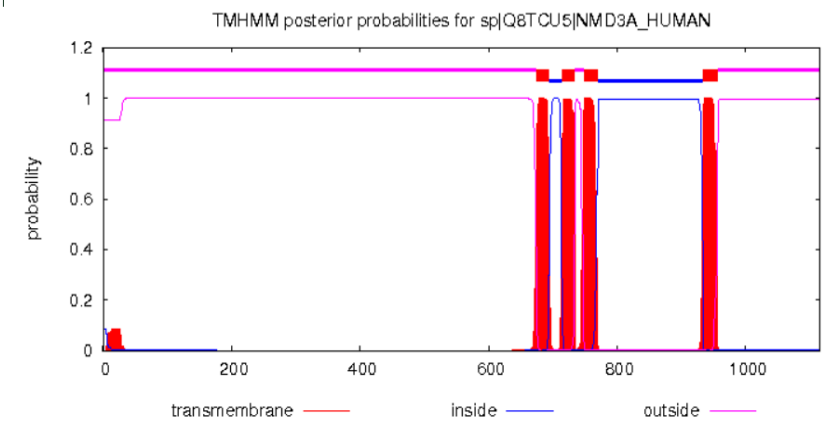
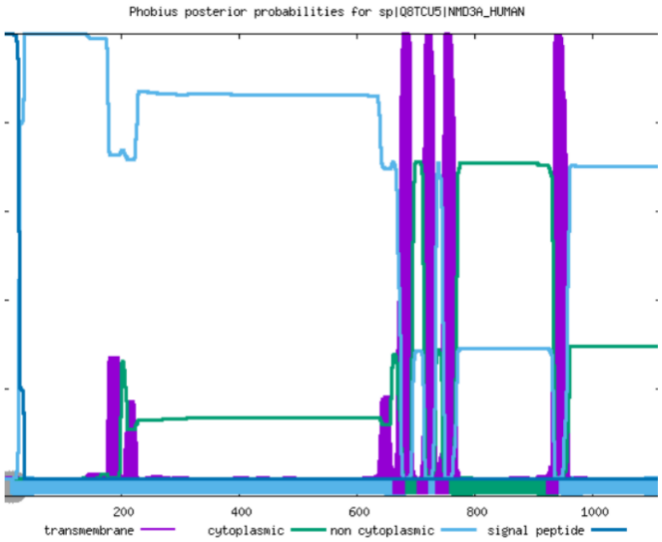


Figura 9. Probabilidad de existencia de hélices transmembrana en la secuencia proteica de Q8TCU5. La leyenda superior al gráfico indica las predicciones de segmentos transmembrana/ citoplasmáticos/ no citoplasmáticos para cada segmento de la proteína, luego el gráfico inferior representa la probabilidad de encontrar regiones transmembrana en determinados segmentos de la secuencia, evidenciándose una mayor probabilidad de encontrar regiones transmembrana en los segmentos 675-694, 714-733, 748-770, y 934-956, lo que podría ser una hélice alfa múltiple. Hay una alta probabilidad de que esta proteína contenga 4 regiones transmembrana.

TMPred es una base de datos de proteínas transmembrana y dominios helicoidales que atraviesan la membrana. TMPred originalmente era una herramienta para analizar las propiedades de las proteínas transmembrana. Se basa principalmente en SwissProt, pero también contiene información de otras bases de datos. TMPred se utilizó con los siguientes parámetros: matriz=MTIDK; Window width: 14,21, 28; Ponderación de las posiciones= no. Este programa indica que hay dos modelos posibles basados en las predicciones de segmentos transmembrana (Figura S8): Modelo 1, con 7 hélices transmembranas (Score=10975); y el modelo 2, con 6 hélices transmembranas (Score=9902).

Adicionalmente se utilizó el servidor Phobius (Figura 10), que sirve para predecir la topología transmembrana y los péptidos señal de la secuencia de aminoácidos de una proteína.




```
ID   sp|Q8TCU5|NMD3A_HUMAN
FT   SIGNAL        1     26
FT   REGION        1      3       N-REGION.
FT   REGION        4     17       H-REGION.
FT   REGION       18     26       C-REGION.
FT   TOPO_DOM     27    674       NON CYTOPLASMIC.
FT   TRANSMEM    675    695
FT   TOPO_DOM    696    715       CYTOPLASMIC.
FT   TRANSMEM    716    734
FT   TOPO_DOM    735    745       NON CYTOPLASMIC.
FT   TRANSMEM    746    769
FT   TOPO_DOM    770    933       CYTOPLASMIC.
FT   TRANSMEM    934    956
FT   TOPO_DOM    957   1115       NON CYTOPLASMIC.
//
```

Figura 10. Predicción de la topología transmembranal. El gráfico se generó calculando la probabilidad total de que un residuo pertenezca a una hélice citoplasmática o no citoplasmática, hélice TM, o péptido señal, sumada en todos los caminos posibles a través del modelo, utilizándose parámetros por default. Aquí uno puede ver posibles hélices TM débiles que no fueron predichas. En la figura se observan al menos 4 segmentos coiled coil hacia el final de la cadena, entre las posiciones que van del 600 a la 800 y de la posición 1000 a la 1115. Esto indica que hay una región citoplasmática en la misma región donde se sugería la existencia de hélices alfa múltiples en TMHMM, es decir, en el segmento desde la posición 675 hasta la 735 y el segmento que va de 770 a la posición 956. Además me indica la posible existencia de una pequeña hélice alfa en el segmento que va desde la posición 4 a la 17. Como conclusión, podemos decir que existen al menos 4 regiones transmembranales (que coinciden con las regiones globulares predichas), y un péptido señal en la región N-inicial (denotada en color azul en la figura).

Predicción de regiones coiled-coil

Para la predicción de segmentos con estructura de coiled-coil se utilizaron los servidores COILS (Lupas, 1991) (Figura 11), Paircoil (Bonnie Berger, 1995) (Figura 12) y Paircoil2 (McDonnell, 2006). COILS es un programa que compara una secuencia con una base de datos de coiled-coils de dos hebras paralelas conocidas y obtiene un score de similitud. Al comparar esta puntuación con la distribución de scores en proteínas globulares y en espiral, el programa calcula la probabilidad de que la secuencia adopte una conformación de hélice.

El programa Paircoil toma tres argumentos: un nombre para la secuencia (opcionalmente), un límite de probabilidad y la secuencia de aminoácidos. El límite de probabilidad determina qué tan estrictamente el programa filtra la secuencia de entrada al detectar la existencia de un dominio de coiled-coil. Se determinó empíricamente, que el valor predeterminado de 0,5 para el límite de probabilidad funciona bien. Por último

se utilizó Paircoil2, que predice el pliegue de coiled-coils a partir de la secuencia utilizando probabilidades de residuos por pares con el algoritmo Paircoil y una base de datos de coiled-coil actualizada.

Por medio del análisis predictivo dado por estos servidores a partir de la secuencia de la proteína Q8TCU5, podemos decir que hay al menos 4 regiones transmembrana en los segmentos 675 al 694, 714-733, 748-770, y el segmento final es una región coiled-coil desde la posición 1000 a la 1115.

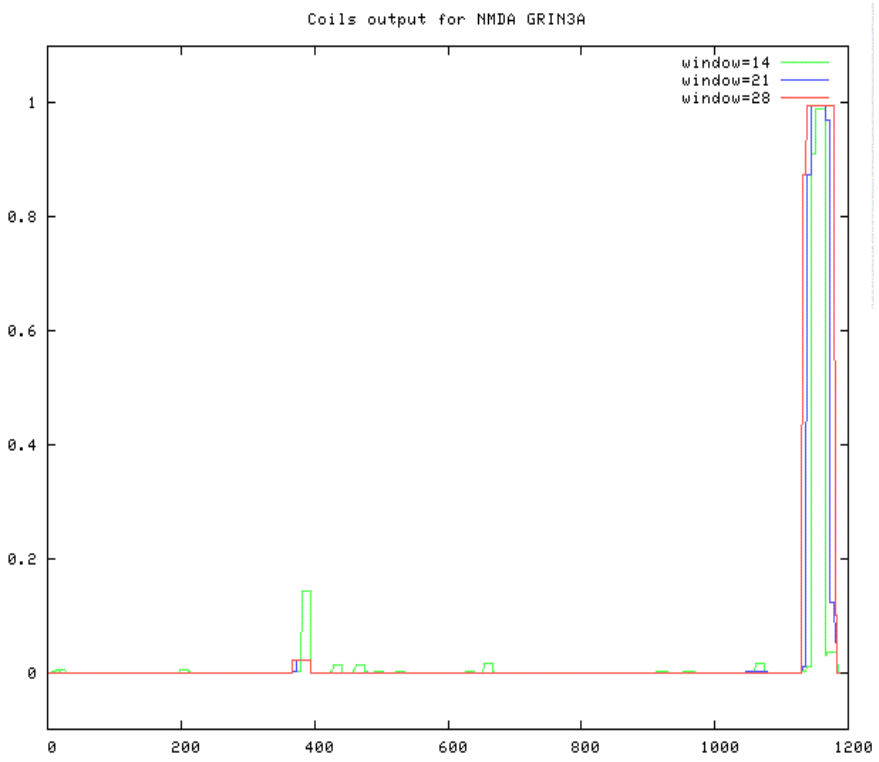


Figura 11. Probabilidad de encontrar una hélice superenrollada, y la probabilidad de formar espirales dobles en la secuencia. En el eje vertical están representadas las probabilidades (scoreadas por MTK, MTIDK, MTK_W y MTIDK_W en ese orden), en el eje horizontal el largo de mi secuencia. Los valores se obtienen con una ventana de escaneo de 21 residuos que detecta los extremos de los segmentos de espirales dobles con más precisión que la ventana de 28. Se observa la presencia de una hélice superenrollada en el segmento final de mi secuencia con un alto nivel de probabilidad. Se utilizaron los parámetros por default: Matriz=MTIDK Ponderación de las posiciones =no NCOILS versión .0 [ISREC-Server].

Predicción de regiones desordenadas

Para la predicción de regiones desordenadas se utilizaron IUPred2A (Bálint Mészáros, 2018) (Figura 13), DynaMine (Figura S9), y MobiDB (Figura S10). IUPred2A es una interfaz web combinada que permite identificar regiones de proteínas desordenadas (no tienen una estructura terciaria bien definida en condiciones nativas) usando IUPred2 y regiones de unión desordenadas usando ANCHOR2.

DynaMine (Elisa Cilia, et. al, 2013) es un predictor de la dinámica de la columna vertebral de las proteínas. Utiliza información secuencial proteica como entrada con un gran potencial para distinguir regiones de diferente organización estructural, como dominios plegados, enlazadores desordenados, glóbulos fundidos y motivos de unión preestructurados de diferentes tamaños. MobiDB (Piovesan D, et. al, 2020) proporciona información sobre regiones intrínsecamente desordenadas (IDR) y características relacionadas de varias fuentes y herramientas de predicción. Los diferentes niveles de confiabilidad y las diferentes características se informan como anotaciones diferentes e independientes.

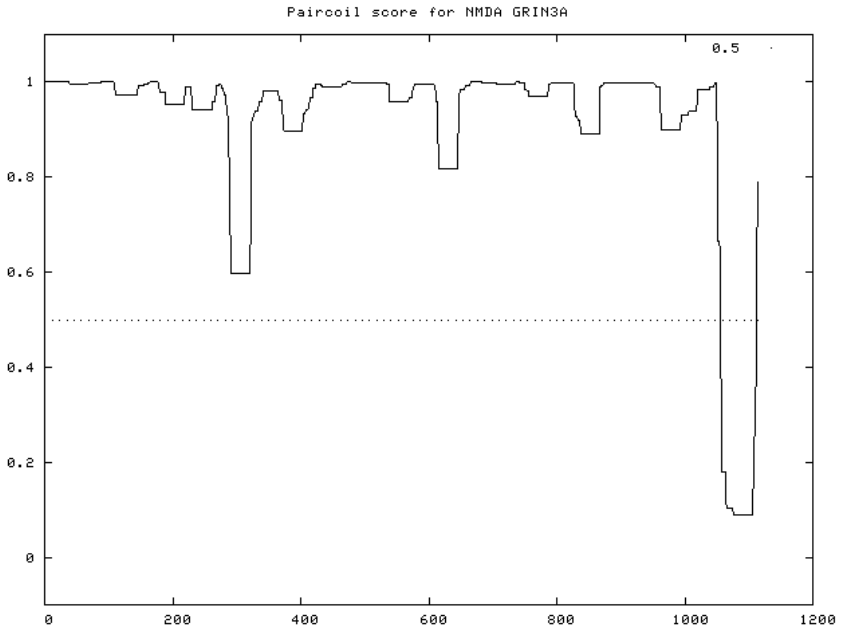


Figura 12. Probabilidad residuo por residuo. Representa la probabilidad residuo por residuo de que ese residuo se encuentre en una coiled-coil. El eje x contiene la ubicación del residuo (desde el principio, siendo 1 el primer residuo en la secuencia), y el eje y contiene la probabilidad de una bobina enrollada. El límite de probabilidad se muestra como una línea discontinua. Este gráfico indica la alta probabilidad de encontrar un segmento coiled coil en la posición 1000 a la 1115. En el resto de la secuencia se observan otros picos de menor probabilidad distribuidos en las regiones que otros programas señalaron como transmembranas. Se utilizaron los parámetros por default.

El análisis a través de los servidores citados (dynamine, MobiDB, IUPred2A) arroja resultados de gran similitud donde sugieren que mi proteína tiene un bajo porcentaje de desorden, excepto por el segmento inicial de mi proteína, de posiciones 50-150 aminoácidos (que es la región que presenta mayor desorden) y el segmento final de 1000-1115 aminoácidos. El resto de la secuencia parece tener una estructura muy conservada y por ende predecible, lo que indica que no guarda grandes distancias evolutivas con sus homólogos. Como conclusión, podemos destacar la presencia de una región de unión desordenada en el segmento que va de la posición 1 a la 200 aproximadamente, región donde se ubicaba el péptido señal anteriormente predicho. Y además se identificó a la región c-terminal como desordenada, lo que convierte a este segmento en una fuente de variabilidad genética importante.

Búsqueda de regiones repetitivas

En cuanto al análisis de la secuencia en la base de datos de proteínas repetitivas en tándem anotadas, RepeatsDB (Paladin L., et. al, 2021) proporciona la posición de la unidad, clasificación y referencia a otras bases de datos) arrojó cero resultados en su búsqueda, lo que sugeriría que esta proteína no posee una unidad repetitiva en su estructura que permita clasificarla como una proteína repetitiva.

Búsqueda de dominios conservados y motivos secuenciales

Se realizó la búsqueda de dominios conservados mediante PFAM (Mistry, J, et. al, 2021), InterPro (Blum et. al, 2020), CCD (Lu S, et. al 2020), CDART (Geer LY, 2002) y Prosite (Sigrist, et. al, 2012). Pfam es una colección de alineamientos secuenciales múltiples y de perfiles de modelos ocultos de Markov (HMM). Cada Pfam HMM representa una familia o dominio de proteínas. Al buscar una secuencia de proteínas en la biblioteca Pfam de HMM, puede determinar qué dominios lleva, es decir, su arquitectura de dominio. Pfam también se puede utilizar para analizar proteomas y arquitecturas de dominio más complejas. Como resultado de la

búsqueda se encontraron 3 dominios (Figura S11): Lig_chan (PF00060, Receptor ionotrópico de glutamato, 674-942); Lig_chan-Glu_bd (PF10613 ,Canal iónico ligado L-glutamato y sitio de unión a glicina, 557-661); ANF_receptor (PF01094, Región de unión del ligando de la familia de receptores, 124-471).

InterPro proporciona análisis funcional de proteínas clasificándolas en familias y prediciendo dominios y sitios importantes. Para clasificar las proteínas de esta manera, InterPro utiliza modelos predictivos (firmas), proporcionados por diversas bases de datos. Para nuestra proteína de interés se encontró la siguiente información: Receptor NMDA 3A, subtipo de receptor NMDA de canales iónicos activados por glutamato con conductancia monocanal reducida, baja permeabilidad al calcio y baja sensibilidad al magnesio dependiente del voltaje. Mediada por glicina. Durante el desarrollo de los circuitos neuronales, juega un papel en el período de refinamiento sináptico, restringiendo la maduración y el crecimiento de la columna. Al competir con la interacción GIT1 con ARHGEF7 / beta-PIX, puede reducir la activación local regulada por GIT1 / ARHGEF7 de RAC1, lo que afecta la señalización y limita la maduración y el crecimiento de las sinapsis inactivas. También puede desempeñar un papel en el mecanismo de señalización mediado por PPP2CB-NMDAR. (Protein family membership: receptor ionotrópico de glutamato, metazoa, IPR001508).

En la base de datos CDD, las secuencias de proteínas de estructuras tridimensionales se incluyen en modelos de dominio siempre que sea posible, uno de los objetivos es hacer que las alineaciones de secuencias múltiples estén de acuerdo con lo que podemos inferir de la estructura tridimensional y la superposición de estructuras tridimensionales, para comprender las relaciones secuencia / estructura / función (Figura 14).

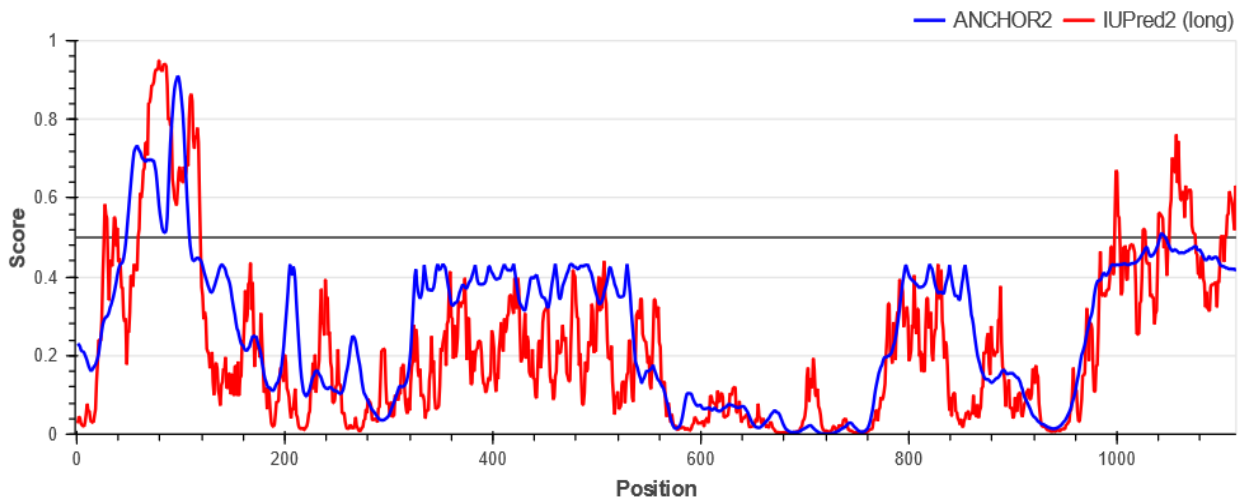


Figura 13. Regiones desordenadas predichas por IUPred2A. Las regiones desordenadas son aquellos picos que superan el valor de score delimitado con una línea negra. Se observan 2 regiones de mayor desorden: segmento que va de la posición 40 a la 160 aproximadamente, y un segundo segmento ligeramente desordenado que va de la posición 960 (aprox) a la 1115. Además se puede observar que aquellas regiones que se clasificaron como transmembrana (600-800 y 800-900) son las regiones con menos desorden, más conservadas y predecibles. También podemos ver que las regiones de unión desordenadas identificadas por el programa tienen un pico en el primer segmento de la secuencia (1-200), que a su vez coincide con la región que contiene el segmento más desordenado de la proteína. Se utilizaron los parámetros por default: IUPred2 long disorder (default).

El Conserved Domain Architecture Retrieval Tool (CDART) encuentra similitudes proteicas a través de distancias evolutivas significativas utilizando perfiles de dominio sensibles en lugar de similitud de secuencia directa. Dada una secuencia proteica de consulta, CDART muestra los dominios conservados que la componen, identificados por RPS-BLAST, y luego enumera las proteínas con una arquitectura de dominio conservado similar. Realizando una búsqueda con parámetros por default con la secuencia del receptor ionotrópico NMDA 3A se encontraron un total de 3161 arquitecturas, donde los primeros 3 hits contenían 3 dominios (Figura S12) :

-Periplasmic binding protein type 1 (intervalo 40-496). Compuesto por la familia de reguladores transcripcionales de los dominios de unión a LacI, proteínas periplasmáticas del transporte ABC, la familia de receptores GPCRs,y familia de receptores NPRs.

-Lig Chan superfamily (intervalo 676-942). Este dominio incluye las 4 regiones transmembranales de los receptores ionotrópicos NMDA.

-Periplasmic binding protein type 2 (intervalo 512-908). Representa los dominios de unión y ligamiento encontrados en proteínas ligadoras de solutos, que sirven como receptores iniciales en el transporte, transducción de señales, y channel gating.

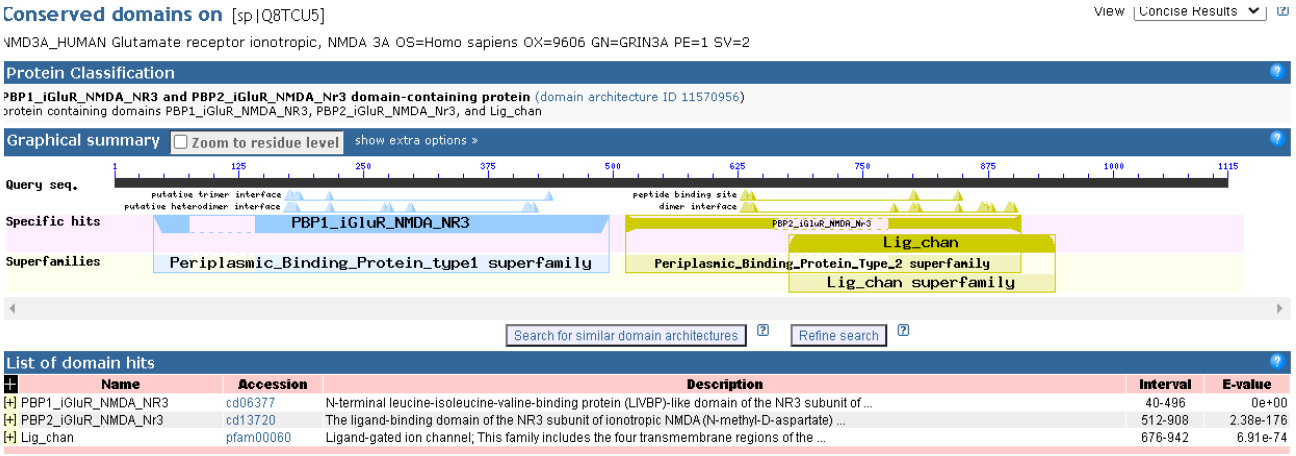
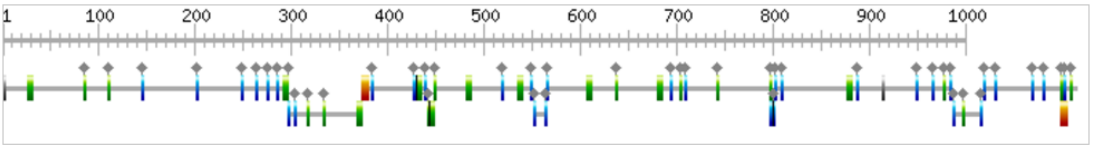


Figura 14. Clasificación dada por CCD para la proteína Q8TCU5. La proteína pertenece a una superfamilia de canales iónicos de ligando controlados que presenta 4 regiones transmembrana de los receptores ionotrópicos de glutamato y los receptores NMDA.También se destaca la superfamilia de pliegues de unión periplásmica tipo 2; este modelo evolutivo y jerarquía representan los dominios de unión de ligandos que se encuentran en las proteínas de unión de solutos que sirven como receptores iniciales en el transporte, la transducción de señales y la activación de canales. El origen del módulo PBP se puede rastrear a través de los filos distantes, incluidos eucariotas, arqueobacterias y procariotas. Además de las proteínas de transporte, la familia incluye receptores de glutamato ionotrópicos y proteínas sensoras no ortodoxas implicadas en la transducción de señales. El dominio de unión al sustrato de los reguladores de la transcripción LysR y los sistemas de transporte de tipo oligopéptido también contienen el pliegue de unión periplásmico de tipo 2 y, por tanto, son significativamente homólogos al de PBP2. Parámetros: Expect Value: 0,01, Maximum number of hits: 500.

Se realizó la búsqueda de motivos secuenciales mediante ScanProsite (De Castro, et. al 2006), un servidor que permite escanear proteínas en busca de coincidencias con la colección de motivos PROSITE, así como con patrones definidos por el usuario. Como resultado del escaneo rápido con parámetros por default (excluyendo motivos con alta probabilidad de ocurrencia) no se encontraron motivos. Sin embargo, al incluir en la búsqueda motivos con alta probabilidad de ocurrencia se obtuvieron los resultados mostrados en la figura 15.

Por último, para obtener la clasificación estructural de Q8TCU5 se realizó una búsqueda secuencial en CATH (Greene, et. al 2007). La base estructural CATH clasifica dominios proteicos basándose en comparación estructural. Los dominios que comparten los primeros 4 números de CATH son homólogos. Esta base de datos está organizada de forma jerárquica, con 9 clasificaciones: clase (proporción de residuos que adoptan la conformación α-hélice o β-plegada), arquitectura (estructura secundaria en el espacio), topología (conectividad y arreglo de la estructura secundaria), superfamilia de homólogos (dominios homólogos), familia secuencial, familia de ortólogos, dominios similares, dominios idénticos, y dominios únicos. Los primeros 4 niveles corresponden a una clasificación del tipo secuencial y estructural, las restantes se basan únicamente en una clasificación secuencial. Se encontraron 84 coincidencias en la búsqueda secuencial de dominios (principalmente regiones de unión a ligandos, desactivación de los sitios activos, etc), y unos 50

resultados en la búsqueda secuencial de familias funcionales (mayormente familias del receptor NMDA, AMPA, y kainato) (Figura 16).



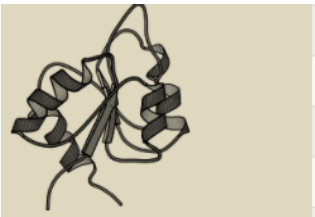
Un solo sitio

PS00004	CAMP_PHOSPHO_SITE	sitio de fosforilación de proteína quinasa dependiente de cAMP y cGMP
PS00007	TYR_PHOSPHO_SITE	Sitio de fosforilación de tirosina quinasa 1
PS00017	ATP_GTP_A	Motivo A del sitio de unión de ATP / GTP (bucle P) ----->Un solo sitio
PS00009	AMIDATION	amidación sitio ----->Un solo sitio

Numerosos sitios

PS00008	MYRISTYL	N-myristoylation sitio
PS00005	PKC_PHOSPHO_SITE	Sitio de fosforilación de la proteína quinasa C
PS00001	ASN_GLYCOSYLATION	Sitio de N-glicosilación
PS00006	CK2_PHOSPHO_SITE	Sitio de fosforilación de caseína quinasa II

Figura 15. Motivos con alta probabilidad de ocurrencia detectados. La figura representa aciertos por patrones con una alta probabilidad de ocurrencia o por patrones definidos por el usuario: [61 aciertos (por patrones distintos) en 1 secuencia]. Los motivos con una alta probabilidad de aparición son en la mayoría de los casos patrones que se encuentran en muchas secuencias de proteínas. Algunos de ellos describen, modificaciones postraduccionales que se encuentran comúnmente, y algunas otras regiones con sesgo de composición.



Level	CATH Code	Description
C	3	Alpha Beta
A	3.40	3-Layer(aba) Sandwich
T	3.40.190	D-Maltodextrin-Binding Protein; domain 2
H	3.40.190.10	Periplasmic binding protein-like II

Figura 16. Clasificación dada por CATH. En el recuadro se muestra las categorías con las que se clasificó a la proteína de acuerdo a la estructura jerárquica de CATH: clase=alfa-beta, arquitectura=3-layer(aba) Sandwich, topología= proteína de unión a D-Maltodextrina; dominio 2.

Asignación de plegamiento

El primer paso del proceso de modelado comparativo está definido como la predicción del plegamiento estructural de la proteína objetivo a partir de su secuencia de aminoácidos mediante la detección de proteínas homólogas de estructura tridimensional conocida. La idea se basa en la hipótesis de Anfinsen de que la estructura de las proteínas en un determinado entorno es a su vez determinada por la secuencia de proteínas. Una vez que se predice el plegamiento de la proteína objetivo, se debe elegir el mejor candidato a utilizar como template, cuya óptima elección es fundamental para asegurar la calidad del modelo a obtener.

A continuación se describen los métodos y bases de datos utilizadas para obtener los homólogos estructurales más cercanos de la proteína GluN3A.

En una primera instancia, se realizó la búsqueda de homólogos en la base de datos PDB (Burley, et. al 2021) (Protein Data Bank). PDB organiza las macromoléculas biológicas por su estructura jerárquica (estructura primaria, secundaria, terciaria, y cuaternaria) para simplificar la búsqueda, proporcionando sus respectivas estructuras obtenidas por técnicas de difracción de rayos X, RMN, microscopía crioelectrónica y modelado teórico. Como resultado no se encontraron entradas específicas del receptor ionotrópico de glutamato NMDA subunidad 3A de humanos, pero sí se encontraron entradas de estructuras homólogas con hasta un 50% ID. De los resultados de la búsqueda por similitud secuencial en la base PDB se obtuvo la siguiente lista

de proteínas: 2RC9_1 (69%ID), 4KCD_1 (69%ID), 2RC7_1 (69%ID), 2RC8_1 (69%ID), 2RCA_1 (57%ID), 2RCB_1 (57%ID). Los resultados se resumen en la tabla 2.

Con el objetivo de buscar diferentes confórmeros estructurales se utilizó la base de datos CoDNaS (Monzon, et. al, 2016). La base de datos CoDNaS es una colección de estructuras cristalográficas redundantes para una proteína dada ampliamente vinculada con información estructural, biológica y fisicoquímica. Varias proteínas depositadas en la base de datos PDB se han cristalizado en diferentes condiciones (por ejemplo con y sin la presencia de un ligando dado, en diferentes estados oligoméricos, con o sin presencia de modificaciones postraduccionales, etc.), que según la teoría de la selección conformacional, son factores que se podrían usar para estudiar los cambios conformacionales y correlacionarlos con la información biológica. Al realizar la búsqueda con la secuencia de la proteína GluN3A no se encontraron resultados. Esto se debe a que esta proteína no está anotada en la base PDB, debido a que su estructura aún no fue dilucidada. En cambio, al buscar el Código de la proteína 2RC9 (homólogo estructural más cercano), la base de datos arrojó un único resultado correspondiente a la proteína 2CR7, que resulta ser otra conformación de la proteína 2RC9, ya que comparten el mismo Código de Uniprot (Bateman, 2021) Q9R1M7, cuyo nombre de entrada es NMD3A_RAT.

Adicionalmente se realizó la búsqueda de homólogos con estructura conocida mediante Psi-blast, HHPred (Zimmermann, et. al, 2018), LOMETS (Wu S, y Zhang Y, 2007) y Phyre2 (Kelley, et. al, 2015). En el primer caso los primeros 8 hits fueron identificados como posibles template por sus elevado %ID y query coverage, bajo valor de e-value, y por poseer una estructura definida en la base PDB (Tabla 4).

Tabla 4. Búsqueda de homólogos de estructura conocida.

Accession number	Código PDB	Código Uniprot
NMDZ1_RAT	4KCC	P35439
GRIA2_RAT	4YU0	P19491
NMD3A_RAT	2RC7	Q9R1M7
NMDE2_RAT	5FXG_B	Q00960
NMDZ1_RAT	6WHR_B	P35439
NMDE2_RAT	6CNA_B	Q00960
NMDZ1_RAT	6WI1_B	P35439
NMDZ1_XENLA	5UOW_D	A0A1L8F5J9

HHpred es un método para la búsqueda de bases de datos de secuencias y la predicción de estructuras que es mucho más sensible que BLAST o PSI-BLAST para encontrar homólogos remotos. La secuencia objetivo o MSA se utiliza para construir un HMM (Hidden Markov Model), que está alineado con todos los HMM que representan proteínas anotadas o dominios con estructura conocida en bases de datos de alineación como Pfam y SMART. Los HMM pueden incluir información SS (experimentalmente determinada o predicha). Como resultado se obtuvieron 377 hits, con una Query MSA diversity (Neff) de 6.28031. Los 10 primeros hits de la búsqueda se muestran en la tabla 5.

Adicionalmente, se realizó una búsqueda con HHPred utilizando el MSA de los 100 primeros hits obtenidos por PSI-BLAST con,los siguientes parámetros por defecto: MSA generation method=HHblits=>UniRef30, MSA generation iterations=3, E-value cutoff for MSA generation=1e-3, Min seq identity of MSA hits with query (%)=20, Min coverage of MSA hits (%)=40, Secondary structure scoring=during_alignment, Alignment Mode: Realign with MAC =global:realign, MAC realignment threshold=0.3, Max target hits=1000, Min probability in hitlist (%)=20. (Tabla 6).

Tabla 5. Resultados de la búsqueda de homólogos remotos, de estructura y/o dominios conocidos mediante HHpred.

Uniprot accession number	Código Uniprot	Código PDB
GRIK5_RAT	Q63273	7KS0_D
GRIA1_MOUSE	P23818	7LDD_C
NMDZ1_RAT	P35439	6WI1_A
GRIK5_RAT	Q63273	7KS0_C
GRIK3_RAT	P42264	6JFY_C
NMDZ1_XENLA	A0A1L8F5J9	4TLL_C
P96404_MYCTU	P96404	6LDZ_A
NMDZ1_XENLA	A0A1L8F5J9	4TLL_D
GRID2_RAT	Q63226	6LU9_C
GRIA2_RAT	P19491	6PEQ_A

Tabla 6. MSA de los primeros 100 hits obtenidos por PSI-BLAST.

Uniprot accession number	Código Uniprot	Código PDB
GRIK5_RAT	Q63273	7KS0
NMDZ1_RAT	P35439	6WI1_A
NMDZ1_XENLA	A0A1L8F5J9	4TLL_D
GRIA1_MOUSE	P23818	7LDD_C
NMDZ1_XENLA	A0A1L8F5J9	4TLL_C
GRIK3_RAT	P42264	6JFY_C
NMDZ1_XENLA	A0A1L8F5J9	5TQ0_B y 5UN1_F
P96404_MYCTU	P96404	6DLZ_A
GRID2_RAT	Q63226	6LU9_C

LOMETS (Local Meta-Threading Server, versión 3) es un predictor de la estructura de proteínas basada en templates y la anotación de funciones basada en la estructura, que integra múltiples métodos de subprocesamiento basados en el deep-learning (CEthreader, DisCovER, EigenThreader, Hybrid-CEthreader, MapAlign) y programas basados en perfiles de última generación (FFAS3D (Xu D, 2013), HHpred, HHsearch, MRFsearch, MUSTER, SparksX). Con parámetros por default se identificaron 10 posibles templates, de los cuales se seleccionaron 10 de ellos para realizar el modelado final de la proteína con MODELLER (Webb B, & Sali A, 2016) y se utilizó L-BFGS para objetivos no homólogos mediante las restricciones de distancia predichas por DeepPotential y calculadas a partir de los templates mejores (top templates).

Phyre2 es un conjunto de herramientas disponibles en la web para predecir y analizar la estructura, función y mutaciones de las proteínas. Utilizando métodos avanzados de detección de homología remota. PHYRE2 construye modelos 3D, predice sitios de unión de ligandos y analiza el efecto de variantes de aminoácidos (por ejemplo, SNP no sinónimos (nsSNP)) para la secuencia de proteínas de un usuario. Proporcionando una secuencia de proteínas se puede obtener: la interpretación de la estructura secundaria y terciaria de sus modelos, la composición de su dominio y la calidad del modelo. La búsqueda de template con este servidor arrojó 6 posibles templates: c6irfD, c5uowB, c6mmiD, c4pe5B, c5kbuA, c4pe5A. La predicción de PHYRE2, con parámetros por defecto, fue realizada con 6 templates con base en métodos heurísticos para maximizar la confianza, el %ID, y el query coverage. El modelo predicho se presenta en el apéndice (Figura S13). Si bien el modelado fue intensivo, hay dos fragmentos sin modelar ya que los templates elegidos no cubren esos fragmentos. Las regiones sin modelar son (1-175) aproximadamente, y la (955-1115), ya que se observan desprovistos de una estructura secundaria determinada por falta de información en el template para su modelado. Hubo un total de 293 posiciones que fueron modeladas ab initio (tener en cuenta que el

modelado ab initio tiene bajo nivel de confianza), por falta de un template adecuado que cubriera estos segmentos. Los candidatos a template se encuentran resumidos en la tabla 7.

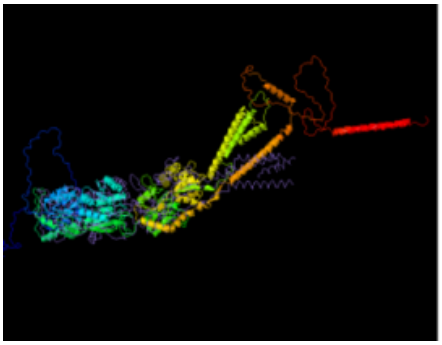
Tabla 2. Búsqueda por similitud secuencial en la base PDB del receptor ionotrópico del glutamato subunidad 3A.

ID de proteína	Descripción	Ligando
4KCD	Crystal Structure of the NMDA Receptor GluN3A Ligand Binding Domain Apo State. *Rattus norvegicus*	Glicerol
2RC9	Crystal structure of the NR3A ligand binding core complex with ACPC at 1.96 Angstrom resolution. *Rattus norvegicus*	ácido 1-aminociclopropano carboxilico
2RC7	Crystal structure of the NR3A ligand binding core complex with glycine at 1.58 Angstrom resolution. *Rattus norvegicus*	Ion bromuro
2RCB	Crystal structure of the NR3B ligand binding core complex with D-serine at 1.62 Angstrom resolution. *Rattus norvegicus*	D-serina
2RC8	Crystal structure of the NR3A ligand binding core complex with D-serine at 1.45 Angstrom resolution. *Rattus norvegicus*	Ion cloruro
2RCA	Crystal structure of the NR3B ligand binding core complex with glycine at 1.58 Angstrom resolution. *Rattus norvegicus*	Glicina

Tabla 3. Resultados de la búsqueda del Código de Uniprot 2RC9.

ID_POOL_CoDNaS	UniProt	#CONF	RMSD min	RMSD max	RMSD avg	Protein Name
2RC7_A	Q9R1M7	8	0.2	1.35	0.7682	Glutamate NMDA receptor subunit 3A

Nota: código correspondiente al homólogo estructural más cercano del que se tiene información, en relación a la proteína Q8TCU5.



Click to view	Rank	RCSB Link	TM-score	RMSD[a]	IDEN[a]	Cov.	Download PDB	Download FASTA
○	1	3kg2	0.504	4.81	0.176	0.58	3kg2A	3kg2A
○	2	7ks0	0.475	4.7	0.178	0.54	7ks0A	7ks0A
○	3	4ugq	0.472	5.02	0.156	0.55	4ugqA	4ugqA
○	4	6ksp	0.39	6.99	0.143	0.51	6kspA	6kspA
○	5	5l2e	0.328	5.67	0.092	0.39	5l2eA	5l2eA
○	6	4pe5	0.314	6.71	0.213	0.39	4pe5A	4pe5A
○	7	7eb2	0.307	6.04	0.068	0.38	7eb2C	7eb2C
○	8	7ca3	0.305	6.4	0.077	0.38	7ca3A	7ca3A
○	9	6uo8	0.302	5.87	0.082	0.37	6uo8A	6uo8A
○	10	5kca	0.29	5.13	0.141	0.33	5kcaA	5kcaA

Figura 17. Modelo final de LOMETS. Construido con Modeller con los 10 templates (detectados por TM-align), estos análogos estructurales no logran una buena cobertura de la proteína ya que apenas logran cubrir un 58% de ella. Los alineamientos estructurales son de baja resolución en general, solo pudiéndose obtener hasta un 4,7 Å de RMSD para el mejor alineamiento estructural (con la proteína 3KG2), además de no poder lograr altos porcentajes de identidad entre los candidatos a template (%ID<20) y la proteína Q8TCU5, lo que posiblemente sea porque hay dominios de la proteína no identificados en las bases de datos, y por lo tanto no cubiertos.

Analizando los parámetros de la tabla 7 se seleccionó como template a la proteína 7ks0 con sus respectivas cadenas (A, B, y C), por su gran porcentaje de cobertura e identidad con la proteína Q8TCU5, además de poseer un score de alineamiento razonable (>800), siendo identificado como el homólogo estructural más cercano mediante la base RCSB PDB.

Modelado por homología utilizando Modeller

Teniendo en cuenta que las proteínas relacionadas evolutivamente tienden a tener estructuras similares, muchas estructuras proteicas pueden modelarse basándose en estructuras ya conocidas de proteínas relacionadas, a esto se le llamó "modelado por homología". Los modelos de la proteína objetivo se obtienen basándose en las coordenadas atómicas de estructuras conocidas, bajo el supuesto de que la estructura de la proteína a modelar es similar a la estructura de la proteína utilizada como template (estructura que servirá de molde para el modelado del objetivo), donde además la proteína objetivo y el template están relacionados evolutivamente (siendo proteínas homólogas), y en el alineamiento entre el template elegido y el objetivo. Uno de los programas más utilizados es Modeller, que modela estructuras tridimensionales de proteínas y sus ensamblajes mediante satisfacción de las restricciones en la estructura espacial de la(s) secuencia(s) de aminoácidos y ligandos a modelar, resultando en una estructura 3D que satisface estas restricciones lo mejor posible. El modelo 3D se optimiza con la función de densidad de probabilidad molecular (molpdf). Para el modelado molecular de la proteína NMD3A_HUMAN (Q8TCU5) los template elegidos fueron: 7KS0_C (del aminoácido 33-845) y 7KS0_C (del aminoácido 21-829). Como resultado se obtuvieron 10 modelos de los cuales, en base al mínimo valor de tanto el DOPE score, como el valor de Molpdf, se eligió el modelo UKNP.B99990006.

Tabla 8. Parámetros utilizados para seleccionar el modelo.

Nombre del modelo	Mol pdf	DOPE score
UKNP.B99990006	7197.50928	-91363.32813

El modelo UKNP.B99990006 fue posteriormente optimizado para obtener un mejor RMSD, removiendo los átomos que no pudieron ser modelados, el primer fragmento removido va del 1 al 44, y el otro fragmento removido va de la posición 957 a la 1016 en el modelo final. Se utilizó Pymol (visor y renderizador molecular) para comparar estructuralmente los templates utilizados contra el modelo obtenido de la proteína objetivo, y se optimizó el alineamiento estructural al remover los átomos no modelados del modelo final, disminuyendo el RMSD del alineamiento. Se obtuvo un RMSD=4.103 para 7KS0_D con el modelo, y un RMSD=21,494 para 7KS0_C y el modelo (Figura S14 (a) y (b)).

Para evaluar la calidad del modelo obtenido se utilizó ProSA (Wiederstein & Sippl, 2007), ya que calcula un puntaje de calidad general para una estructura de entrada específica. Si esta puntuación está fuera de un rango característico de las proteínas nativas, es probable que la estructura contenga errores. Para el análisis del modelo UKNP.B99990006 obtenido con Modeller se obtuvieron los que se muestran en la figura 18 (a), (b), y (c).

Para el análisis funcional se utilizó nuevamente Pymol para observar la superficie de la proteína modelada y sus interacciones (Figura S15).

Tabla 7. Selección de template.

Candidato a template	Codigo PDB	Resolución (Å)	Identidad	Query coverage	Score aligment (T-coffe)
GRIK5_RAT	7ks0_C	5,3	22%	33%	804
GRIK5_RAT	7ks0_A	5,3	22%	33%	804
GRIK5_RAT	7ks0_D	5,3	40%	40%	845
GRIK5_RAT	7ks0_B	5,3	40%	40%	845
NMDZ1_RAT	4pe5_A	3,96	24%	54%	878
GRIA2_RAT	3kg2_A	3,60	21%	33%	786
GRIA1_MOUSE	7ldd_A	3,40	20%	72%	850

Nota: Para la elección de los templates adecuados se utilizó como criterio los parámetros que se muestran en la tabla, priorizando aquellas entradas pdb que cuentan con una resolución más cercana a 2 Å, contando con un buen coverage de la proteína objetivo, y un alto valor para su porcentaje de identidad. Se han suprimido los candidatos que contaban con un gran número de gaps (como por ejemplo la proteína 5uow).

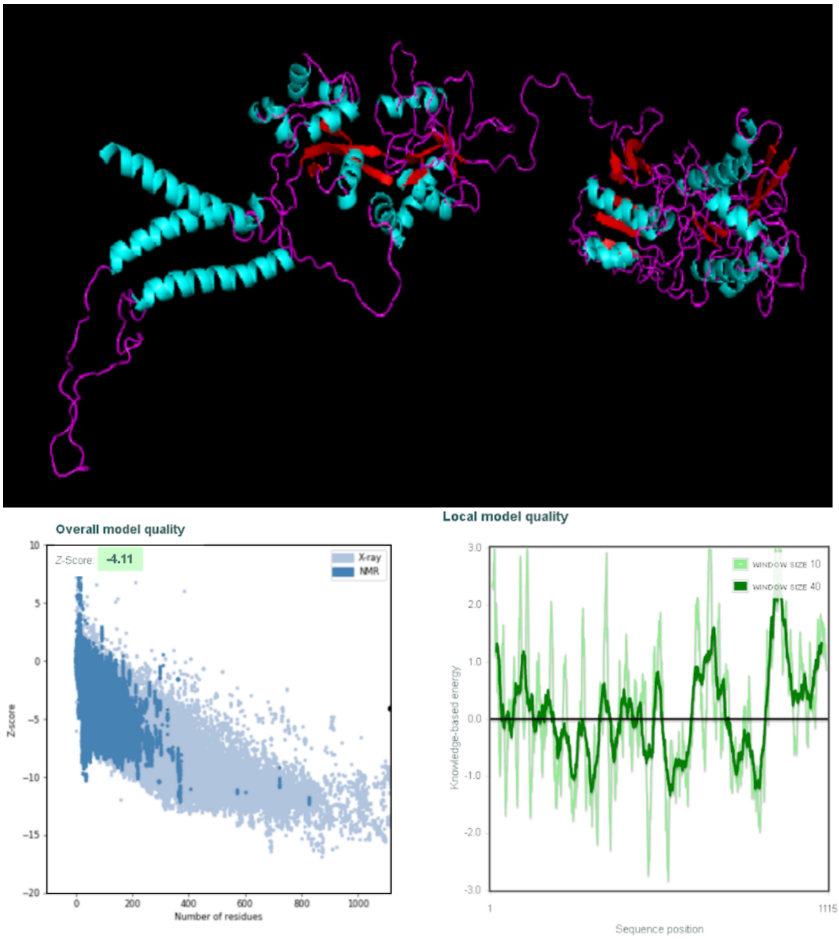


Figura 18 (a). Modelo refinado de Q8TCU5. En la figura se observa la estructura del modelo UKNP.B99990006 refinado con su correspondiente estructura secundaria: en color celeste se indican las hélices alfa, en color rojo se indican las hojas beta plegadas, y en fucsia los loops.(b). Calidad general del modelo. La figura izquierda representa la calidad general del modelo en función de los valores del Z-score obtenidos para estructuras determinadas experimentalmente (rayos X, NMR), como el valor de Z-score obtenido para el modelo UKNP.B99990006 (punto negro con un Z-score= -4.11) esta muy alejado del rango de valores para las estructuras experimentales que comparten la misma longitud para su secuencia, se puede decir que el modelo obtenido contiene grandes fallas. (c). Calidad del modelo local. La figura de la derecha muestra la calidad del modelo local, donde la linea color verde oscura simboliza la energía promedio sobre cada fragmento de 40 residuos, mientras que la linea de color verde claro representa la energía promedio sobre cada fragmento de 10 residuos. En general, los valores positivos corresponden a partes problemáticas o erróneas de la estructura de entrada. De esta manera se puede observar cuáles regiones del modelo obtenido presentan mayores problemas de modelado, en este caso se puede observar que solo una pequeña fracción de la proteína fue correctamente modelada (solo dos dominios) ya que prevalecen los picos de valores positivos.

Estimación filogenética

Como primera etapa en el estudio filogenético se obtuvo un alineamiento múltiple con T-Coffee (Notredame, et. al 2000) (Version_11.00, Cedric Notredame) utilizando secuencias de las siguientes

especies: *Homo sapiens*, *Pan troglodytes*, *Gorilla gorilla gorilla*, *Theropithecus gelada*, *Mandrilus leucophaeus*, *Papio anubis*, *Chlorocebus sabaeus*, *Cercocebus atys*, *Macaca Nernestrina*, *Macaca fascicularis*, *Hylobates moloch*, *Propithecus coquereli*, *Rhinopithecus roxellana*, *Rhinopithecus bieti*, *Colobus angolensis palliatus*, *Pongo abelii*, *Piliocolobus tephrosceles*, *Nomascus leucogenys*, *Trachypithecus francoisi*, *Aotus nancymaae*, *Cebus imitator*, *Saimiri boliviensis boliviensis*, *Sapajus apella*, *Callithrix jacchus*, *Rhinolophus ferrumequinum*, *Microcebus murinus*, *Otolemur garnettii*, *Carlito syrichta*. (N° de secuencias alineadas=35), (Figura S16).

Analizando el receptor ionotrópico del glutamato NMDA subunidad 3A, perteneciente a especies a la orden de los primates, se encontró una gran conservación de la proteína en general, siendo la proteína mayormente conservada, logrando un 99% ID promedio, y con un query coverage del 100%. Únicamente en el bloque n°12 que comprende el intervalo (1000-1080), hubo mayores diferencias con un total de 9 posiciones similares pero no idénticas, que es una de las regiones que presenta mayor desorden. En cuanto el resto de la proteína las variaciones por bloque eran mínimas (solo posiciones variantes en Q8TCU5) , lo que indica una gran conservación de la proteína receptora del glutamato (subunidad 3A) en primates.

A continuación se utilizó Modeltest es un subprograma del software HYPHY (Pond, 2005) para seleccionar el modelo de proteínas que mejor representa las secuencias bajo estudio. Por medio de un análisis estándar se generó una reconstrucción filogenética con las 35 secuencias homólogas de primates obtenidas con Blast, por el método de distancia (calculando una distancia genética entre un par de especies) de Neighbour Joining. Con este árbol se realizó una comparación entre modelos, obteniéndose como mejor modelo HIV between+F. Finalmente, se utilizó el software PHYML (Guindon, & Gascuel, 2003) para la inferencia de la filogenia basado en el principio estadístico de Maximum likehood (Parámetros utilizados: Model of amino-acids sustitution=HIVw, Amino-acids frequencies=empirical, Number of sustitution rate=4, Gamma distributed rate across sites=yes, Tree topology search options= Best of NNI and SPR, Non parametric boostrap analysis=yes, -Number of replicates=100, el resto de los parámetros se utilizaron por defecto) (Figura 19).

Predicción de función

Se realizó una búsqueda bibliográfica acerca del receptor ionotrópico de glutamato NMDA subunidad 3A. En la tabla 9 se resume parte de la información obtenida.

Se utilizó el servidor ConSurf (Ashkenazy, et. al 2016) para estimar la conservación evolutiva de aminoácidos basada en las relaciones filogenéticas entre secuencias homólogas. El grado en el que una posición se conserva evolutivamente (tasa de evolución estimada por método bayesiano o método de máxima verosimilitud) depende en gran medida de su importancia estructural y funcional, por lo que el análisis de conservación de posiciones entre miembros de la misma familia a menudo puede revelar la importancia de cada posición para la estructura o función de la proteína. Se obtuvieron múltiples output de MSA que mostraban la conservación o variabilidad de aminoácidos en las 35 secuencias de primates analizadas, así como también se obtuvo un árbol filogenético, obtenido por ML (Figura 20).

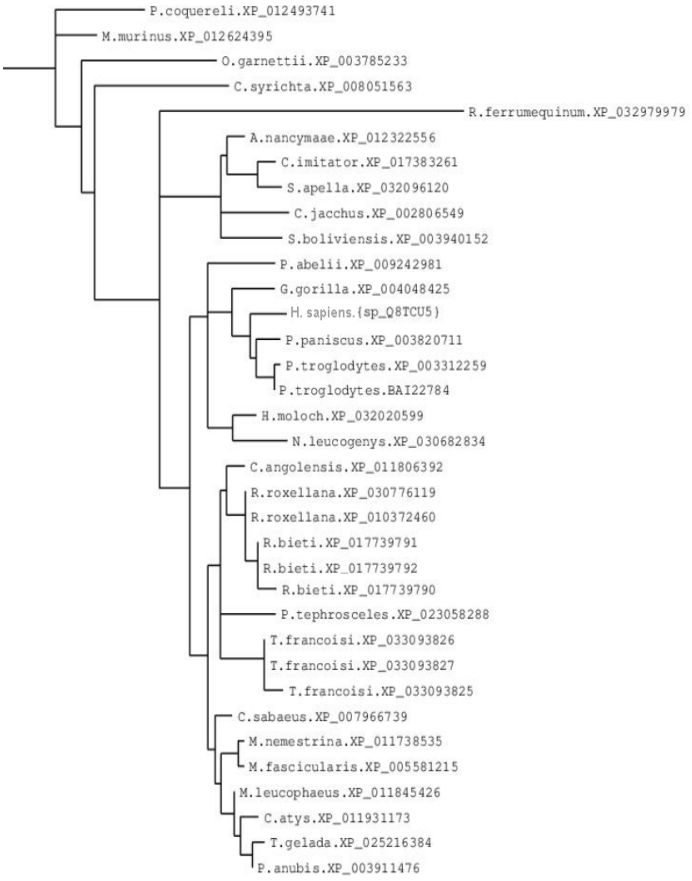


Figura 19. Árbol filogenético obtenido con PhyML. En la figura se muestra el árbol rooteado obtenido (con los 35 homólogos cercanos obtenidos anteriormente en Blast) en PhyML y posteriormente modificado con SNAD para una mejor visualización, se analizó la divergencia evolutiva del receptor ionotrópico del glutamato NMDA 3A. Se observa que la proteína NMDA GRIN3A de *H. sapiens*, tiene una distancia evolutiva menor con la proteína expresada en *P. paniscus* (chimpancé gracil), y con la proteína expresada en *P. troglodytes* (chimpancé común), lo cual es esperable ya que son las especies evolutivamente más cercanas a los *H. sapiens* a nivel genético. La especie con mayor divergencia secuencial respecto a la proteína encontrada en *H. sapiens* es *Papio anubis*.

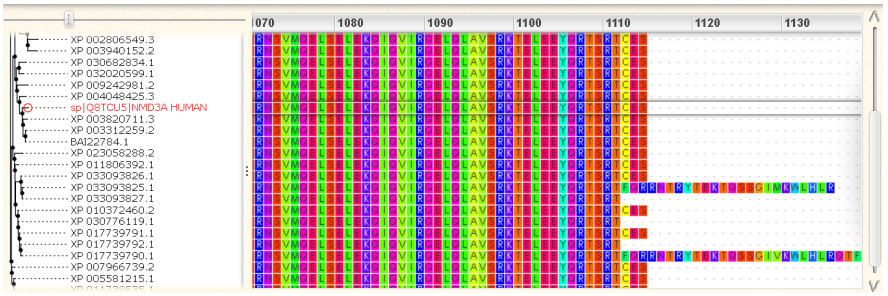
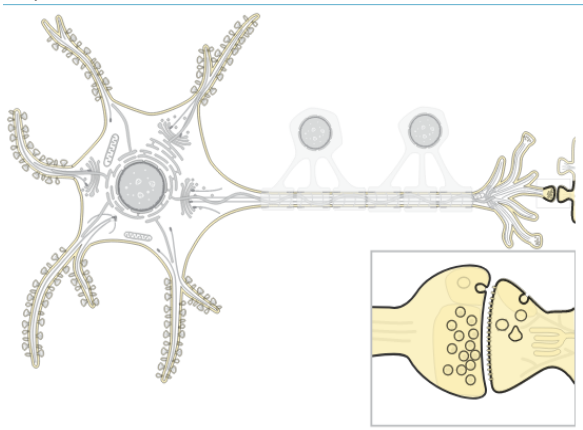


Figura 20. Árbol filogenético construido por ConSurf visualizado con Wasabi (Veidenberg, et. al, 2016). Se observa en la figura la variabilidad secuencial entre las secuencias de primates, donde existen dos segmentos variables no compartidos entre primates, y a su vez estos segmentos terminales tienen mayor longitud que en el resto de los primates. En el margen izquierdo se muestra el árbol construido por máxima similitud.

Tabla 9. Información general acerca del receptor ionotrópico NMDA 3A.

Nombres:	Glutamate receptor ionotropic, NMDA 3A/ GluN3a/NMDAR3A/NR3A
Función molecular	Canal de iones , canal iónico de apertura por ligando , Receptor
Proceso biológico	Transporte de iones , Transporte
Ligando	Calcio , magnesio
Nombres de genes	Nombre: GRIN3A
Organismo	*Homo sapiens* (humano)
Identificador taxonómico	9606 [NCBI]
Linaje taxonómico	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo
Proteomas	UP000005640 Componente i : Cromosoma 9
Función molecular	Actividad del canal iónico controlado por ligando. Fuente: GO_Central Actividad del receptor de glutamato NMDA Fuente: UniProtKB Unión a proteína fosfatasa 2A. Fuente: UniProtKB Actividad del receptor de señalización. Fuente: GO_Central Actividad del canal iónico dependiente del transmisor involucrado en la regulación del potencial de membrana postsináptica. Fuente: Ensembl Transporte de iones de calcio Fuente: UniProtKB Desarrollo de dendrita Fuente: Ensembl Vía de señalización del receptor de glutamato ionotrópico Fuente: GO_Central Regulación negativa del desarrollo de la columna dendrítica Fuente: UniProtKB Inhibición prepulso Fuente: Ensembl Regulación de la exocitosis de vesículas sinápticas Fuente: Ensembl Respuesta al etanol
Ubicación subcelular (Uniprot)	 Membrana celular Por similitud ; Pro membrana de múltiples pasadas Por similitud membrana celular postsináptica Por similitud densidad postsináptica Por similitud Nota: Enriquecido en membrana plasmática densidades postsinápticas. Requiere la pres apuntar a la membrana plasmática (por sim Por similitud

Para tratar de establecer una función para el dominio C-terminal se utilizó PSIPRED. Al realizar una búsqueda mediante este servidor para el fragmento C-terminal de la proteína Q8TCU5, es decir, de la posición 750 a la 1115, para el cual no se pudieron obtener predicciones fiables, se reconocieron dos posibles dominios en la región c-terminal (Figura 5, 6, 7, 31, y 32). De esta forma se detectó un segmento terminal extracelular.

Para la predicción de la función de la proteína objetivo se utilizó un método de aprendizaje automático, el método Support Vector Machine (Cai, et. al, 2001) para predecir la clase funcional de las proteínas y péptidos a partir de sus propiedades secuenciales, independizandose de la similitud secuencial. Las predicciones concuerdan ampliamente con las derivadas de InterPro, pudiendo resumir los resultados en la tabla 10.

CONCLUSIONES Y DISCUSIÓN

Se puede decir, dado los sucesivos análisis, que el receptor NMDA GRIN3A está altamente conservado en primates, posee al menos 4 hélices transmembrana, además de tener una región coiled-coil hacia el final de su secuencia (1000-1115) que es un segmento sin dominios detectables por su alta variabilidad secuencial, además de ser una de las regiones de mayor desorden junto con el segmento que va del 1 al 200 (posible

péptido señal). El resto de la proteína tiene una estructura rígida, por ende está bien conservada por lo que debe ser una proteína con una función biológica bien definida.

Tabla 10 . Resumen de los GO terms más relevantes.

Proceso biológico	Función molecular	Componente celular
Transporte de iones (GO:0006811)	actividad del receptor de glutamato ionotrópico (GO: 0004970)	membrana (GO: 0016020)
proceso biosintético de macromoléculas celulares (GO:0034645)	actividad del canal iónico controlado por ligando (GO: 0015276)	componente intrínseco de la membrana (GO:0031224)
Vía de señalización del receptor de superficie celular (GO:0007166)	actividad del canal iónico (GO: 0005216)	componente integral de la membrana (GO:0016021)
transporte de iones transmembrana (GO:0034220)	actividad del receptor de señalización (GO: 0038023)	periferia celular (GO:0071944)

Nota: Resumen de los GO terms asociados más relevantes recopilados de Interpro, QuickGO, Uniprot, y el software AmiGO de GeneOntology (GO).

Para el alineamiento múltiple de secuencias con la herramienta T-Coffee se seleccionaron 35 secuencias c/u perteneciente a las especies del orden de los primates, donde las especies con menor distancia evolutiva son: Homo sapiens, Pan troglodytes, Gorilla gorilla gorilla. La proteína está bien conservada, por lo que seguramente cumplirá la misma función biológica en estos 3 organismos (por transferencia). Se encontró en el MSA que los segmentos de las secuencias alineadas que muestran mayor número de gaps (mayor número de mutaciones, y por ende menor conservación) son los que corresponden a los segmentos de mayor desorden en NMD3A _ HUMAN (donde se predice una hélice transmembrana). Se puede observar que hay tres especies de primates con una región C-terminal diferente: *Carlito syrichta* (especie más alejada evolutivamente del resto de los primates), *Trachypithecus francois*, y *Rhinopitehcus bieti* que poseen una región C-terminal de mayor longitud que la esperada. Todas las secuencias de homólogos analizadas pertenecen al mismo cluster por guardar una alta similitud secuencial, estructural, y funcional.

No se han encontrado unidades de repetición por lo que no hay evidencias para decir que esta sea una proteína repetitiva. Al buscar dominios conservados se encontraron 4 dominios: la familia Lig_chan que es un receptor ionotrópico de glutamato en el segmento (674-942), la familia Lig_chan-Glu_bd que es una región con canal iónico ligado L-glutamato y sitio de unión a glicina en el segmento (557-661), la familia ANF_receptor que es una región de unión del ligando de la familia de receptores en el segmento (124-471), y el dominio de proteínas de unión periplásmica tipo 1 y 2 (60-690) aproximadamente. Solo se encontraron motivos secuenciales que se encuentran en una gran mayoría de proteínas por ser muy comunes. Se observó una gran cantidad de regiones de baja complejidad en el fragmento N-inicial y el C-terminal de la secuencia de la proteína Q8TCU5, lo que implica que estas dos regiones son fuente de alta variabilidad genética, propensas a aumentar la frecuencia de desarrollo de enfermedades neurodegenerativas, y podrían estar implicados en los procesos de adaptación (evolución del cerebro social humano).

En cuanto a la función molecular es un receptor de NMDA de canales iónicos activados por glutamato, de baja permeabilidad al calcio, por lo cual se predice que es un componente intrínseco de la membrana. Su función biológica está vinculada al refinamiento sináptico, restringiendo la maduración y el crecimiento de la columna, así como se relaciona a la poda de las sinapsis inactivas.

BIBLIOGRAFÍA

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Andres Veidenberg, Alan Medlar, Ari Löytynoja, Wasabi: An Integrated Platform for Evolutionary Sequence Analysis and Data Visualization. Molecular Biology and Evolution, Volume 33, Issue 4, April 2016, Pages 1126–1130, https://doi.org/10.1093/molbev/msv333.

A.V. McDonnell, T. Jiang, A.E. Keating, B. Berger. Paircoil2: Improved prediction of coiled coils from sequence. Bioinformatics Vol. 22(3) (2006).

Aron Marchler-Bauer, Huaiyu Mi, Darren A Natale, Marco Necci, Christine A Orengo, Arun P Pandurangan, Catherine Rivoire, Christian J A Sigrist, Ian Sillitoe, Narmada Thanki, Paul D Thomas, Silvio C E Tosatto, Cathy H Wu, Alex Bateman, Robert D Finn Nucleic Acids Research (2020), gkaa977, PMID: 33156333

Ashburner et al. Gene ontology: tool for the unification of biology. Nat Genet. May 2000;25(1):25-9.

Ashkenazy H., Abadi S., Martz E., Chay O., Mayrose I., Pupko T., and Ben-Tal N. 2016, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl. Acids Res. 2016; DOI: 10.1093/nar/gkw408; PMID: 27166375 [ABS], [PDF]

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

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

Buchan DWA, Jones DT (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz297

B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.

Cai, YD., Liu, XJ., Xu, Xb.et al. Support Vector Machines for predicting protein structural class. BMC Bioinformatics 2, 3 (2001). https://doi.org/10.1186/1471-2105-2-3.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, 50. AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics. Jan 2009;25(2):288-289.

CoDNaS 2.0: A comprehensive database of protein conformational diversity in the native state.Alexander M. Monzon; Cristian O. Rohr; María Silvina Fornasari; Gustavo Parisi. Database. Oxford Journals. Accepted on March 2016.DOI:10.1093/database/baw038. Pubmed

De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch 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]

DTU Health Tech, (2017). TMHMM: Prediction of transmembrane helices in proteins, Server v. 2.0. Center for Biological Sequence Analysis, the bioinformatic unit, Technical University of Denmark.

Elisa Cilia, Rita Pancsa, Peter Tompa, Tom Lenaerts, and Wim Vranken, From protein sequence to dynamics and disorder with DynaMine. Nature Communications 4:2741 doi: 10.1038/ncomms3741 (2013).

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

Greene, L. H., Lewis, T. E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., et al. (2007). The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Research, 35(Database issue), D291-7. doi:10.1093/nar/gkl959; https://www.cathdb.info/.

Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res. 2002 Oct;12(10):1619-23.

Greenwood Tiffany A, Lazzeroni Laura C, Murray Sarah S, Cadenhead Kristin S, Calkins Monica E, Dobie Dorcas J, Green Michael F, Gur Raquel E, Gur Ruben C, Hardiman Gary, Kelsoe John R, Leonard Sherry, Light Gregory A, Nuechterlein Keith H, Olincy Ann, Radant Allen D, Schork Nicholas J, Seidman Larry J, Siever Larry J, Silverman Jeremy M, Stone William S, Swerdlow Neal R, Tsuang Debby W, Tsuang Ming T, Turetsky Bruce I, & Freedman R. & Braff David L (2011). Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. The American journal of psychiatry. DOI: 10.1176/appi.ajp.2011.10050723

GRIN3A, Uniprot: the universal protein knowledgebase (2021) . https://www.uniprot.org/uniprot/Q8TCU5

Guindon S., Gascuel O. PhyML : "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology. 2003 52(5):696-704.

Gurillo Muñoz P. (2016). Revisión de los Trastornos del Espectro Psicótico: Sus características genéticas y función de las neurexinas. GRIN Verlag. https://www.grin.com/document/439363

Hoffman, K & Stoffel, W (1993). TMbase - A database of membrane spanning proteins segments, Biol. Chem. Hoppe-Seyler 374,166.

Kelley, L., Mezulis, S., Yates, C. et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053

Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003 Jul 1;31(13):3701-8. doi: 10.1093/nar/gkg519. PMID: 12824398; PMCID: PMC169197.

Lisanna Paladin, Martina Bevilacqua, Sara Errigo, Damiano Piovesan, Ivan Mičetić, Marco Necci, Alexander Miguel Monzon, Maria Laura Fabre, Jose Luis Lopez, Juliet F Nilsson, Javier Rios, Pablo Lorenzano Menna, Maia Cabrera, Martin Gonzalez Buitron, Mariane Gonçalves Kulik, Sebastian Fernandez-Alberti, Maria Silvina Fornasari, Gustavo Parisi, Antonio Lagares, Layla Hirsh, Miguel A Andrade-Navarro, Andrey V Kajava, Silvio C E Tosatto, RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D452–D457, https://doi.org/10.1093/nar/gkaa1097

Lukas Käll, Anders Krogh and Erik L. L. Sonnhammer. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server, Nucleic Acids Res., 35:W429-32, July 2007 (doi) (PubMed)

Lupas, A., Van Dyke, M., and Stock, J. (1991). COILS: Predicting Coiled Coils from Protein Sequences, Science 252:1162-1164.

MAPPING and MOLECULAR GENETICS. http://www.omim.org /entry/181500

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1). https://doi.org/10.1093/nar/gkaa913; http://pfam.xfam.org/.

M.Torrisi, M.Kaleel, G.Pollastri. Deeper Profiles and Cascaded Recurrent and Convolutional Neural Networks for state-of-the-art Protein Secondary Structure Prediction, (2019) Scientific Reports, 9: 12374, 2019, doi: 10.1038/s41598-019-48786-x

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

Piovesan D, Necci M, Escobedo N, Monzon AM, Hatos A, Mičetić I, Quaglia F, Paladin L, Ramasamy P, Dosztányi Z, Vranken WF, Davey N, Parisi G, Fuxreiter M and Tosatto SCE, (2020), MobiDB: intrinsically disordered proteins in 2021. Nucleic Acid Research. gkaa1058. PubMed

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy : hypothesis testing using phylogenies. Bioinformatics, 21(5), 676-679. doi:10.1093/bioinformatics/bti079.

Sánchez de las Matas Martín, María del Carmen. Teoría de la mente y esquizofrenia: aspectos conceptuales y evolutivos. InterSedes: Revista de las Sedes Regionales, vol. XV, núm. 30. (2014), pp. 169-196 Universidad de Costa Rica Ciudad Universitaria Carlos Monge Alfaro, Costa Rica.

S.C. Potter, A. Luciani, S.R. Eddy Y. Park, R. Lopez and R.D. Finn. (2018) HMMER web server: Nucleic Acids Research. Web Server Issue 46:W200-W204.

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE. Nucleic Acids Res. 2012; doi: 1093/nar/gks1067. PubMed:23161676 [Full text] [PDF version]

Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sai Ganesan, David S Goodsell, Sutapa Ghosh, Rachel Kramer Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Catherine L Lawson, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Persikova, Chris Randle, Alexander Rose, Yana Rose, Andrej Sali, Joan Segura, Monica Sekharan, Chenghua Shao, Yi-Ping Tao, Maria Voigt, John D Westbrook, Jasmine Y Young, Christine Zardecki, Marina Zhuravleva, RCSB Protein Data Bank:

powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D437–D451, https://doi.org/10.1093/nar/gkaa1038

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE. Nucleic Acids Res. 2012; doi: 1093/nar/gks1067. PubMed:23161676 [Full text] [PDF version]

Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sai Ganesan, David S Goodsell, Sutapa Ghosh, Rachel Kramer Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Catherine L Lawson, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Persikova, Chris Randle, Alexander Rose, Yana Rose, Andrej Sali, Joan Segura, Monica Sekharan, Chenghua Shao, Yi-Ping Tao, Maria Voigt, John D Westbrook, Jasmine Y Young, Christine Zardecki, Marina Zhuravleva, RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D437–D451, https://doi.org/10.1093/nar/gkaa1038

The InterPro protein families and domains database: 20 years on. Matthias Blum, Hsin-Yu Chang, Sara Chuguransky, Tiago Grego, Swaathi Kandasaamy, Alex Mitchell, Gift Nuka, Typhaine Paysan-Lafosse, Matloob Qureshi, Shriya Raj, Lorna Richardson, Gustavo A Salazar, Lowri Williams, Peer Bork, Alan Bridge, Julian Gough, Daniel H Haft, Ivica Letunic, Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D265-D268. doi: 10.1093/nar/gkz991. (Epub 2019 Nov 28.) [PubMed PMID: 31777944] [Full Text at Oxford Academic]

The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. Jan 2021;49(D1):D325-D334.

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, https://doi.org/10.1093/nar/gkaa1100

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.

Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Research. 35, 3375-3382 (2007).

Xu D., Jaroszewski L., Li Z., Godzik A. (2013). FFAS-3D: Improving fold recognition by including optimized structural features and template re-ranking. Bioinformatics (2013) doi: 10.1093. PubMed

Zhang M, Liu D, Tang J, Feng Y, Wang T, Dobbin KK, Schliekelman P, Zhao S. SEG - A Software Program for Finding Somatic Copy Number Alterations in Whole Genome Sequencing Data of Cancer. Comput Struct Biotechnol J. 2018 Sep 7;16:335-341. eCollection 2018. https://doi.org/10.1016/j.csbj.2018.09.001

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.. J Mol Biol. 2018 Jul 20. S0022-2836(17)30587-9.

Zhu F, Han LY, Chen X, Lin HH, Ong S, Xie B, Zhang HL, Chen YZ. Homology-free prediction of functional class of proteins and peptides by support vector machines. Curr Protein Pept Sci. 2008 Feb;9(1):70-95. doi: 10.2174/138920308783565697. PMID: 18336324.