

PROTEINA DKFZp434P0714 DE *Homo sapiens* SOBRE EL ALZHEIMER: MODELO ESTRUCTURAL Y REGIONES FUNCIONALES ASOCIADAS

Edison Sesquilé¹, Diana Vargas¹, George Barreto², Janneth González²

¹ Pontificia Universidad Javeriana, Colombia

janneth.gonzalez@javeriana.edu.co.

RESUMEN

Predecir computacionalmente la estructura tridimensional de la proteína DKFZp434P0714 de *Homo sapiens* e inferir las regiones funcionales asociadas importantes para la enfermedad del Alzheimer. Se realizó un análisis computacional de la estructura primaria de DKFZp434P0714 a partir de los servidores BLAST, PROTPARAM, PROTSCALE, TMHMM2 y PROSITE. La estructura secundaria se obtuvo por PSIPRED. La predicción de la estructura terciaria de la proteína se realizó mediante el programa Phyre2. El modelo final de la estructura terciaria fue visualizado y editado con el programa PyMol v.0.98. Se validó la estructura terciaria mediante el programa RAMPAGE. La predicción de sitios activos se realizó mediante el servidor Q-sitefinder. Por último, se realizó un modelamiento con la proteína homóloga del BIAST armadillo repeatcontaining 2, isoformCRA_b [*Homo sapiens*] con el programa CLUSPRO 2.0. Se propone un modelo computacional, se demuestra la existencia de patrones que podrían tener un papel importante en el Alzheimer; se encontró el dominio (ARM). También un dominio receptor LDL-receptor class A, con un receptor neuronal LRP1 unido a APP, afectando el tráfico celular que induce el desarrollo del Alzheimer. La transcripción Tcf/LEF-1 mediado por β -catenina presente en ARM, está involucrada en la determinación del destino de células progenitoras gliales en la neocorteza humana, estas con fenotipo antígeno NG2+ predominan en cerebros lesionados como en el Alzheimer. La coincidencia entre segmentos estructurales y funcionales sugiere que el modelo puede usarse para predecir

ciertos aspectos del comportamiento biológico de la proteína en cuanto a los mecanismos de activación de la enfermedad de Alzheimer.

INTRODUCCIÓN

La enfermedad de Alzheimer es una afección neurodegenerativa, ésta ataca principalmente la memoria del ser humano haciendo que pierda paulatinamente la actividad cerebral que se ve reflejado en la pérdida de neuronas, baja actividad sináptica, aparición de placas seniles y degeneración neurofibrilar. Esto afecta las capacidades neuronales básicas como razonar, hablar, actuar, vestirse, caminar y recordar. (Peña, 1999; Lanfranco et al. 2012; Donoso, 2003).

Los síntomas del Alzheimer se clasifican en macroscópicos y microscópicos, en los primeros se evidencian atrofas cerebrales predominantemente en la corteza temporal. En los microscópicos se observan dos síntomas importantes situados en las placas amiloides y los ovillos neurofibrilares que son producto de la agregación de proteínas anormales de A β y TAU. Otras neuropatológicas incluyen la pérdida neuronal, inflamación y angiopatía amiloide (Shepherd et al. 2009; Lopera, 2012).

Los pacientes afectados pierden progresivamente su independencia ante la imposibilidad de llevar su vida autónomamente, requiriendo asistencia constante de terceros; por esta razón es una de las enfermedades más temidas, por su carácter irreversible e incurable, convirtiéndola en un proceso devastador tanto para el paciente como para su familia (Peña, 1999; Fuentes & Prato, 2012). Alzheimer es un tipo de demencia que en la última década ha sido el centro de atención de los medios de comunicación, toda persona ha oído sobre la enfermedad, tiene algún familiar afectado o conoce alguien que la padece, porque la persona empieza a perder memoria y comportarse de forma inapropiada; con casos reportados de personas saliendo a la calle en pijama, utilizando erróneamente el dinero o la ropa (Peña, 1999).

Esta patología es la más frecuente en los adultos mayores con un porcentaje del 50- 60% del total de los casos, esta enfermedad no es común en edades tempranas (antes de los 60 años) y cuando se presenta está entre el 1% y 3% de los casos. A los 65 años tiene un porcentaje del 5 al 8%, duplicándose cada 5 años y llegando a cifras del 30 al 50% en personas mayores de 85 años; Aunque el Alzheimer familiar de inicio precoz que es antes de los 65 años, tiene un patrón de herencia mendeliano autosómico dominante y puede ser causado por mutaciones en el gen de la PPA, PS1 O PS2 (Fuentes & Prato, 2012; Lanfranco et al. 2012; Lopera, 2012).

Actualmente se estima que más de 35 millones de personas alrededor del mundo están afectadas, es muy dependiente de la región donde se presenta la patología, siendo mayor en los países desarrollados de América del Norte como Europa occidental y una segunda parte en los países en desarrollo en Asia y América Latina (Ballard et al. 2011; Querfurth & Laferla, 2010).

No se han realizado estudios sobre las funciones estructurales de la proteína DKFZp434P0714, lo que nos lleva a indagar en métodos computacionales que nos permitan hacer una aproximación al conocimiento estructural de la proteína, razón por la cual han tomado importancia para el entendimiento de los mecanismos de acción de las proteínas a partir de su estructura tridimensional (Xu et al. 2007). Por lo anterior, el objetivo de este estudio consistió en predecir computacionalmente la estructura tridimensional de la proteína hipotética DKFZp434P0714 e inferir regiones funcionales asociadas importantes para la enfermedad del Alzheimer. Los resultados evidencian la primera aproximación existente de la estructura tridimensional de la proteína DKFZp434P0714 y el encuentro de patrones funcionales asociados a ésta que podrían ser importantes para los mecanismos de la enfermedad del Alzheimer.

MATERIALES Y MÉTODOS

Análisis computacional de la estructura primaria de DKFZp434P0714

La secuencia de la proteína hipotética DKFZp434P0714 (Acc. No. |52545596|emb|CAB66688.2|) fue obtenida a través del GenBank en el Centro Nacional de Información para Biotecnología (NCBI), la búsqueda de similitud con otras secuencias reportadas actuales se llevó a cabo con BLAST disponible en NCBI. Las propiedades fisicoquímicas, tales como el peso molecular, el índice de estabilidad, el punto isoeléctrico, el tiempo de vida, el índice alifático y el promedio general de hidropatía, fueron calculadas con PROTPARAM (<http://expasy.org/tools/>). Para el análisis del índice de hidrofobicidad se usó la herramienta PROTSCALE y el algoritmo de Kyte y Doolittle del servidor Expasy proteomic tolos (<http://expasy.org/tools/>). La predicción de segmentos transmembranales se obtuvo a partir del programa TMHMM2 del Centro de Análisis de Secuencias Biológicas (CBS) de la Universidad Técnica de Dinamarca (<http://www.cbs.dtu.dk/services/>). La identificación de motivos se realizó con el servidor PROSITE.

Predicción de estructuras secundaria y terciaria de DKFZp434P0714

La predicción de la estructura terciaria de la proteína se realizó mediante el programa Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>). El modelo final de la estructura terciaria fue visualizado y editado con el programa PyMol v.0.98 (<http://www.pymol.org>). Se validó la estructura terciaria mediante el programa RAMPAGE para establecer si nuestra la estructura es o no estable. Por último, se realizó una modelamiento de nuestra proteína hipotética DKFZp434P0714 con la proteína homóloga del BLAST armadillo repeat containing 2, isoformCRA_b [Homo sapiens] con el programa CLUSPRO 2.0 (<http://cluspro.bu.edu/login.php>).

RESULTADOS

Análisis computacional de la estructura primaria de DKFZp434P0714

El alineamiento local de secuencias realizado con BLAST permitió la identificación de una proteína con secuencia de aminoácidos similar en un 100.0% a la secuencia de DKFZp434P0714: La proteína armadillo repeat containing 2, isoformCRA_b (Número de acceso: ref|EAW48370.1) de Homo sapiens las cuales comparten la superfamilia de dominios ARM (Tabla. 1). Teniendo en cuenta que la identidad de cada proteína se basa en el orden preciso de cada uno de sus 16 aminoácidos, el porcentaje de identidad o similitud entre las secuencias DKFZp434P0714 y armadillo repeatcontaining 2, isoformCRA_b permite deducir que éstas son homólogas e idénticas en cuanto a la conformación de su estructura tridimensional y por tanto en sus funciones.

Además, mostro un dominio homólogo conservado del sistema informático BLAST, el cual fue identificado como armadillo/beta-catenina repeat (ARM), que es un dominio característico de las proteínas armadillo y según lo reportado por Noll et al. (2000) y Levesque et al. (1999) se establece que la proteína hipotética podría estar implicada en el desarrollo temprano del alzheimer, ya que estos dominios ARM interactúan con la proteína presenilina (PS). Esta última tiene una alta correlación con la aparición temprana del alzheimer en seres humanos, las mutaciones en presenilina 1 (PS1) y presenilina 2 (PS2) son las principales responsables de generar esta patología, contando con más de 50 mutaciones descritas. La relación de las proteínas armadillo con presenilina están ligados a los bucles de las presenilina dentro del ARM, porque tienen un dominio altamente conservado entre los residuos 372-399 del extremo C-terminal en el gran bucle

citoplásmico de PS1; lo que sugiere que forman un sitio de unión funcional importante para presenilinas.

La β -catenina la cual se encuentra también dentro del dominio ARM ubicándose específicamente en el citoplasma celular, cumpliendo funciones de morfogénesis y mantenimiento de tejidos celulares, para esto es necesario una adhesión molecular con proteínas cadherina del citoplasma y citoesqueleto; esto forma complejos como el de cadherina-catenina, que juega un papel en la regulación de señales que controlan la adhesión y movilidad celular, la fosforilación de tirosina que hace parte del gen ARM. β -catenina forma un complejo con Tcf/LEF-1, que tiene función de transcripción y que las lleva al núcleo en donde se une a sitios específicos del ADN, lo que activa la expresión génica (Huber et al. 1997). Esto repercute en el Alzheimer por la presencia de depósitos de β -amiloide extracelular, resultado del procesamiento anormal de la proteína precursora de amiloide (APP) y dos enzimas: β -secretasa, y γ -secretasa en las placas neuronales, paredes de los vasos sanguíneos, ovillos neurofibrilares intraneuronales, estas condiciones en conjunto resultan en la pérdida acelerada de neuronas y la demencia (Abisambra et al. 2010).

Por otro lado, estudios como el de He & Shen (2009) indican que la cascada de transcripción Tcf/LEF-1/ de β -catenina presente en ARM, está involucrada en la determinación del destino de células progenitoras gliales ubicadas en la neocorteza humana. Estas células gliales con fenotipo antígeno NG2+ predominan en cerebros lesionados, como es el caso de la enfermedad de Alzheimer; en la que genera gliogenesis en la neocorteza degenerada en etapas tempranas de desarrollo como en adultos, desmintiendo la idea que las neuronas no se regeneran en el hipocampo y epéndimo de pacientes adultos con trastornos neurodegenerativos crónicos, como el Alzheimer, analizando que la vía señalización β -catenina es un factor muy importante en la expresión de los genes proneurales. Nuevamente, se establece otra función que podría tener nuestra proteína hipotética en relación con el Alzheimer.

Las propiedades fisicoquímicas obtenidas mediante PROTPARAM (Tabla. 2) permitieron establecer el peso molecular, el índice de estabilidad, el punto isoeléctrico, el tiempo de vida, el índice alifático y el promedio general de hidropatía de DKFZp434P0714. El punto isoeléctrico es donde las cargas positivas de la proteína se igualan a las negativas anulando la existencia de movimiento en un campo eléctrico representa por tanto el pH al cual DKFZp434P0714 podría mostrar un mínimo de solubilidad durante ensayos experimentales, facilitando su aislamiento en un campo eléctrico durante los mismos experimentos (Werner, 2007). El índice alifático permite inferir que la proteína es altamente termoestable por su valor de 105.34, lo que obedece a las observaciones de

Ikai en las cuales las proteínas de origen mesofílico con alto índice alifático suelen ser más termoestables (Hooda, 2011).

El análisis hidrofóbico de ProtScale (Figura. 1) reveló que es altamente se afirma que proteína hipotética DKFZp434P0714 es altamente hidrofílica y se encuentra fuera de la membrana celular lo que nos predice que esta proteína es globular.

En cuanto a los motivos de PROSITE se identificó un dominio receptor LDL-receptor class A, el cual tiene un receptor neuronal LRP1 que se une a APP, esta asociación afecta el tráfico celular induciendo el desarrollo de la enfermedad de Alzheimer. Siendo APP una proteína genéticamente implicada en el desarrollo en esta enfermedad, el procesamiento proteolítico de esta catalizado cuando el dominio LDL receptora class A se expresa afectando la señalización de la proteína (Ranganathan et al. 2011).

Análisis computacional de la estructura secundaria de DKFZp434P0714

La predicción de la estructura secundaria se realizó con el servidor PSIPRED (Figura 4). Este reveló la existencia de 40 hélices alfa entre los aminoácidos 11-14, 46-51, 82-88, 97-100, 102-111, 114-125, 127-143, 151-165, 169-184, 186-201, 208-212, 215-223, 230-245, 248-258, 260-277, 286-300, 303-311, 314-322, 328-342, 345-368, 372-385, 390-401, 403-419, 441-455, 462-467, 471- 481, 484-502, 508-526, 530-544, 548-557, 559-567, 571-585, 588-598, 601-611, 614-630, 637-640, 644-654, 657-662, 667-681, 683-695; 1 lámina alfa entre los aminoácidos 17-21 y los aminoácidos restantes fueron regiones no estructuradas.

Análisis computacional de la estructura terciaria de DKFZp434P0714

Los resultados del tándem del sistema phyre2 de nuestra proteína hipotética mostró que podría estar asociada con funciones de adhesión celular y ciclo celular, lo que corroboramos con lo encontrado en la literatura de Huber et al. (1997); como es el caso del mantenimiento y morfogénesis de tejidos ya que juega un papel muy importante en control de las cascadas de señalización para una adhesión y movilidad celular, que son dos procesos muy importantes dentro de cualquier ciclo celular ya que son la base de muchos procesos celulares como es el de la diferenciación celular (organogénesis).

La validación del modelo de DKFZp434P0714 evaluado con RAMPAGE cumplió los parámetros estereoquímicos de una estructura estable con 354 (95,4%) residuos en regiones favorecidas, 10 (2,7%) residuos en regiones permitidas y 5 (1,4%) residuos fuera de regiones favorecidas (Figura. 5). En base a la probabilidad que la combinación de los pares de ángulos sean posibles en esta proteína DKFZp434P0714, las regiones

favorecidas corresponden a los sitios en donde la probabilidad que los ángulos que conforman la estructura proteica sea posible está más cercana a uno; mientras que las regiones permitidas y las no permitidas corresponden a sitios con menor probabilidad de acercamiento al ángulo correcto. Todo esto nos lleva a inferir que el modelo establece un acercamiento confiable a la estructura terciaria real de la proteína.

La predicción de sitios activos se realizó con el programa Q-sitefinder (Figura 7) reveló la posición de 10 sitios activos probables en base a la energía en la unión de sondas hidrófobas de (CH₃) a la proteína. Los 10 sitios de unión más favorables están representados con áreas de color turquesa, verde claro, verde oscuro, azul oscuro, oscuro claro, púrpura, lila, marrón, naranja y café. El volumen aproximado de la cavidad junto a las coordenadas calculadas (Tabla 4) señalan que las coordenadas mínimas (33, -46, -70) y máximas (47, -31, -56), salvo para el sitio activo 9 cuyas coordenadas mínimas son (35, -32, -76) y máximas (49, -15, -57).

CONCLUSIONES

La coincidencia entre segmentos estructurales y funcionales del modelo la proteína hipotética sugiere que este puede usarse para predecir ciertos aspectos del comportamiento de la proteína en cuanto a los mecanismos de activación de la enfermedad de Alzheimer. La proteína DKFZp434P0714 podría estar implicada en el desarrollo temprano del Alzheimer debido a su relación con los bucles de la proteína presenilina (PS), nuestra proteína forma un sitio de unión funcional importante para PS en el dominio Armadillo para la activación de la enfermedad; dentro del dominio ARM se encuentra la β -catenina, que en conjunto con proteínas cadherinas colaboran en el procesamiento anormal de la proteína precursora amiloide y de las enzimas β -secretasa, γ -secretasa en las placas neuronales, resultando en la pérdida acelerada de neuronas y la demencia, además la cascada de transcripción Tcf/LEF-1/ de β -catenina presente en ARM genera gliogenesis en la neocorteza degenerada en pacientes con Alzheimer, presentando la posibilidad de regenerar neuronas en el hipocampo y epéndimo en adultos de edades avanzadas; lo que termina siendo un problema no solo para las persona que padece la patología, sino para toda su familia por las consecuencias que se derivan de esta enfermedad degenerativa.

BIBLIOGRAFÍA

1. Peña J. Enfermedad del Alzheimer. Del diagnóstico a la terapia: conceptos y hechos. Fundación la Caixa. España. 1999, 137p.
2. Lanfranco R, Manríquez-Navarro P, Avello L, Canales-Johnson, A. Evaluación de la enfermedad de Alzheimer en etapa temprana: biomarcadores y pruebas neuropsicológicas. Rev Med Chile 2012; 140 : 1191-1200.
3. Fuentes P, Patro J. 2012. Comunicación diagnóstica en enfermedad de Alzheimer. Reflexión y propuesta. Rev Med Chile 2012; 140: 811-817.
4. Donoso A. La enfermedad de Alzheimer. Enfermedad de alzheimer: una nota histórica. Rev chil neuro-psiquiat 2003; 41(2): 13-22.
5. Levesque G, Yu G, Nishimura M, Zhang D, Levesque L, et al. Presenilins Interact with Armadillo Proteins Including Neural-Specific Plakophilin-Related Protein and b-Catenin. Journal of Neurochemistry 1999; 72(3): 999-1008.
6. Noll E, Medina M, Hartley D, Zhou J, Perrimon N, Kosik K. Presenilin Affects Arm/b-Catenin Localization and Function in Drosophila. Developmental Biology 2000; 227(2): 450-464.
7. Shepherd C, McCann H, Halliday G. Variations in the neuropathology of familial Alzheimer's disease. Acta Neuropathologica 2009; 118: 37-52.
8. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer's disease. Lancet 2011; 377(9770): 1019-1031.
9. Querfurth H, LaFerla F. Alzheimer's disease. New England Journal of Medicine 2010; 362: 329-344.
10. He P, Shen Y. Interruption of β -Catenin Signaling Reduces Neurogenesis in Alzheimer's Disease. The journal of neuroscience 2009; 29 (20) 6545-6557.
11. Huber A, Nelson J, Weis W. Three-Dimensional Structure of the Armadillo Repeat Region of β -catenin. Cell 1997; 90 (5): 871-882.

12. Lopera F. Enfermedad de Alzheimer Familiar. Revista Neuropsicología. Neuropsiquiatría y Neurociencias. 2012: 12(1): 163-188.
13. Abisambra J, Fiorelli T, Padmanabhan J, Neame P, Wefes I. LDLR Expression and Localization Are Altered in Mouse and Human Cell Culture Models of Alzheimer's Disease. PLoS ONE 2010: 5(1): e8556.
14. Ranganathan S, Noyes N, Leon M, Winkles J, Battey F. LRAD3, A Novel Low-Density Lipoprotein Receptor Family Member That Modulates Amyloid Precursor Protein Trafficking. The journal of neuroscience 2011: 31(30): 10836-10846.
15. Xu Y, Xu D, Liang D. Computational Methods for Protein structure Prediction and Modeling. Springer. Springer. USA. 2007, 335 p.
16. Werner MS. Biología: Fundamentos para Medicina y Ciencias de la Vida. Reverté. Barcelona, España. 2007, 661p.
17. Hooda V. Physicochemical, Functional and structural characterization of wheat germin using in silico methods. Current Research Journal of Biological Sciences 2011: 3 (1): 35-41.

ANEXOS

Tabla 1. Resultados alineamiento local de secuencias de proteínas.

Descripción	Score	Identidad	e-value	Numero de acceso
armadillo repeatcontaining 2, isoformCRA_b [Homo sapiens]	1456	100%	0.0	EAW48370.1
unnamed protein product [Homo sapiens]	1450	99%	0.0	BAG53010.1
unnamed protein product [Homo sapiens]	1436	99%	0.0	BAG57766.1
unnamed protein product [Macaca fascicularis]	1335	96%	0.0	BAE00686.1

Tabla 2. Propiedades físico-químicas de la proteína hipotética DKFZp434P0714.

Numero de aminoácidos	708
Peso molecular	79265.6
IP y pH teórico	6.57
Número total de residuos cargados negativamente (Asp + Glu)	83
Número total de residuos cargados positivamente (Arg + Lys)	79
Vida media estimada	1 hora en reticulocitos de mamífero 30 minutos en levadura en vivo 10 horas en <i>Escherichia coli</i> en vivo
Índice de inestabilidad	40.57
Índice alifático	105.34
Gran media de hidrofobicidad	-0.111

Tabla 3. Porcentaje y residuos de aminoácidos en la proteína

Aminoácido	Numero de residuos	Porcentaje de residuos (%)
Ala (A)	39	5,5
Arg (R)	26	3,7
Asn (N)	46	6,5
Asp (D)	39	5,5
Cys (C)	21	3
Gln (Q)	33	4,7
Glu (E)	44	6,2
Gly (G)	36	5,1
His (H)	18	2,5
Ile (I)	46	6,5
Leu (L)	104	14,7
Lys (K)	53	7,5
Met (M)	12	1,7
Phe (F)	27	3,8
Pro (P)	19	2,7
Ser (S)	53	7,5
Thr (T)	34	4,8
Trp (W)	4	0,6
Tyr (Y)	12	1,7
Val (V)	42	5,9
Pyl (O)	0	0
Sec (U)	0	0

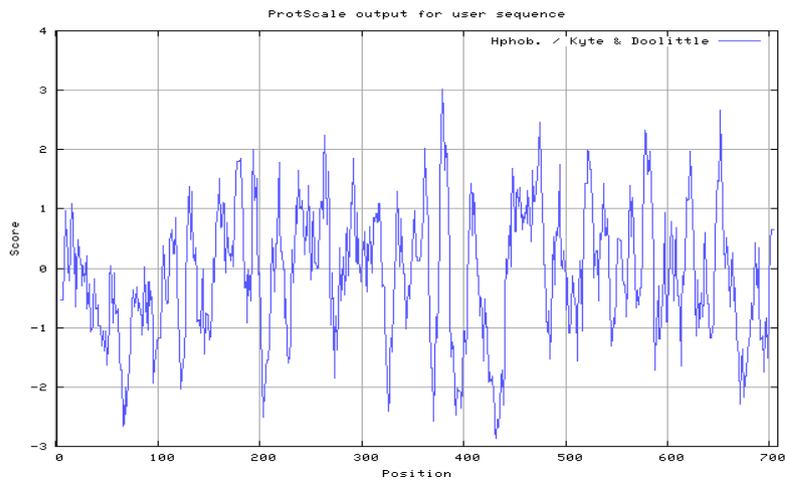


Figura 1. Análisis de hidrofobicidad PROTSCALE

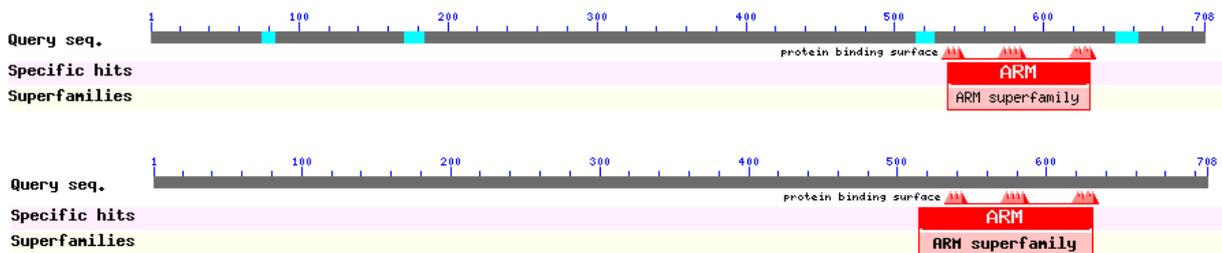


Figura 2. Dominios conservados NCBI DC y BLAST

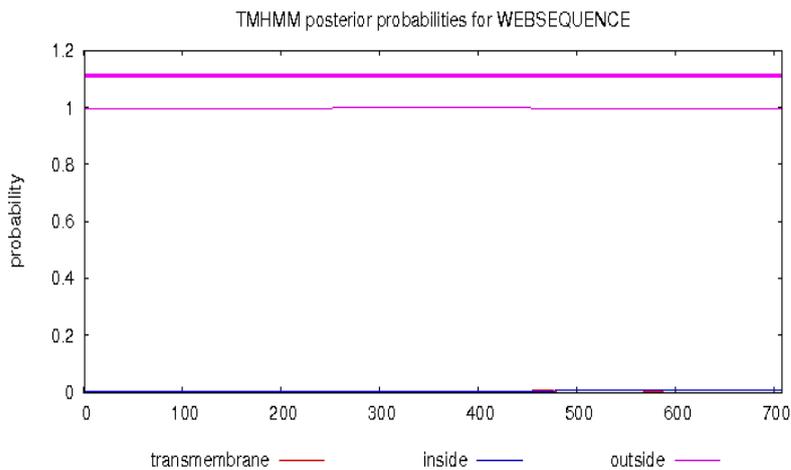


Figura 3. Ubicación de la proteína en la célula en TMHMM del CBS

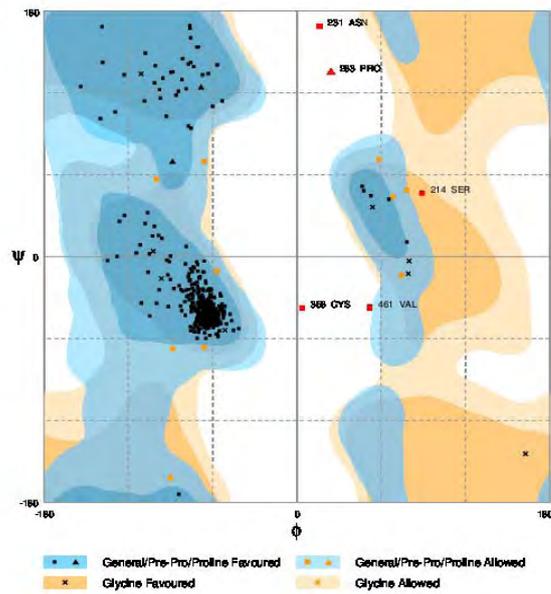


Figura 5. Validación de la estructura terciaria. Distribución de la combinación de ángulos para la estructura DKFZp434P0714 de alta resolución. El área azul más intenso corresponde a las regiones más favorecidas, las de menos intensidad corresponden a las regiones permitidas. El área naranja más intenso corresponde a las regiones más favorecidas para Glicina, que normalmente es altamente conservado en secuencias proteicas permitiendo la combinación de ángulos que no son posibles para otros aminoácidos.



Figura 6. Modelamiento de la proteína hipotética DKFZp434P0714 con la proteína homologa del BIAST armadillo repeatcontaining 2, isoformCRA_b [Homo sapiens] con el programa CLUSPRO 2.0v (docking).

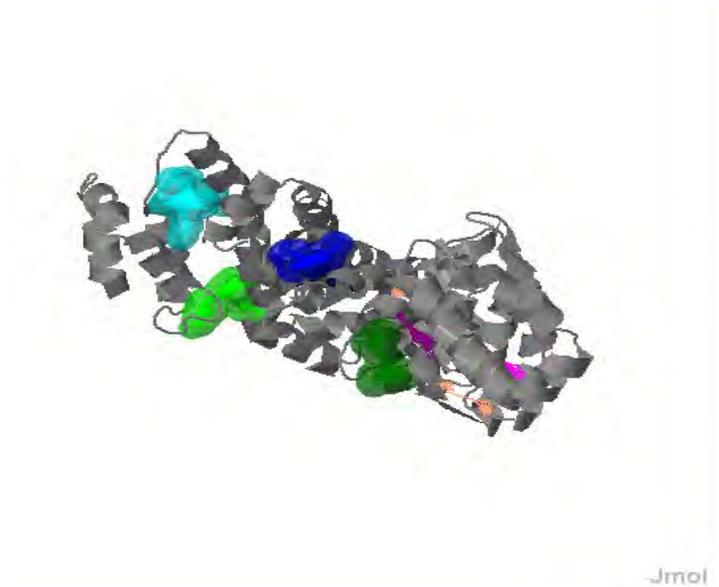


Figura 7. Modelamiento de los sitios activos de la proteína DKFZp434P0714 con el programa Q-sitefinder. El área turquesa corresponde a la unión de sondas hidrófobas de (CH₃) a la proteína con energía más favorable, el siguiente sitio activo más probable de un sitio activo corresponde al área verde claro, seguido por el área verde oscuro y por último el área azul.

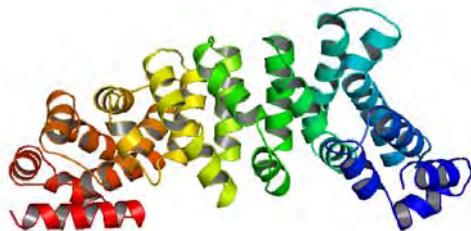


Figura 8. Proteína DKFZp434P0714 editada y arreglada en PyMOL.