

## **PREDICCIÓN IN-SÍLICO DE ASPECTOS ESTRUCTURALES Y FUNCIONALES DE UNA PROTEÍNA HIPOTÉTICA RELACIONADA CON ESCLEROSIS LATERAL AMIOTRÓFICA.**

Nathali Coral Acosta<sup>1</sup>, María Carolina Arevalo, Isis Londoño Chaparro, Jessica Santander Guzman, George Barreto, Janneth González<sup>1</sup>

Pontificia Universidad Javeriana. Facultad de ciencias

\* janneth.gonzalez@javeriana.edu.co

### **RESUMEN**

**Objetivo.** Se realizaron predicciones estructurales y funcionales de la proteína hipotética CAD89935.1 para observar su relación con ELA. **Materiales y métodos.** Se utilizaron varias herramientas de la bioinformática para la caracterización de la estructura primaria, secundaria y terciaria de la proteína hipotética. **Resultados.** La proteína es altamente hidrofóbica, lo que indica que es transmembranal. Perteneciente a la familia MILTON y presenta regiones HAP1\_N que hacen de la proteína un regulador tipo inhibidor de la neurotransmisión. Utilizando la herramienta LOMETS se realizó la predicción de la estructura terciaria y por medio de Pymol se ubicó el sitio del ligando, ubicado entre los residuos 121 al 256. **Conclusión.** El modelo tridimensional obtenido coincide con la funcionalidad de las regiones encontradas en la proteína. La proteína es un regulador tipo inhibidor de la neurotransmisión y es muy probable que esté relacionada con enfermedades neurodegenerativas como la esclerosis lateral amiotrófica y el Huntington.

### **INTRODUCCIÓN**

En el año de 1989, el médico neurólogo Jean-Martin Charcot describió la condición patológica caracterizada por el deterioro de las neuronas motoras al que denominó esclerosis lateral amiotrófica (ELA), conocida también como la enfermedad de Lou-Gehrig. Ésta es una enfermedad neurodegenerativa que se caracteriza por la pérdida de actividad y capacidad motora partiendo de un síndrome neurológico mixto de las motoneuronas

superiores, comprometiendo los núcleos basales y la corteza cerebral, las columnas grises de los cordones medulares espinales y el tallo cerebral (1).

En el 2005 estudios relacionados con esta enfermedad lograron aislar la proteína hipotética CAD 89935.1 del tejido de la médula espinal de un ser humano afectado por ELA. Aunque no se ha caracterizado a nivel estructural esta proteína, los datos del NCBI (National Center for Biotechnology Information) (2) indican que pertenece a la familia de proteínas Milton relacionadas con esta enfermedad neurodegenerativa.

Se ha encontrado que los defectos en el transporte axonal implicados con esta enfermedad generan una disfuncionalidad sináptica, debido a que se presenta una degeneración en el transporte retrógrado de los axones y provocan el deterioro de las cadenas pesadas de las kinesinas, la kinesina 1a (conocida también como KIF5A), provocando una mutación que implica neuropatías en los mamíferos. Las alteraciones en las actividades de las proteínas kinesinas conllevan a mecanismos potencialmente patogénicos y representan otras enfermedades neurodegenerativas como el Alzheimer y Huntington (3).

Por otra parte las células han desarrollado diversos sistemas de entrega para realizar el transporte intracelular, controlados por las interacciones entre microtúbulos y proteínas en donde se encuentran sitios altamente conservados que codifican el transporte de organelos a diferentes velocidades y direccionalidades. Existen dos tipos de motores implicados en este proceso: los anterógrados, que dirigen el transporte desde el cuerpo celular hasta la periferia de los axones y los motores retrógrados, característicos de las enfermedades neurodegenerativas, que transportan desde la periferia hasta el cuerpo celular. Por lo tanto una alteración de una proteína asociada a microtúbulos provoca que el transporte intracelular se vea afectado de manera drástica (4).

Aunque se han realizado muchos estudios respecto a la esclerosis lateral amiotrófica a nivel fisiológico, no está comprobado científicamente qué genes pueden estar involucrados en la mutación de las proteínas asociadas a ELA (3). Por esta razón, se hace necesario un mayor número de estudios que precisen las causas a nivel molecular que ocasionan el mal funcionamiento de estas proteínas.

Por medio de herramientas bioinformáticas se realizó una predicción estructural de la proteína hipotética CAD8935.1 (conformada por 779 aminoácidos) con el objetivo de analizar regiones funcionales y sus potenciales sitios de unión. Esta caracterización aportará conocimiento sobre la fisiopatología molecular de ELA y otras enfermedades

como el Huntington que afecta a los seres humanos. Este estudio puede ser útil para que se hagan futuros tratamientos relacionados con esta enfermedad

## **MATERIALES Y MÉTODOS**

Utilizando las bases de datos del NCBI (National Center for Biotechnology Information), se buscó la proteína hipotética relacionada con la enfermedad de Esclerosis lateral amiotrófica en humanos con el código de acceso CAD8935.1 y se obtuvo su secuencia de aminoácidos en formato FASTA.

### **Análisis de la estructura primaria**

A partir de la secuencia FASTA se determinaron las propiedades físico-químicas de la molécula (peso molecular, punto isoeléctrico, perfiles de hidrofobicidad, coeficiente de extinción, inestabilidad y vida media) utilizando la herramienta Protparam del portal de recursos bioinformáticos ExPASy (Gasteiger 2005). El gráfico de perfiles hidrofóbicos se realizó con la herramienta ProtScale que utiliza el algoritmo Kyte-Doolittle (6) (escala de -3.5 a 4.2, en el eje de las ordenadas y posición de los aminoácidos en el eje de las abscisas).

Posteriormente se sometió la secuencia a un BLAST (Basic Local Alignment Search Tool), herramienta perteneciente al NCBI (2), lo cual realizó la búsqueda de proteínas homólogas para inferir características de la molécula (familia a la que pertenece y función hipotética), teniendo en cuenta los mayores score y e-value. Quince de estas secuencias fueron escogidas para realizar un alineamiento múltiple mediante el servidor T-Coffee (7).

Para la identificación de la familia a la que pertenece la proteína CAD89935.1, se utilizó PFAM (8) y de esta manera se logró identificar las regiones en donde posiblemente se encuentran los dominios conservados dados por BLAST(2).

### **Predicción de la estructura secundaria**

Se utilizó el programa Phyre (9), que utiliza las bases de datos de SCOP (Structural Classification Of Proteins) y PDB (Protein Data Base) para obtener mediante un Psi-Blast secuencias homólogas de la cadena query. A partir del resultado del alineamiento utiliza los programas Psi-Pred, SPRO y JNet para realizar con cada uno la predicción de la hélice  $\alpha$ , cadena  $\beta$  y coils respectivamente, finalmente realiza un consenso de los 3 programas, el cual se tomó para el análisis.

## **Predicción de la estructura terciaria**

Mediante el programa de LOMETS(10) se obtuvieron las posibles estructuras terciarias de la proteína CAD89935.1. que se validaron con el programa MolProbity(11) para obtener el Ramachandran Plot, y teniendo en cuenta este resultado, se escogió la proteína modelada 1wa5C. Además se realizó un alineamiento tridimensional con la proteína de mayor hit en el programa I-TASSER.

## **Predicción de sitios de interacción**

Se utilizó la base de datos String (12) para predecir las interacciones que la proteína hipotética tiene con otras proteínas, como lo son la homología, co-expresión y co-ocurrencia utilizando un nivel de confianza medio mayor de 0.4.

La secuencia de la proteína estudiada se sometió al programa 3DLigandsite (13) para identificar los posibles sitios de unión; a partir de estos resultados, por medio de la herramienta Pymol, se realizó una visualización tanto la proteína como del ligando. Finalmente, se utilizó el servidor Pocket-Finder(14) que utiliza el algoritmo propuesto por Hendlich et al. (1997) para identificar bolsillos que sirven como sitios de unión para la interacción de la proteína con otras moléculas.

## **RESULTADOS**

### **Análisis de la estructura primaria**

La secuencia primaria obtenida en el NCBI (2) contiene 779 a.a. Los resultados de ProtParam (5) arrojan que la proteína posee un punto isoeléctrico (pI) de 5.54, un coeficiente de extinción de 28140 (Los coeficientes de extinción están dados en unidades de  $M^{-1} cm^{-1}$ , a 280 nm medido en agua) y un peso de 86782.2 gr. El índice de inestabilidad corresponde a 65.11, lo que indica que la molécula es altamente inestable.

En cuanto a su composición, los aminoácidos predominantes corresponden a Leucina (Leu) y Serotonina (Ser) con porcentajes de 12.3% y 11.6% respectivamente. La molécula posee un total de residuos cargados negativamente (Ala+Glu) de 104 y cargados positivamente (Arg+Lys) de 85. Los perfiles obtenidos a partir de la herramienta ProtScale (5) mediante el algoritmo de Kyte-Doolittle se representan en una gráfica (Figura 1) el cual emplea una escala de hidropaticidad para cada uno de los aminoácidos, con variaciones entre -3,5 (Glu) y 4,2 (Val)(6).

En la gráfica, se observa que las zonas con mayor hidrofobicidad se encuentran en las posiciones 762 a 764 con índices entre -2689 a -2.478 y en las posiciones 305 a 308 con índices entre -2456 y -1556, representando un menor valor. Por otro lado, las zonas altamente hidrofílicas se encuentran entre los aminoácidos 328 a 331 con valores entre

1233 y 1911, representando un valor más alto y en los aminoácidos 83 a 85 con valores entre 1533 y 1022 (5).

Al realizar el alineamiento local en BLAST(2), se detectó una región conservada identificada con el nombre de HAP1\_N y dominios conservados putativos para la super familia MILTON; a partir de este resultado, se decidió primero encontrar la familia a la cual pertenece la proteína CAD89935.1 mediante el uso de la herramienta PFAM(8), lo que corroboró que la proteína presenta una región HAP1\_N ubicada entre los aminoácidos (a.a) 1 al 259 aproximadamente. Se encontraron dos regiones identificadas para la familia MILTON, la primera región ubicada entre los aminoácidos 138 a 473 y la segunda, entre los aminoácidos 591 a 753.

El resultado de BLAST(2) permitió también la identificación de 100 secuencias de a.a con un query cover por encima de 95%, en donde las secuencias se presentan para primates y otros animales; a partir de ello, se realizó un alineamiento múltiple mediante el servidor T-Coffee (7), para ello, se tomó la secuencia FASTA de la proteína CAD89935.1 (Homo sapiens) junto con 15 secuencias FASTA de organismos diferentes.

Las 15 secuencias utilizadas se encontraron pertenecen a *Trichechus manatus latirostris* (Manatí), *Dasyus novemcinctus* (Armadillo), *Orcinus orca* (Orca), *Myotis brandtii* y *Pteropus alecto* (Murcielagos), *Microtus ochrogaster* y *Rattus norvegicus* (roedores), *Mustela putorius furo* (Hurón), *Alligator sinensis* (Caimán), *Chrysemys picta bellii* (Tortuga), *Anolis carolinensis* (Iagartija), *Ficedula albicollis* y *Zonotrichia albicollis* (aves paseriformes) y *Anas platyrhynchos* (ave anseriforme). Los resultados de T-Coffee (7) muestran también la presencia de la región HAP1\_N.

De la base de datos de String (12) al colocar el nivel de confianza medio mayor 0.4 se tuvo que la proteína hipotética presentó homología con TRAK1, co-expresión con STRADB y relación con 6 seis proteínas más, entre ellas HGS, KCNJ2 y OGT.

#### **Análisis estructura secundaria.**

A partir del consenso realizado por Phyre (9) se encontró que el 4% de la secuencia pertenecen a coils, 50,8% a hélices  $\alpha$  y 45,1% a cadenas  $\beta$ . En la Figura 2 se puede observar su distribución en la cadena, mostrando en gris las zonas de coil, en rojo hélices  $\alpha$  y en azul cadenas  $\beta$ .

#### **Análisis estructura terciaria.**

La predicción de la estructura terciaria de la proteína CAD89935.1 se obtuvo mediante el programa meta-servidor LOMETS(10), en donde realiza el modelamiento a partir del alineamiento de fragmentos de la secuencia FASTA de la proteína sobre plantillas estructurales con pliegues que se han reportado en bases de datos biológicas (15). En los resultados de LOMETS, se encogieron 2 modelos con base en la obtención de un alto Score confidence, el Coverage y por la longitud de alineamiento (Tabla 1).

La validación del modelo protéico de CAD 89935.1 se obtuvo mediante el servidor de MolProbity, en donde se tiene el Plot de Ramachandran. También se realizó una predicción de la función de la proteína mediante la herramienta informática CombFunc (15) (Tabla 2), lo que indicó con una probabilidad de 0,961 que la proteína está asociada a un receptor GABA.

El alineamiento 3D se realizó con la proteína 1awc que tenía una analogía estructural con un T-M score de 0.884. I

### **Predicción de sitios de unión**

Mediante el servidor 3DLigandSite (13), se colocó el archivo pdb de la proteína 1wa5C, para identificar los sitios de unión con los que interactúa la proteína con otras moléculas, a partir de esto, se obtuvo un sitio de unión junto con un ligando que se encuentra en interacción. (Figura 3), en donde los aminoácidos Lisina (posición 254), Glutamina (posición 219) y Arginina (posición 163) se encuentran en dicha interacción.

En la figura visualizada en Pymol (16), se indican también las distancias de las cargas polares con el ligando; en donde para Arginina corresponde una distancia de 2,6 Amstrongs; para Glutamina presenta 2,3 Amstrongs y para Lisina presenta 1,9 Amstrongs. Finalmente, mediante Pocket-Finder (14) por medio de la opción de CHIME, se encontró 10 cavidades, en donde se encuentra una región comprendida entre los residuos 121 hasta el 256; dentro de estos residuos se pueden encontrar los residuos Arginina en la posición 163, Glutamina 219 y Lysina 254 y esto apunta a que probablemente estos sitios de residuos identificados podrían corresponder a los sitios de interacción con el ligando encontrado.

De la base de datos de String (12) al colocar el nivel de confianza medio mayor 0.4 se tuvo que la proteína hipotética presentó homología con TRAK1 (MILTON), co-expresión con STRADB y relación con 6 seis proteínas más, entre ellas HGS, KCNJ2 y OGT.

### **CONCLUSIONES**

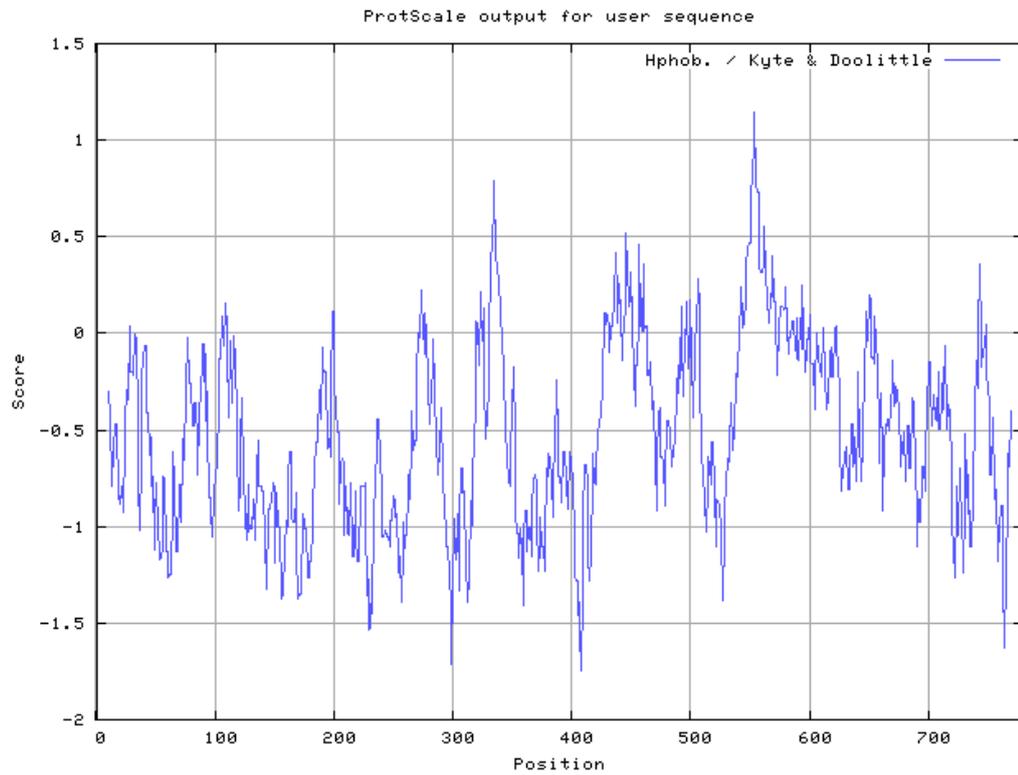
El modelo tridimensional obtenido coincide con la funcionalidad de las regiones encontradas en la proteína que pertenece a la familia MILTON y posee una región conservada HAP1\_N, y teniendo en cuenta su alta hidrofobicidad y la predicción funcional realizada por CombFunc (15), es posible que se trate de un regulador tipo inhibidor de la neurotransmisión y puede estar asociada a enfermedades neurodegenerativas como la esclerosis lateral amiotrófica y el Huntington.

## BIBLIOGRAFÍA

1. Rodriguez A, Clavijo D, Mejía R y col. "Bases biológicas y patobiológicas humanas de la esclerosis lateral amiotrofica". Universitas Médica. 2006 .Vol 47
2. National Center of Biotechnology Information (2013) [en línea], disponible en <http://www.ncbi.nlm.nih.gov/>. Consultado por última vez el 28/11/2013
3. Morfini G, Burns M, Stenoien D, Brady S. "Axonal Transport" American Society for Neurochemistry 2012. Published by Elsevier Inc. All rights reserved.
4. Hirokawa N and Noda Y. "intracelullar transport and Kinesin superfamily proteins KIFs: structure, function and dynamics.
5. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel RD, Bairoch A, Protein identification and analysis tools on the ExpASy Server.
6. Kyte J, Doolittle RF. Amino acid scale: Hydropathicity. J. Mol. Biol. 1982 ;( 157)
7. Poirot O, O'Toole E, Notredame C. Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res. 2003 Jul 1;31(13):3503-6. PMID: 12824354 [PubMed - indexed for MEDLINE]
8. The Pfam protein families database: M. Punta, P.C. Coghill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn. Nucleic Acids Research (2012) Database Issue 40:D290-D301
9. Kelley LA, Sternberg JE. Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc. 2009.
10. S. Wu, Y. Zhang. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Research 2007; 35: 3375-3382
11. Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.

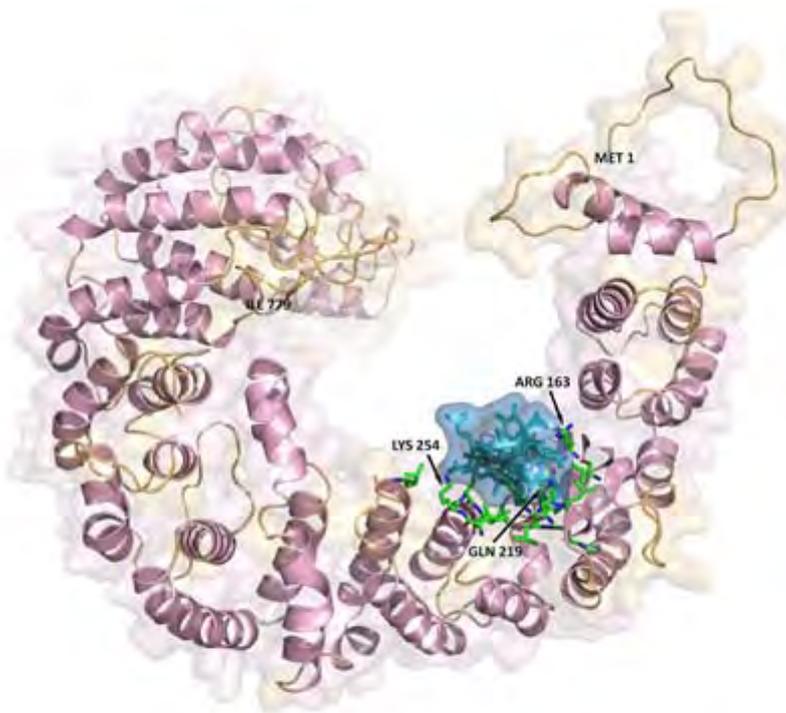
12. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. "String v9.1: protein-protein interaction networks, with increased coverage an integration" 2013 Jan;41(Database issue):D808-15. doi: 10.1093/nar/gks1094. Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Switzerland.
13. Wass MN, Kelley LA & Sternberg MJ (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. NAR 38 Suppl:W469-73.
14. Petrucci, R; Hardwood, W; Herring, G. 2006. General Chemistry: Principles and Modern Applications. 9th Edition. Prentice Hall.
15. Wass MN, Sternberg MJE (2008) ConFunc--functional annotation in the twilight zone. Bioinformatics 24: 798–806
16. Schrödinger (2013) Pymol (versión 1.5)
17. Koutspoulos O, Laine D, Chudakov D, Parton R, Frazier A, Ryan M. "Human Mitons associate with mitochondria and induce microtubule-depend remodeling of mitochondrial networks" Biochimica et Biophysica Acta 1803 (2010) 564–574
18. Saxton W and Hollenbeck P. "The axonal transport of mitochondria"Journal of Cell Science 125, 1–1. 2012. Published by The Company of Biologists Ltd. doi: 10.1242/jcs.053850
19. McGuire J, Rong J, Li S, Li X. "Membrane transport, structure, function an biogenesis: Interactin of Huntin-associated Protein-1 with Kinesin Light chain, implications in intracellular trafficking in neurons" doi: 10.1074/jbc.M509806200 originally published online December 8, 2005
20. Creighton, T. 1997. Proteins: Structures and Molecular Propieties. 2th Edition. European molecular Biology Laboratory. W. H. Freeman and Company. New York.
21. Petrucci, R; Hardwood, W; Herring, G. 2006. General Chemistry: Principles and Modern Applications. 9th Edition. Prentice Hall.

## ANEXOS

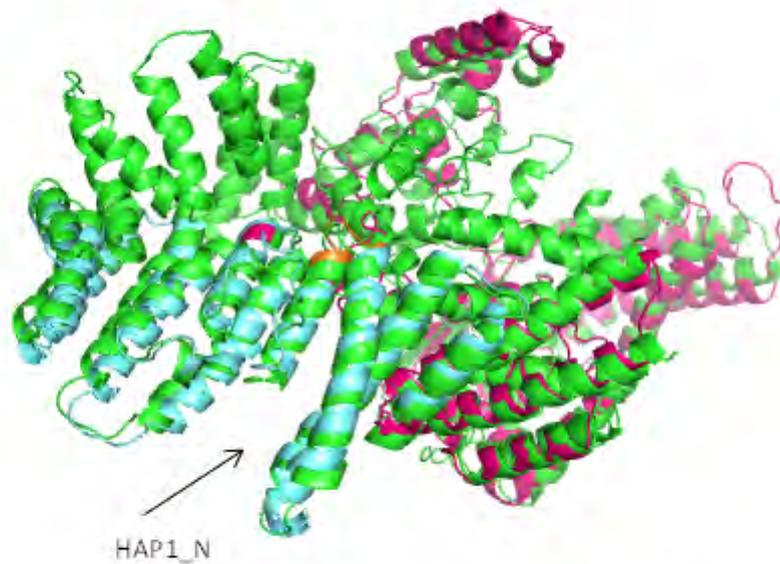


**Figura 1.** Perfiles de hidrofobicidad y polaridad obtenidos mediante la herramienta bioinformática ProtScale (5). Hidrofobicidad según algoritmo de Kyte-Doolittle (6) (escala de -3.5 a 4.2, en el eje de las ordenadas y posición de los aminoácidos en el eje de las abscisas).





**Figura 3.** Visualización en Pymol (16) del modelo obtenido computacionalmente. Se muestra en morado la proteína y en azul el ligando en interacción



**Figura 4.** Alineamiento en 3D de la proteína. 4j SPb (verde) y CAD8935.1 (Azul), se visualiza una región altamente conservada HAP1\_N)

Resultados de LOMETS			
Template	Align_lenght	Coverage	Confidence Score
1wa5C	730	0,937	High
4kf7A	744	0,955	High

**Tabla 1.** Resultados obtenidos a partir de la herramienta LOMETS (10), se seleccionaron 2 modelos con base en los valores del Score confidence, Coverage y la longitud de alineamiento.

Gen Ontology			
GO Term	Descripción	Número	SVM Probabilidad
GO: 0050811	Receptor de GABA de unión	10	0,961
GO: 0006493	aminoácidos de la proteína de glicosilación ligada a O	10	0,728
GO: 0006357	regulación de la transcripción del promotor de la ARN polimerasa II	10	0,728
GO: 0006605	focalización de proteínas	10	0,649
GO: 0006836	transporte de neurotransmisores	10	0,344

**Tabla 2.** Resultados de predicción funcional realizada en el programa CombFunc (15). Los colores rojo, naranja y amarillo representan probabilidades altas, medias y bajas respectivamente.