

PREDICCIÓN DE ASPECTOS ESTRUCTURALES Y FUNCIONALES DE LA PROTEÍNA NO CARACTERIZADA C10orf25 EN *Homo sapiens*

Juan Fernando Carrascal, Dennis Castillo, Laura Viviana Soto, Daniel Felipe Bohorquez, Sebastián Herrera, George Barreto, Janneth González¹

janneth.gonzalez@javeriana.edu.co¹

RESUMEN

Objetivo. Caracterizar y predecir la estructura terciaria de la proteína C10orf25 en *Homo sapiens* analizando las posibles regiones funcionales implicadas con la Enfermedad de Alzheimer (EA). **Materiales y métodos.** Se realizó un análisis computacional de la proteína hipotética a partir de los servidores BLAST, PROTPARAM, PROTSCALE, STRING, SWISS-Model, Predictprotein, InterProScan, EXPASY, RAMPAGE, y análisis de acoplamiento. **Resultados.** Se hizo la predicción de las estructuras secundarias, terciarias e interacciones de la proteína C10orf25; Se generaron las interacciones de la proteína hipotética que podrían tener una función y papel importantes en los procesos celulares. Se hizo el análisis de acoplamiento **Conclusiones.** La proteína hipotética C10orf25 puede estar involucrada en las primeras etapas de la fase degenerativa del Alzheimer principalmente en el sistema motor y en los procesos cognitivos, sin embargo requiere mayor investigación para el conocimiento y posibles tratamientos de esta enfermedad.

INTRODUCCIÓN

La enfermedad de alzheimer (EA) es la causa de demencia degenerativa primaria más frecuente en el mundo, dado que puede llegar a perjudicar aproximadamente a 56 pacientes de cada 1000 personas al año (Kukull et al 2002). Actualmente se cree que la EA afecta a cerca de 35 millones de personas en todo el planeta y se estima que para el año 2030 esta cifra aumentará a 115 millones de enfermos (Dartigues 2009). Dentro de las graves consecuencias de la EA se encuentra la alteración de las regiones de hipocampo, lo cual deteriora la memoria (Braak H & Braak E. 1991), daños en el metabolismo de la proteína precursora amiloide (APP) así como en varios neurotransmisores (dopamina, serotonina, adrenalina, acetilcolina, y glutamato), y la generación de un conjunto de reacciones

inflamatorias, oxidativas y hormonales (Gazulla & Cavero-Nagore 2006). Dado lo anterior, la EA dificulta el nivel funcionamiento normal del paciente y la ejecución de trabajos o actividades sociales habituales (Knopman, Boeve & Ronald 2003).

Los principales factores de riesgo de la EA son la senectud, esencialmente en personas con edades superiores a los 60 años y la historia familiar de la enfermedad, siendo los genes relacionados aquellos que codifican para la proteína precursora de péptido Beta- amiloide (APP) los genes de preselininas (PSEN1 y PSEN2), y el gen de la apolipoproteína E (APOE). A pesar de que se conocen estos y otros genes implicados con la EA, aún no se sabe con total certeza su arquitectura genética debido principalmente a la ausencia de estudios en lo concerniente a la epigenética y la regulación de la expresión (Seto-Salvia & Clarimón 2010).

Dado lo anterior, el conocimiento de las bases moleculares de la enfermedad es trascendental para el desarrollo de nuevas propuestas terapéuticas que corrijan la EA en estadios previos a su formación (Fernandez-Valencia et al 2009). Con base en esto, el objetivo del artículo fue caracterizar y predecir la estructura terciaria de la proteína C10orf25 en *Homo sapiens* analizando las posibles regiones funcionales implicadas con la EA

MATERIALES Y MÉTODOS

Análisis computacional de la estructura primaria de C10orf25

La secuencia de la proteína hipotética no caracterizada C10orf25 (Acc. No. gi |282847365| ref| NP_001034469.2|) fue obtenida a través de GenBank del Centro Nacional de Información para Biotecnología (NCBI). Los análisis de similitud fue contrastada con otras secuencias del Banco de Datos de proteínas (PDB) a través de BLASTP, disponible en NCBI, para la detección de regiones homólogas

Las propiedades fisicoquímicas, como el punto isoeléctrico (pI), el peso molecular, la composición de aminoácidos, el índice de estabilidad (IE), el índice alifático, el tiempo de vida y el promedio general de hidropatía, PROTPARAM (<http://www.expasy.org/ tools/>). Para el análisis del índice de hidrofobicidad se utilizó la herramienta PROTSCALE del servidor Expasy Proteomic Tools (<http://www.expasy.org/ tools/>).

Predicción de la estructura secundaria, terciaria y tridimensional de C10orf25

La predicción de la estructura secundaria y la accesibilidad al solvente se obtuvieron en base a alineamientos múltiples de secuencias y un sistema multinivel (Rost&Sander 1993) a través del servidor PredictProtein (<https://www.predictprotein.org/>). La predicción de la estructura terciaria se llevó a cabo mediante la búsqueda de dominios a través del servidor en EBI (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) empleando InterProScan 4., herramienta que utiliza 13 diferentes programas, tales como PFRAM, PROSITE, PHANTHER, GO, entre otros. De la misma forma, se utilizó el servidor ProDom (<http://prodom.prabi.fr/prodom/current/html/home.php>), una base de datos completa de dominios de familias de proteínas generados a partir de la comparación global de todas las secuencias de proteínas disponibles. (Bru *et al* 2005)

La predicción de la estructura tridimensional (3D) se elaboró a través del servidor web LOMETS (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>), mediante la recopilación de alineamientos con secuencias de alta puntuación relacionados con nueve programas instalados localmente. La validación estereoquímica y energética del modelo en 3D se realizó a través de la herramienta de modelamiento automático Swiss-Model del servidor Expasy Proteomic Tools (<http://www.expasy.org/tools/>) y el servidor Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). Por último para la visualización en 3D de la proteína hipotética se utilizó PyMOL v. 0,98 (<http://www.pymol.org/>).

La predicción de sitios activos de la proteína hipotética no caracterizada se utilizó la plataforma Q-SiteFinder (<http://www.modelling.leeds.ac.uk/qsitefinder/>) para obtener los archivos del Protein Data Bank con la predicción gráfica de los sitios activos de la proteína hipotética.

Análisis de acoplamiento

El acoplamiento proteína-proteína se obtuvo por ClusPro (Comeau *et al* 2004) (<http://cluspro.bu.edu/publications.php>), el cual es un programa web totalmente automatizado basado en CAPRI (Evaluación Crítica de Predicción de Interacciones).

Predicción de interacciones Proteína-Proteína

La predicción de interacción proteína-proteína se realizó por medio de la Herramienta de búsqueda para la recuperación de interacciones genes/proteínas, STRING (<http://string-db.org/>), base de datos que predice las interacciones entre proteínas, incluyendo asociaciones directas o físicas e indirectas o funcionales, que se derivan de cuatro fuentes: Contexto genómico, experimentos de alto rendimiento, Co-expresión (Conservado) y el conocimiento previo.

RESULTADOS

Análisis computacional de la estructura primaria de C10orf25

El alineamiento local de secuencias realizado con BLAST no presento resultados concluyentes en la búsqueda de secuencias homólogas. Las propiedades fisicoquímicas obtenidas mediante PROTPARAM 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 C10orf25 (Tabla 1). El análisis hidrofóbico de PROTSCALE (Figura 1) reveló que la proteína C10orf25 es en su mayoría hidrofílica, sin embargo presenta regiones hidrofóbicas principalmente en los primeros 22 aminoácidos.

El peso molecular calculado de la proteína C10orf25 fue 14440.8 D. El punto isoeléctrico (pI), el cual es definido como el pH o el punto en el que las cargas contrarias u opuestas de los grupos básicos y ácidos se equilibran o se vuelven neutras (Werner, 2008; Sivakumar, 2007), arrojó un resultado de 10.26 lo que revela el carácter básico de la proteína. El índice alifático que se define como el volumen relativo ocupado por las cadenas laterales alifáticas según Ikai 1980, mostró un resultado de 70.33. Dado lo anterior es posible que el valor sea alto, lo cual significa que la proteína presenta estabilidad en un amplio rango de temperaturas (Sivakumar, 2007). El índice de hidropatía es una métrica empleada para la energía [kJ/mol] la cual se usa para transferir un segmento de secuencia de longitud definida de un medio hidrofóbico a un medio hidrofílico (Muller- Esterl, 2008). Por lo tanto, un alto valor del índice de hidropatía (Figura 1) es evidencia de una región transmembranaria, mientras que valores bajos como el del promedio general (GRAVY) de la proteína C10orf25 (Tabla 1), son muestra de regiones de la proteína que están localizadas dentro de la membrana. El tiempo de vida de la proteína es de 30 horas, esto quiere decir que la proteína, luego de ser sintetizada por una célula desaparece en el transcurso de tiempo mencionado (Barreto, 2012)

Estructura secundaria y terciaria

El resultado de PredictProtein reveló una proporción de 45.08% de hélices y 54.92% de Bucles para la estructura secundaria de la proteína hipotética no caracterizada C10orf25. La accesibilidad al solvente para la proteína C10orf25 presento un 47, 54% expuesto, 12,30% intermedio y 40,16 inmersa.

El análisis realizado en InterProScan (Figura 2) reveló la presencia de un péptido señal no integrado entre los aminoácidos 1 y 28 predichos mediante SignalP, lo que significa que posteriormente se pierde. Por otra parte la predicción de dominios por el servidor ProDom revela la caracterización de una glicoproteína.

Predicción de la estructura 3D.

Las probabilidades de la predicción de la estructura 3D (Figura 3) muestran que entre los aminoácidos 5 y 21 existe una alta probabilidad de encontrar Hélices dentro de la estructura, de igual forma en los aminoácidos 42 a 45, 61 a 67, 84 a 95 y 101 a 104 (Wu & Zhang 2007)

Las probabilidades de encontrar colas en la predicción terciaria de la estructura son igualmente altas en los aminoácidos 21 a 31, 39 a 41, 49 a 51 y del 111 al 122. Caso contrario de las Hélices beta que presentan una baja probabilidad de aparición en toda la predicción 3-D, donde la mas alta probabilidad es de 0,2 entre el aminoácido 71 y 73.

Validación de la estructura 3D

La comparación de la estructura 3-D con las bases no redundantes muestra que la estructura predicha de la proteína Hipotética C10orf25 se encuentra fuera de los promedios del Z-score presentando un valor de QMEAN de -5.38 (Figura 4), lo que indica que no se encuentran homologías predicha en los set no redundantes del Protein Data Base (PDB).

El modelo adquirido por SWISS-MODEL se comparó con la proteína obtenida mediante la plataforma de LOMETS en PyMol v.99. Tanto la proteína predicha como el modelo obtuvieron estructuras similares (Figura 5). El Ramachandran plot (Figura 6), muestra que la predicción de la proteína Hipotética C10orf25 presentan un alto porcentaje en regiones permitidas (93.3%), un porcentaje en regiones favorecidas de 5.8% y un bajo porcentaje en regiones no permitidas de Glicina (0.8%) (Lovell et al 2002).

Predicción de sitios activos

La predicción de los sitios activos de la proteína hipotética C10orf25 (Figura 7) muestra que se encuentran 3 sitios activos principales de leucina, Valina y Cisteína además de otros 4 sitios activos que se encuentran al interior de la proteína hipotética y que no presentan una superficie expuesta a para la interacción. Los sitios activos representados con la malla dentro de la proteína hipotética C10orf25 muestra la superficie expuesta para la interacción de estos sitios activos.

Red de interacción proteína-proteína

Las interacciones proteína-proteína muestran que la proteína Hipotética C10orf25 (Figura 8), interactúa directamente con la proteína C1orf25; las cuales están directamente relacionadas en la coordinación motora y en el comportamiento exploratorio.

Análisis de acoplamiento

La cadena C del péptido 2OTK interactúa con la proteína Hipotética C10orf25 por su sitio activo prioritario y de mayor superficie siendo un sitio activo de Cisteína, Valina y Leucina,

el Docking flexible muestra que este es el sitio más óptimo para la interacción con el péptido 2OTK (Figura 9) (Kozakov et al 2013; Kozakov et al 2006; Comeau et al 2004; Comeau et al 2004b)

Los resultados de hidrofobicidad predichos por PROTSCALE para los primeros 22 aminoácidos son congruentes con la predicción realizada por InterProScan el cual muestra la presencia en esta misma región de un pepetido señal. Probablemente esta proteína esta relacionada con el paso de la proteína a través de una membrana (Rastogi, 2005). Lo anterior se relaciona con los resultados de PROTPARAM el cual predijo regiones de la proteína que están localizadas dentro de la membrana, lo que concuerda con los resultados de ProDom con la predicción de dominios caracterizados de una glicoproteína.

Las interacciones proteína-proteína muestran la interacción cercana de la proteína hipotética C10orf25 con la proteína hipotética C1orf25 (Fig.), la proteína C1orf25 está involucrada directamente en el sistema motor y en el comportamiento exploratorio en humanos, por lo que la proteína hipotética C10orf25, al igual que la proteína hipotética C1orf25, estén implicadas en los primeros estados del alzheimer dados sus efectos sobre los sistemas motores y cognitivos exploratorios (Sood et al 2001).

A demás de la interacción con la proteína C1orf25, la proteína hipotética C10orf25 mediante el acoplamiento mediante Dockin con el ligando 2OTK_C mostro que este ligando interactúa por el sitio activo de Cisteína, Valina y Leucina. El ligando 2OTK_C es una cadena C de un péptido de unión involucrado en el proceso del Alzheimer por lo que se puede asumir que la proteína hipotética C10orf25 puede estar involucrada en los proceso degenerativos iniciales del Alzheimer (Gronwall et al 2007).

La proteína hipotética C10orf25, al ser una proteína de bajos estudios y aparentemente bajas interacciones, no presenta homologías distinguibles con otras estructuras descritas por cristalografía o por otros métodos *in silico* (Fig.) por lo que la información de la predicción de las estructuras 3D y de las interacciones de la proteína hipotética son escasas y poco estudiadas.

CONCLUSIONES

La proteína hipotética no caracterizada C10orf25 puede estar involucrada en las primeras etapas de la fase degenerativa del Alzheimer principalmente en el sistema motor y en los

procesos cognitivos por lo que el análisis de cristalografía y predicciones de nuevas proteínas hipotéticas al ser tan escasos requieren de mayor cuidado e investigación para el conocimiento y posibles tratamiento de esta enfermedad.

BIBLIOGRAFÍA

Barreto L, Barreto G, Morales L, Acevedo O, Gonzales J. 2012. Proteína LIC10494 de *Leptospirainterrogans* serovar *Copenhageni*: modelo estructural y regiones funcionales asociadas. *Universitas scientiarum*. Pag. 1-12

Benkert, P., Schwede, T. and Tosatto, S.C.E. 2009. QMEANclust: Estimation of protein model quality by # combining a composite scoring function with structural density information. *BMC Struct Biol*. 20; 9:35.

Braak H, Braak E. 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol (Berl)*; 82: 239-59.

Bru, C. et al. 2005. The ProDom database of protein domain families: More emphasis on 3D. *Nucl. Acids Res*. 33: 212–215.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. a. 2004. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. b. 2004 ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res*. 32:W96–9.

Dartigues J. 2009. Alzheimer's disease: a global challenge for the 21st century. *Lancet Neurol*; 8: 1082-3.

Fernández-Verdecia CI, Díaz del Guante M, Castillo-Díaz L, Álvarez-Blanco J.2009 Neurogénesis como diana terapéutica para la enfermedad de Alzheimer. *Rev Neurol*; 49: 193-201.

Gazulla,J. & Cavero-Nagore,M. 2006. Glutamato y enfermedad de Alzheimer. *REV NEUROL* 42 (7): 427-432

Gronwall,C., Jonsson,A., Lindstrom,S., Gunneriusson,E., Stahl,S. and Herne,N. 2007. Selection and characterization of Affibody ligands binding to Alzheimer amyloid beta peptides. *J. Biotechnol.* 128 (1), 162-183

Knopman D, Boeve F, Ronald C. 2003. Petersen Essentials of the Proper Diagnoses of Mild Cognitive Impairment, Dementia, and Major Subtypes of Dementia. *Mayo Clin Proc.*;78:1290-1308

Kozakov D, Beglov D, Bohnuud T, Mottarella S, Xia B, Hall DR, Vajda, S. 2013. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*,

Kozakov D, Brenke R, Comeau SR, Vajda S. 2006. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins*. Aug 24

Kukull W, Higdon R, Bowen J, McCormick W, Teri L, Schellenberg G. 2002.Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol*; 59: 1737-46.

Melo, F. and Feytmans, E. 1998. Assessing Protein Structures with a Non-local Atomic Interaction Energy. *Journal of Molecular Biology* 277, 1141-1152

Muller- Esterl W. 2008. Bioquímica: fundamentos para medicina y ciencias de la vida. Editorial Reverté, Barcelona, España.Pp.352.

Rastogi S C. 2005. Cell Biology. Third Edition. New Age International Publishers p. 360

S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson. 2002. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*. 50: 437-450.

Sivakumar K, Balaji S, Gangaradhakrishnan. 2007. In silico characterization of antifreeze proteins using computational tools and servers. *Journal of Chemical Sciences* 576.

Seto-Salvia N & Clarimón J 2010. Genética en la enfermedad de Alzheimer *Rev Neurol*; 50 (6): 360-364

Sood,R., Bonner,T.I., Makalowska,I., Stephan,D.A., Robbins,C.M., Connors,T.D., Morgenbesser,S.D., Su,K., Farugue,M.U., Pinkett,H., Graham,C., Baxeavanis,A.D., Klinger,K.W., Landis,G.M., Trent,J.M. and Carpten,J.D. 2001. Cloning and characterization of 13 novel transcripts and the human RGS8 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus. *Genomics* 73 (2), 211-222

Werner MS. 2008. *Biología: Fundamentos para Medicina y Ciencias de la Vida*. Reverté, Barcelona, España. 32p.

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

ANEXOS

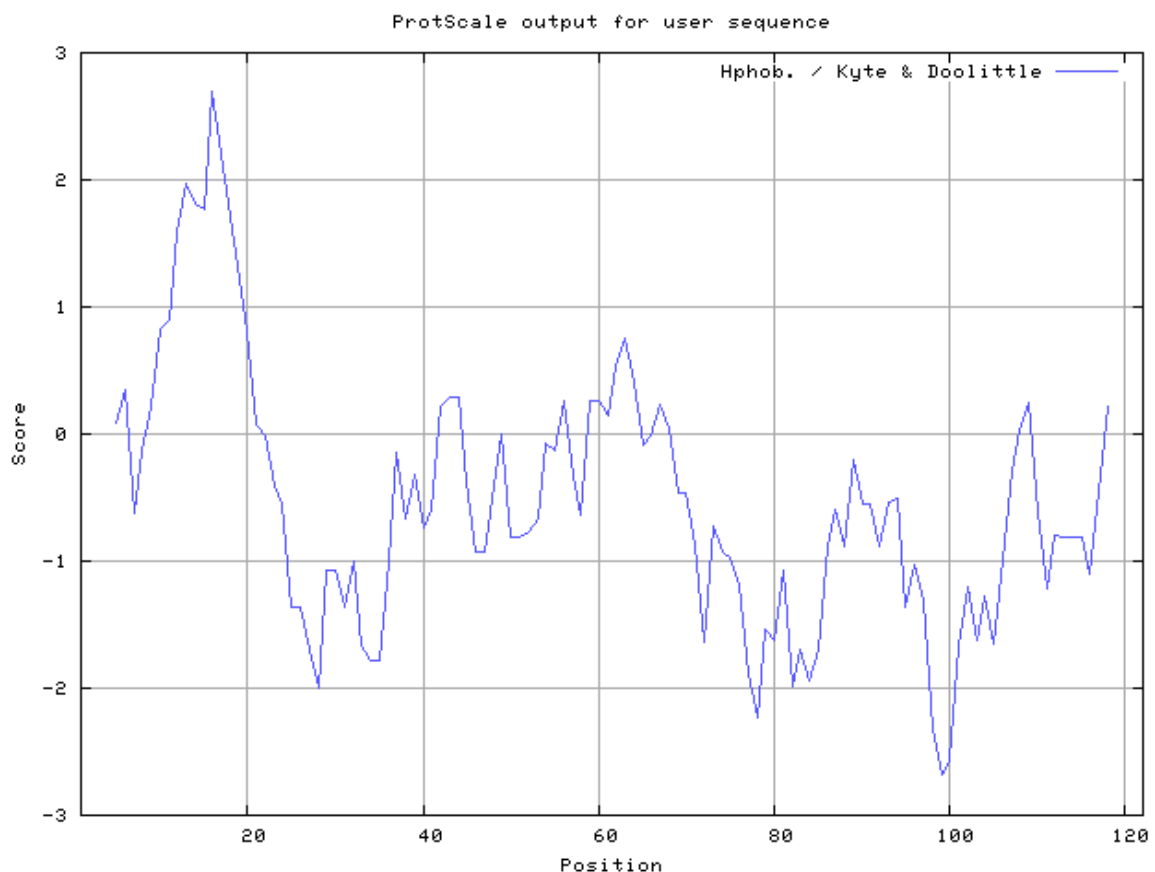


Figura 1. Análisis hidropático de Smlt1810 basándose en los valores de Kyte y Doolittle

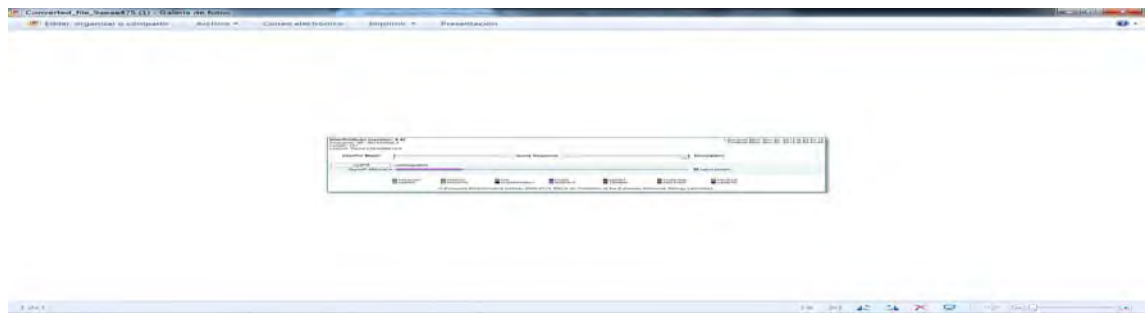


Figura 2. Ubicación del péptido señal en la proteína hipotética C10orf25 predicha por InterProScan

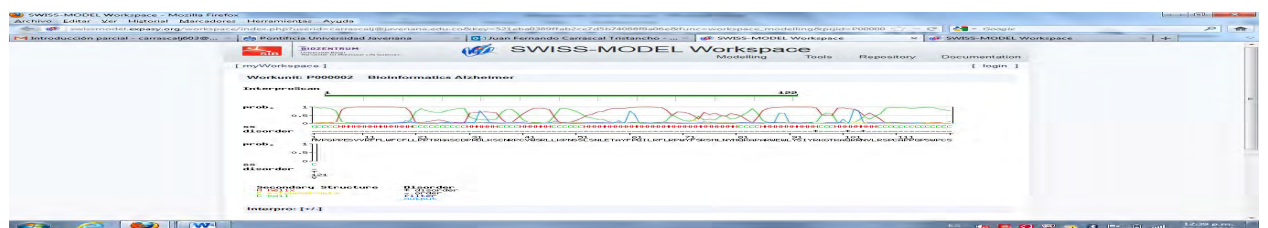


Figura 3. Probabilidades de presencia de Hélices, colas y Hélices Beta en la proteína Hipotética C10orf25, con probabilidades entre 0 y 1 a lo largo de los 122 aminoácidos que conforman la proteína hipotética.

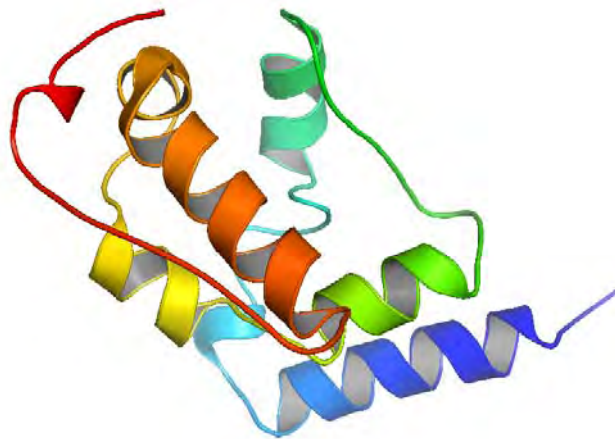


Figura 4. Modelo tridimensional generado por LOMETS y visualizado con PyMOL v,99. Las hélices de la proteína están ubicadas entre los aminoácidos 6-13; 22-24; 36-45; 62-67; 73-78; 84-102; 117-119. Validación del modelo tridimensional generado por SWISS-MODEL y visualizado con PyMOL V,0.99. Muestra que las dos

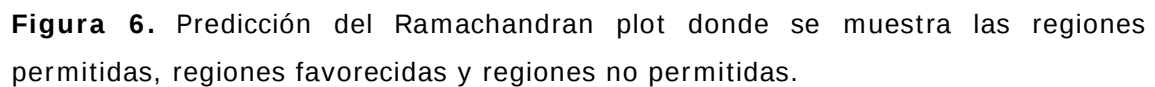
Comparison with non-redundant set of PDB structures

normalised QMEAN6 score

protein size

Z-score QMEAN: -5.38
 Z-score Cbeta: -3.20
 Z-score all_atom: -2.83
 Z-score solvation: -2.05
 Z-score torsion: -3.33
 Z-score SSE agree: -2.67
 Z-score ACC agree: -3.44

Legend:
 ■ $|Z\text{-score}| < 1$
 ■ $1 < |Z\text{-score}| < 2$
 ■ $|Z\text{-score}| > 2$
 ■ query model



PyMOL for evaluation only.
Contact sales@delsci.com.

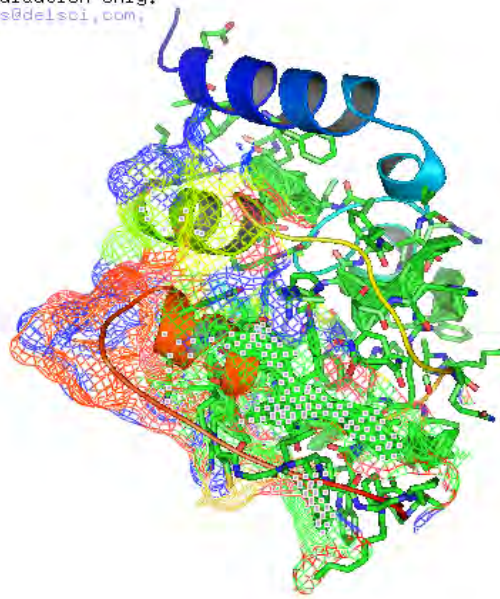


Figura 7. Predicción de los sitios activos de la proteína C10orf25 donde se observan las zonas donde se encuentran los sitios activos de Cisteína, Leucina, Valina.

PyMOL for evaluation only.
Contact sales@delsci.com.

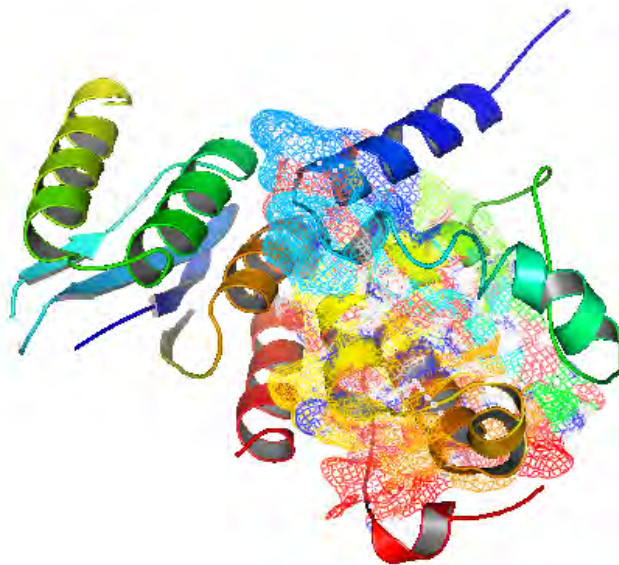


Figura 8. Interacción de la proteína hipotética C10orf25 por sus sitios activos de Cisteína, Valina y Leucina con la cadena C del péptido Ab (20TK), diseñada como proteína de unión.

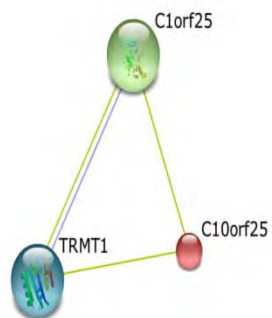


Figura 9. Red de interacción proteína-proteína de la proteína hipotética C10orf25 interactuando con sus proteínas más cercanas funcionalmente.

Tabla 1. Propiedades fisicoquímicas de C10orf25 reveladas por Protparam

Propiedades		Fisicoquímicas C10orf25
de		
Peso Molecular (Da)		14440.8
Índice de Inestabilidad		70.40
Punto	Isoeléctrico	10.26
Tiempo de Vida (Horas)	30	
Índice Alifático		70.33
Promedio General de Hidropatía	-0.432	
(GRAVY)		