

## Modelo molecular teórico del receptor serotoninérgico 5HT<sub>2A</sub> acoplado a proteína G

Rafael Eduardo Malagón Bernal<sup>1</sup>, Manuel Alejandro Fernández Navas<sup>1</sup>, Orlando Emilio Acevedo Sarmiento<sup>2</sup>

<sup>1</sup> Facultad de Ciencia y Tecnología, Departamento de Química Farmacéutica, Universidad de Ciencias Aplicadas y Ambientales, Bogotá D.C., Colombia.

<sup>2</sup> Departamento de Física, Pontificia Universidad Javeriana, Bogotá D.C., Colombia.

\*ramalagon@udca.edu.co

Recibido: 29-03-2012; Aceptado: 12-08-2012

### Resumen

**Objetivo.** Construir un modelo molecular teórico de la estructura terciaria del receptor 5HT<sub>2A</sub> de *Homo sapiens* a partir de estructuras obtenidas experimentalmente como plantillas. **Materiales y métodos.** Para la realización del modelo teórico se contempló el protocolo establecido por Ballesteros y Weinstein para la construcción del receptor acoplado proteína G, por medio de alineamiento de la secuencia de aminoácidos, perfiles de hidrofobicidad, refinamiento de bucles por restricciones espaciales, y minimización de energía con el campo de fuerza OPLS\_2005. **Resultados.** El modelo obtenido fue validado por el gráfico de Ramachandran con un 91,7% de aminoácidos dentro de los límites establecidos para ángulos phi y psi, y un RMSD de 0,95 Å con respecto a rodopsina de bovino. **Conclusiones.** Se obtuvo un modelo teórico validado, útil para realización de estudios de acoplamiento ligando-receptor.

**Palabras clave:** receptor proteína G, perfil hidrofobicidad, gráfico Ramachandran, enlace ortoestérico, modelaje molecular.

### Abstract

**Theoretical molecular model of the G protein-coupled 5HT<sub>2A</sub> serotonergic receptor.** **Objective** Build a theoretical molecular model of the tertiary structure of the *Homo sapiens* 5HT<sub>2A</sub> receptor from experimentally obtained structures as templates. **Materials and methods** In the construction of the theoretical model we considered the protocol established by Ballesteros and Weinstein for the construction of the G-protein coupled receptor, by the alignment of the amino acid sequence, hydrophobicity profiles, refinement of loops by spatial restrictions and energy minimization with the force field OPLS\_2005. **Results** The resulting model was validated by the Ramachandran plot with 91.7% of amino acids within the limits set for angles phi and psi and a RMSD of 0.95 Å with respect to bovine rhodopsin. **Conclusions** We obtained a validated theoretical model useful in studies of ligand-receptor docking.

**Key words:** G protein receptor, hydrophobicity profile, Ramachandran plot, orthosteric site, molecular modelling.

### Resumo

**Modelo molecular teórico do receptor serotoninérgico 5-HT<sub>2A</sub> acoplado à proteína G.** **Objetivo.** Construir um modelo molecular teórico da estrutura terciária do receptor 5HT<sub>2A</sub> de *Homo Sapiens*, com estruturas obtidas experimentalmente como moldes. **Materiais e métodos.** Para a elaboração do modelo teórico se utilizou o protocolo estabelecido por Ballesteros e Weinstein para a construção do receptor acoplado à proteína G, por intermédio de alinhamento da seqüência de aminoácidos, perfis de hidrofobicidade, refinamento de bucles por restrições espaciais e minimização da energia com o campo de força OPLS\_2005. **Resultados.** O modelo obtido foi validado pelo gráfico de Ramachandran com 91,7% dos aminoácidos dentro dos limites estabelecidos para os ângulos phi e psi, e um RMSD de 0,95 Å com respeito à rodopsina de bovino. **Conclusões.** Obteve-se um modelo teórico validado, útil para a realização de estudos de acoplamento ligante-receptor.

**Palavras-chave:** receptor proteína G, perfil de hidrofobicidade, gráfico Ramachandran, união ortoestérica, modelo molecular.

## Introducción

En la búsqueda constante por identificar nuevas moléculas de fuentes naturales o sintéticas, que generen actividad terapéutica, cumpliendo un papel agonista o antagonista frente a los procesos mediados por receptores, la identificación del blanco molecular se plantea como una de las primeras etapas en la terapia dirigida a blanco, además de la evaluación de su posible sitio de interacción, donde se generará el efecto y las características de unión de la nueva molécula con su blanco terapéutico.

Para entender el proceso de unión del ligando con el receptor se hace indispensable conocer el rol que desempeña el ligando con los diferentes blancos biológicos en la fisiología y bioquímica. Dentro de los diferentes blancos existen proteínas con funciones de: estructura de la membrana celular, mediadores de electrolitos como canales iónicos, sistemas enzimáticos que al estimularse catalizan reacciones, o también aquellos que modifican la producción y/o estructura de diversas proteínas modificando los procesos de transcripción y síntesis proteicas (1). Entre estos blancos existen los sistemas de proteínas G, los cuales pueden ejercer todas las funciones mencionadas anteriormente mediante la estimulación puntual de sus blancos moleculares activos. Estos sistemas son de interés farmacológico debido a que son sitios clave donde los fármacos pueden interactuar con el organismo. La acción final del ligando como fármaco se ve mediada por la activación de la proteína G generando segundos y terceros mensajeros (1). Por tal razón, el conocimiento de la estructura tridimensional es vital si se desea comprender a fondo el mecanismo de interacciones moleculares y así formular un posible diseño racional de fármacos de nuevas moléculas que ligadas a su receptor ejerzan una actividad terapéutica deseada.

En la determinación experimental de proteínas existen tres métodos para la elucidación de estructuras terciarias y cuaternarias con resoluciones que permiten identificar las posiciones de las moléculas que lo conforman tales como lo son la resonancia magnética nuclear (RMN), la difracción de rayos X y la criomicroscopía con difracción de electrones; sin embargo, éstos presentan problemas puntuales como: La técnica de RMN, que se focaliza en su baja resolución al determinar estructuras de proteínas mayores a 35 KDa; la cristalografía con difracción de rayos X, que requiere en primer lugar, cristales con un alto grado de ordenamiento lo cual implica una gran dificultad en el caso de proteínas de membrana, debido a las fuertes interacciones hidrofóbicas presentes en las proteínas y la necesidad de trabajar en medios con agentes tensoactivos; la criomicroscopía con difracción de electrones, produce imágenes que se obtienen de promediar varias imágenes independientes; para el caso de

proteínas de membrana éste último se ha utilizado haciendo uso de cristales bidimensionales, el problema se define en su baja resolución (2).

Hoy en día, las estructuras tridimensionales obtenidas experimentalmente y publicadas se pueden obtener de la base de datos de Brookhaven conocida como PDB (Protein Data Bank), donde diariamente surgen investigaciones con nuevas estructuras; en la actualidad poseen 79356 (Febrero de 2012) (3). No obstante, la gran mayoría de estructuras que se encuentran en la base de datos de PDB corresponden a proteínas inmersas en medios acuosos y tan solo una porción de ellas se refiere a proteínas que se encuentran dentro de una membrana, lo cual dificulta el estudio de interacciones de algunos principios activos con sus blancos terapéuticos. Es por esto que, al no disponer de una estructura obtenida experimentalmente, se acude a técnicas de modelaje molecular con las cuales se pretende obtener un modelo teórico de la estructura de la proteína o receptor validado para el estudio de interacciones moleculares aplicado al diseño de nuevos fármacos, esto conlleva a los métodos de predicción de proteínas como alternativa (4).

Hasta 2007 la estructura de rodopsina de bovino (1U19-1GZM), fue la única proteína disponible para la realización de modelos homólogos de receptores acoplados a proteína G (GPCR); hoy en día se dispone de más estructuras de GPCR, tales como el receptor  $\beta_2$  adrenérgico de humano (2RH1), rodopsina de calamar (2Z73), receptor  $\beta_1$  adrenérgico de pavo (2V74), opsina de bovino (3CAP) y receptor de adenosina  $A_{2A}$  de humano (3EML), las cuales dan amplia información de la estructura y diferencias entre rodopsina de bovino con respecto a las otras clases de GPCR subtipo A, particularmente en la orientación y posición de las hélices transmembranales y en la región de la estructura de los bucles (5). El uso de rodopsina de bovino fue planteado por Ballesteros y Weinstein en 1995, donde se contempla el uso del mapa de densidad electrónica de rodopsina de bovino como topología transmembranal, identificando aminoácidos conservados en la familia de GPCR subclase A (6).

El modelaje de proteínas G acopladas a receptor es un proceso incierto, por lo tanto, es importante evaluar la calidad del modelo obtenido para evitar errores en la conformación obtenida debidos a malos alineamientos, dificultad en el modelamiento de regiones de bucles, así como también en el proceso de refinamiento utilizado. La calidad estereoquímica del modelo puede ser evaluada mediante gráficos de Ramachandran, los cuales nos indican la correcta posición de los aminoácidos de acuerdo al estándar de los ángulos phi y psi de éstos para la formación de hélices alfa ( $\alpha$ ), hojas Beta ( $\beta$ ) y bucles (7). Es así que el modelo es válido para estudios biomoleculares y/o aceptado con un 90 % de

sus aminoácidos en regiones permitidas según el gráfico de Ramachandran (5, 7, 8).

Entre los receptores acoplados a proteínas G se encuentran los receptores de serotonina, estudios realizados han identificado numerosos receptores de serotonina en el sistema nervioso central (SNC), desde el 5HT<sub>1</sub> hasta 5HT<sub>7</sub>. Los receptores de serotonina difieren en su función, en su respuesta a diversos compuestos, en sus efectos sobre el sistema de segundo mensajero o en su capacidad excitatoria o inhibitoria; en específico, los receptores 5HT<sub>2A</sub> se expresan en grandes cantidades en diversas regiones corticales, estos receptores son esenciales para la acción de la serotonina 5HT en un gran número de procesos del SNC en donde se ven involucrados la regulación de la conducta, la agresión, estado de ánimo, la modulación de la percepción, el dolor y la ansiedad, entre otros procesos fisiológicos. Debido a que las drogas alteran la conducta, se ha pensado que dichas alteraciones son producto de una modificación en la comunicación intraneuronal; gran cantidad de psicofármacos, incluyendo antipsicóticos atípicos, antidepresivos y ansiolíticos, actúan sobre 5HT<sub>2A</sub>, en ello radica la importancia de conocer el mecanismo por el cual se pueda generar un efecto agonista o antagonista para regular la expresión de dicho receptor con respecto a aminas biogénicas en enfermedades que involucran trastornos psiquiátricos; por tal motivo se hace relevante la investigación y diseño de nuevos fármacos con el fin de brindar ayuda a aquellos que padecen este tipo de patologías. (9, 10, 11). Basados en la información obtenida, el blanco terapéutico central a elucidar teóricamente en la presente investigación es la estructura terciaria del receptor 5HT<sub>2A</sub> por medio de métodos computacionales validados.

En este trabajo se ha construido un modelo teórico del receptor 5HT<sub>2A</sub>, utilizando múltiples alineamientos de la secuencia con estructuras  $\alpha$  y  $\beta$  adrenérgicas obtenidas experimentalmente como plantillas, fusión de dos propuestas de modelos basados en su estructura y sometido a su respectiva validación.

## Materiales y métodos

Para el modelo teórico se partió de la identificación de la secuencia de aminoácidos del receptor 5HT<sub>2A</sub> de *Homo sapiens*. (NCBI Reference Sequence: NP\_000612.1 PUBMED <http://www.ncbi.nlm.nih.gov/pubmed>), se abordó la técnica de topología de membrana debido a que ésta contempla la realización de modelos homólogos con un porcentaje de identidad inferior al 30% entre la secuencia en estudio con respecto a estructuras obtenidas experimentalmente (4, 5, 6, 8)

## Alineamiento de la secuencia de aminoácidos

Se identificaron los residuos más conservados usando el esquema numérico establecido por Ballesteros y Weinstein, (Tabla 1) (6, 12, 13, 14). Para la generación del modelo se empleó como herramienta el servidor de predicción de proteínas I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/> [Fecha de cálculo: 09/03/2011]) del cual se ha comprobado su eficacia en los experimentos de CASP7, CASP8, CASP9 (15,16). Para iniciar el proceso de modelado de la secuencia de aminoácidos se registra en el servidor, donde se establecen restricciones en los aminoácidos CYS 110 y CYS 187 en la estructura de rodopsina de bovino (1GZMA) como posible plantilla propuesta desde el usuario al servidor para conservar el puente disulfuro. El servidor contempla alineamientos múltiples con diferentes plantillas para la construcción de un esqueleto base (Tabla 2), conformado por carbonos  $\alpha$  y sus cadenas laterales que vienen de la identificación de regiones o sectores homólogos de la secuencia con las plantillas y que tomará su forma y estructura por la superposición de la secuencia con respecto a las regiones identificadas de estructuras obtenidas experimentalmente y las regiones no alineadas con estructuras experimentales son calculadas por métodos ab initio (especialmente las regiones de los bucles) estructuradas por intercambio de simulaciones de Monte Carlo; 160 modelos son generados por 8 servidores contenidos en LOMETS -con sus algoritmos propios de cálculo de probabilidad cada servidor- (herramienta de I-TASSER) donde cada servidor genera 20 modelos ordenados por sus Z-scores en cada algoritmo. Los diez mejores modelos se escogen de acuerdo a una función de puntuación con respecto a su geometría, desviación estándar y diferencia entre el cálculo de la estructura nativa y el cálculo de la estructura propuesta (17).

También en este protocolo se contempló el análisis de perfiles de hidrofobicidad para identificar regiones inmersas en la membrana y regiones expuestas a un medio hidrofílico. De los cinco modelos obtenidos por I-TASSER del receptor 5HT<sub>2A</sub> los cuales provienen de alineamientos múltiples (Tabla 2), se evaluó su conformación estereoquímica por medio de gráficos de Ramachandran y se fusionaron los modelos 1 y 2 manteniendo en la fusión la estabilidad del puente disulfuro (Maestro, versión 9.1, Schrödinger, LLC, New York, NY, 2010.), (18).

## Refinamiento de bucles

Se consideró la predicción de los bucles (bucle del citoplasma CL, bucle Extracelular EL) debido a que sus posiciones conformacionales dentro del gráfico de Ramachandran se

encontraban en regiones de alta energía en el medio siendo no permitidas para las posiciones estándar de los ángulos phi y psi (7), usando el método de restricciones espaciales propuesto por Andrei Fiser y colaboradores en el 2003, contenido en el programa ModLoop-MODELLER (<http://modbase.compbio.ucsf.edu/modloop/> [Fecha de Cálculo: 15/03/2012]) (19), se refinaron bucles extracelulares (BE1 137-147, BE2 211-227A 227-235B, BE3 349-360) y bucles citoplasmáticos (BC1 101-108, BC2 177-189, BC3 256-315) con una longitud máxima de 12 aminoácidos por segmento a excepción del bucle extracelular No 2 el cual contiene 24 aminoácidos y la cisteína presente en este bucle es la que enlaza a la hélice transmembranal No 3 formando el puente disulfuro. Se adicionaron protones en la estructura utilizando la herramienta de MOLPROBITY previo análisis de gráficos de Ramachandran.

### Minimización de energía

El proceso de minimización se llevó a cabo luego de haber obtenido una estructura con arreglos conformacionales y/o refinamiento de ésta, con el fin de disminuir los impedimentos estéricos de la estructura de la proteína obtenida. La minimización de energía del receptor se realizó por mecánica molecular, utilizando como campo de fuerza OPLS-2005 (20) en 100 pasos con radio de corte de 10 Å para interacciones no enlazantes, el cual permite arreglar la proteína haciendo un reordenamiento espacial a su geometría disminuyendo así la energía en su conformación; lo que finalmente se busca es que más del 90% de los aminoácidos de la estructura se encuentren en regiones permitidas, en la evaluación por gráficos de Ramachandran (<http://molprobity.biochem.duke.edu/> [Fecha de cálculo: 21/03/2011]) (5, 6, 7, 21).

### Resultados y discusión

La construcción del Modelo 3D Teórico de la estructura terciaria del receptor 5HT<sub>2A</sub> contempla la topología de membrana de rodopsina de bovino y múltiples alineamientos. Los aminoácidos más conservados de la secuencia en estudio y de las plantillas son parte esencial para la identificación de las regiones biológicamente más conservadas y homólogas en las proteínas G acopladas a receptor. De acuerdo con los alineamientos propuestos por el servidor se observa la conservación de los aminoácidos establecidos según Ballesteros y Weinstein, incluyendo el puente disulfuro, indicando que las plantillas seleccionadas se acomodan para la construcción del esqueleto base utilizando estructuras experimentales de última generación las cuales han mejorado su resolución (**Tabla 1**).

Generalmente, el primer paso en el reconocimiento de estructuras es la búsqueda sistemática por algoritmos heurísticos, los cuales se generan por el programa BLAST con la base de datos del PDB, mostrando las plantillas que representan resultados significativos en su búsqueda, identificando sus porcentajes de identidad y las más homólogas de acuerdo al logaritmo empleado. Sin embargo, en estas búsquedas se encontró que estas estructuras presentan el inconveniente que su identidad está focalizada a algunas regiones de la secuencia problema no cubriendo en su totalidad los 471 aminoácidos del receptor (5HT<sub>2A</sub> *Homo Sapiens*); o cuando cubre gran parte de la secuencia problema su identidad se encuentra por debajo del 30% (alrededor de 21 %), dificultando realizar una superposición, ya que estructuralmente se estaría representado la plantilla (6); también en algunos casos no conservan los aminoácidos propuestos por Ballesteros y Weinstein ni el puente disulfuro, es así que empleando la técnica de múltiples alineamientos de secuencias se escogen las proteínas que contengan mayor estructura y que conserven los aminoácidos de acuerdo al esquema establecido por Ballesteros y Weinstein, para tomar de cada una de ellas un segmento de la estructura experimental y acoplarla a la secuencia en estudio (receptor 5HT<sub>2A</sub>) (**Figura supl. A**), hasta cubrir gran parte de la secuencia problema, para dar un acercamiento a la representación teórica con respecto a estructuras experimentales. Debido a que estas proteínas G comparten características estructurales evolutivas, se propone el uso del mapa de densidad electrónica de rodopsina --la cual ha sido preestablecida por años como plantilla de proteínas G, para el estudio de proteínas transmembranales debido a su dificultad en la extracción (8)-- basándose principalmente en la conservación de dominios específicos para cada hélice transmembranal como huella dactilar y patrón evolutivo (**Tabla 1**), (5, 6, 12, 13, 14).

De acuerdo con los alineamientos propuestos por el servidor, calculados con el algoritmo TM-ALIGN, se observa la conservación de los aminoácidos propuestos por Ballesteros y Weinstein incluyendo el puente disulfuro, indicando que las plantillas seleccionadas se acomodan para la construcción del esqueleto base utilizando estructuras experimentales de última generación diferentes de las usualmente utilizadas: las rodopsinas. Las identidades por debajo del 30 % y la cobertura que no supera el 85%, la gran mayoría se encuentran entre el 75% y 68%, sobre la secuencia 5HT<sub>2A</sub> son muy pequeñas para hacer uso de una sola plantilla como estructura base, por tal razón se contemplan múltiples plantillas basadas en el parámetro Z-score, definiendo así la energía de separación entre el plegamiento nativo de la estructura experimental con respecto al conjunto de unidades calculadas del plegamiento de la secuencia 5HT<sub>2A</sub> y usando

Tabla 1. Residuos más conservados localizados en la familia de las rodopsinas.

Resultados modelo 5HT <sub>2A</sub>			
Dominio hélice transmembranal (TM)	Tipo de residuo	Número de residuo en rodopsina (1)	Número de residuo en 5HT <sub>2A</sub>
TM1	N	55	92
TM2	D	83	120
<b>TM3</b>	(D/E)RY	134, <b>135</b> , 136	172, <b>173</b> , 174
TM4	W	161	200
TM5	P	215	246
<b>TM6</b>	CWXP	264, 265, 266, <b>267</b>	335, 336, 337 <b>338</b>
TM7	NPXXY	304, 303, 305, 306, 307	376, 377, 378, 379, 380
PUENTE S-S	C	110-187	148-227

-Residuos en **negrilla** números y letras, siempre se tienen que conservar.

Tabla 2. Plantillas alineadas de secuencias con mayor porcentaje de similitud con respecto a secuencia 5HT<sub>2A</sub>

Rango a	Secuencia identificada PDB	iden1 b	iden2 c	Cov d	Norm. Z-score e
1	2ks9A Receptor de Sustancia P	18%	19%	76%	2,29
2	3pb1A Plasminógeno Activador inhibidor 1	32%	25%	67%	5,39
3	2y00A Receptor β1 Adrenérgico	35%	23%	59%	4,98
4	3em1A Nucleósido difosfato kinasa	22%	22%	69%	3,67
5	2ks9A Receptor de sustancia P	17%	19%	76%	3,58
6	2rh1A Receptor β2 Adrenérgico	28%	24%	68%	4,53
7	3d4sA Receptor β2 Adrenérgico	3%	24%	68%	11,29
8	2rh1A Receptor β2 Adrenérgico	19%	24%	85%	3,83
9	3ny8A Receptor β2 Adrenérgico	28%	24%	68%	4,98
10	2vt4B Receptor β1 Adrenérgico	37%	23%	57%	3,95

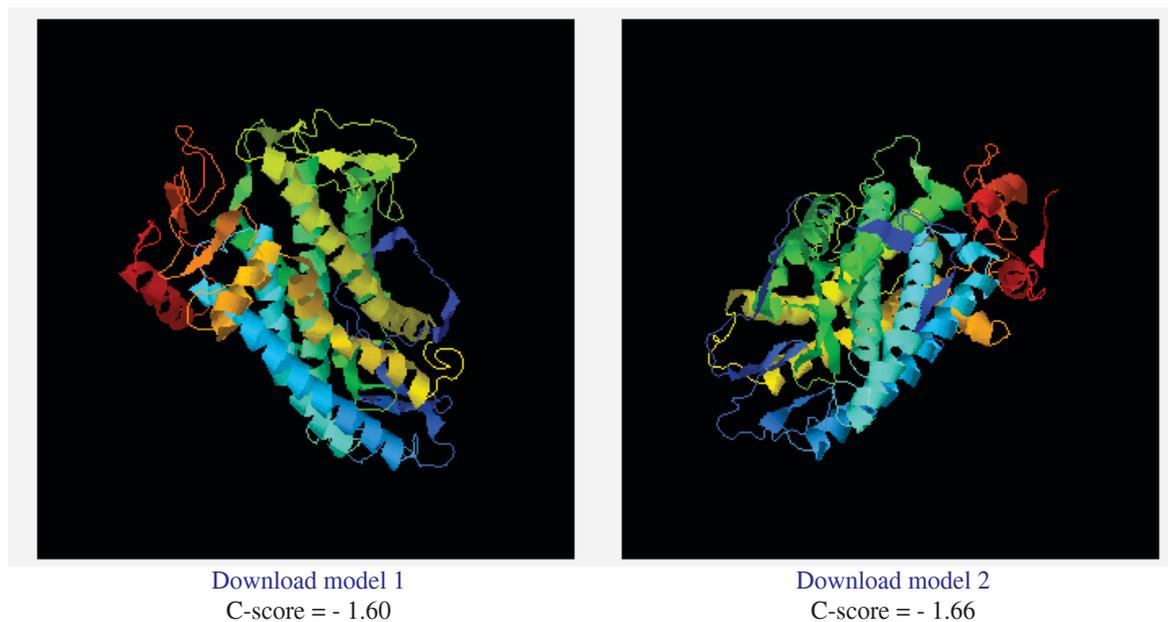
a) El rango de las plantillas, representan las mejores diez plantillas utilizadas por I-TASSER para realizar múltiples alineamientos. b) Iden1: Es el porcentaje de identidad de la secuencia consultada con respecto a las plantillas en regiones específicas. c) Iden2: Es el porcentaje de identidad de la secuencia consultada con respecto a toda la cadena de las secuencias de aminoácidos alineados. d) Cov: representa la cobertura de los alineamientos propuestos y es igual al número de residuos alineados dividido por la longitud de la proteína consultada. e) norm z-score: Es el estándar del z-score de los aminoácidos introducidos. Alineamientos con un z-score estandarizado > 1 significa un buen alineamiento y viceversa.

como algoritmo TM-Align contemplando alineamientos globales y locales para el ensamble de la estructura base donde extraen de cada proteína experimental fragmentos altamente relacionados con la secuencia adaptando los plegamientos con mayor similitud estructural (**Tabla 2**), (**Figura supl. A**).

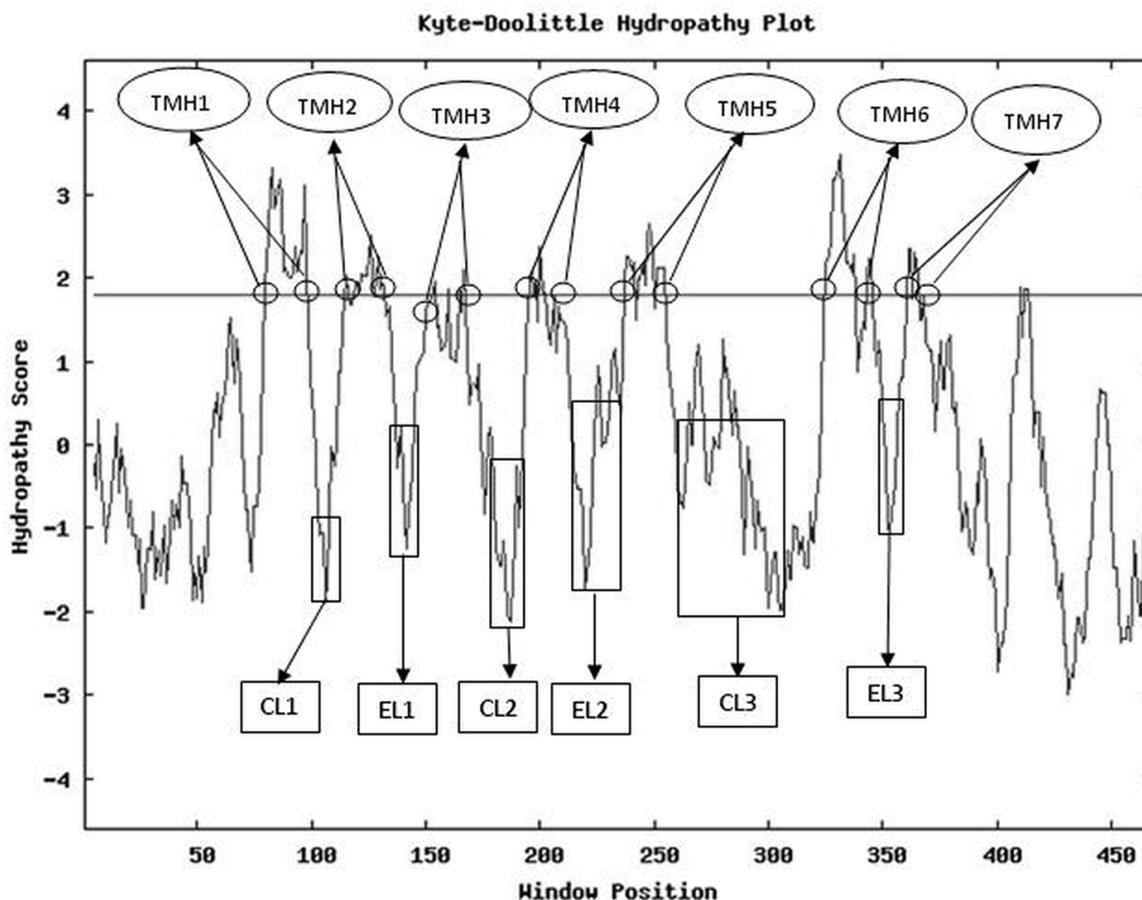
Una vez obtenidos los modelos se observa que el primero de ellos con C-score = -1.60, presenta una desviación estándar de  $10,9 \pm 4,6 \text{ \AA}$ , debido a las características estructurales que conservan sus aminoácidos, los cuales originan orientaciones diferentes en los segmentos transmembranales modificando la posición de los bucles, que es la región donde es más notoria la desviación estándar con respecto a las estructuras experimentales (**Figura 1**). También, se debe tener en cuenta que la estructura 5HT<sub>2A</sub> no tiene relación estructural cercana con las plantillas utilizadas a lo largo de toda la secuencia, éstas conservan el sitio de enlace ortoestérico que en su estado inactivo se encuentra enlazado formando el puente disulfuro y que su ruptura conlleva a cambios conformacionales activando el receptor desde los sitios alostéricos generando cascadas de señalización como segundos mensajeros, pero comparadas con rodopsina, el sitio de enlace es más grande y abierto al espacio extracelular. Por ésto se realiza una

comparación con el mapa de densidad electrónica de la familia de rodopsina de bovino, se hace una comparación estructural con proteínas relacionadas disminuyendo el valor de RMSD (Raíz Media Cuadrática de la Desviación Estándar) a 1.2 Å en promedio, acercándose el modelo a estructuras experimentales sin sobreponer sus plantillas como homólogos. (**Figura supl. B**)

Se plantea el análisis del perfil de hidrofobicidad, representado en el gráfico de Kyte & Doolittle, éste contempla las características fisicoquímicas de los aminoácidos que conforman la proteína dándole función o estructura y establecen los límites para la generación de estructura (hélices a y bucles) (**Tabla 3**), donde regiones por encima de cero son hidrofóbicas y regiones por debajo de cero son hidrofílicas, se observa la probabilidad propuesta por el perfil de hidrofobicidad que los aminoácidos de la secuencia poseen características fisicoquímicas que determinan la conformación estructural, confirmando que la secuencia estudiada (5HT<sub>2A</sub>) contiene propiedades de proteína integral con 7 hélices identificadas por su carácter hidrofóbico, y los respectivos bucles citoplasmáticos y extracelulares por sus características hidrofílicas (**Figura 2**).



**Figura 1.** Modelo 1 y 2 predicho por I-TASSER el cual se mide la calidad con base al sistema de puntos representado por el C-Score, resultado que estima la calidad del modelo predicho, éste es calculado basado en la relevancia de los entrelazamientos de los alineamientos de las plantillas y el parámetro de la simulación en la convergencia de la estructura ensamblada. C-score se encuentra en el rango de [-5,2], donde un C-score de mayor valor significa un buen modelo y viceversa (15,16). Exactitud estimada del Modelo 1: ( $0,52 \pm 0,15$  TM-score) (RMSD  $10,9 \pm 4,6 \text{ \AA}$ ).



**Figura 2.** Representación del perfil de hidrofobicidad Kyte & Doolittle 5HT<sub>2A</sub> humano CL – bucle citoplasmático, EL – bucle extracelular, TMH – hélice transmembranal (<http://web.expasy.org/protscale/> [calculado: 09/03/2011] Modificado)

Por otro lado, del modelo obtenido con C-Score= -1.60, se obtuvo los límites transmembranales propuestos para cada hélice y para los respectivos bucles dentro de la conformación se encontraron 3 bucles citoplasmáticos y 3 bucles extracelulares enlazados a las hélices de transmembrana; uno de los bucles (BE 2), compuesto por 24 aminoácidos, el cual representa un reto en el modelaje, puesto que se enfrenta en modelar adecuadamente el bucle evitando impedimentos estéricos y en mantener el enlace puente disulfuro, uno de los más importantes en la conformación de la proteína, pues, es en éste donde se presenta el enlace propuesto para la formación de cistina (enlace ortoestérico) en los aminoácidos CYS 148 de la HTM 3 y la CYS 227 del bucle BE2ab que inicia en la HTM 4 y termina en la HTM 5; la importancia del enlace ortoestérico radica en que es un sitio activo estructural que enlazado se encuentra en su estado inactivo pero al estimular el receptor en regiones alostéricas sufre ruptura del puente disulfuro activando el receptor, modificando la

orientación de las hélices transmembranales y generando alguna cascada de señalización para segundos mensajeros o apertura de canales iónicos. En otras investigaciones han contemplado el modelaje directo de los aminoácidos que conforman el puente desde la HTM 3 y el bucle BE2a Y BE2b construyendo un nuevo residuo CYS-S-S-CYS ensamblándolos con restricciones en las HTM para asegurar la conformación (22).

En este trabajo se planteó la fusión de dos modelos generados por I-TASSER (modelo 1 C-score = -1,60 y modelo 2 = -1,66, ya que se encontraba en cada uno de ellos estructuras que mantenían el puente disulfuro formado entre las CYS 148 y CYS 227 a una distancia de 2.05 Å; uno de los hallazgos fue la identificación del puente sujetando la CYS227 sobre una hoja β en el modelo 2 ya que en el modelo 1 las hojas β se encontraban, pero eran más pequeñas y no mantenían la CYS sobre la hoja β. Es así que se observó una mayor estabilidad sobre la hoja β del modelo 2 que evita interacciones con

los otros bucles debido a los arreglos conformacionales y la longitud de dicha hoja  $\beta$ , impidiendo así la ruptura del puente disulfuro que se puede generar en la minimización de energía. En simulaciones previas se encontraron estos inconvenientes pero fueron solucionados por la hoja  $\beta$ , (**Tabla 3, Figura 3, Figura supl C**).

### Refinamiento estructural

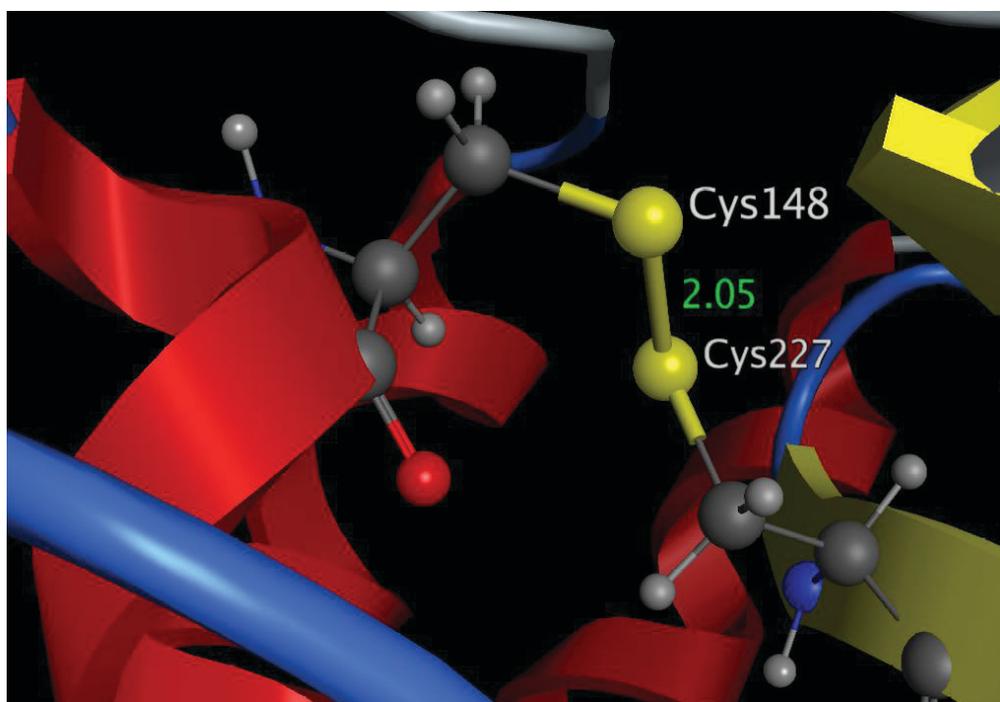
En este proceso encontraron todos aquellos aminoácidos que generan distorsiones e impedimentos estéricos, se identifican por medio de gráficos de Ramachandran que miden la posición de los aminoácidos en ángulos phi y psi, prediciendo las regiones favorables en la conformación geométrica. El modelo 1 obtenido por el servidor, fue verificado por gráficos de Ramachandran (modelo con C-Score = -1,60), obteniéndose 83,4% de aminoácidos en regiones favorables de acuerdo a parámetros del gráfico de este modelo preliminar. Se identifica que la gran mayoría de aminoácidos presentes en estas regiones no favorables corresponden a los bucles, los cuales, por falta de forma definida y sus características fisicoquímicas como el carácter hidrofílico, determinan su flexibilidad en medios polares (**Figura Supl D**).

Uno de los problemas que se presentó en el modelo 1 al realizarle arreglos conformacionales de los bucles fue

la pérdida de estabilidad en el sitio ortoestérico (puente disulfuro), sufriendo ruptura. De su estabilidad depende su forma para ser sometido a estudio de acoplamiento a receptor ya que se requieren los sitios alostéricos identificados sin impedimentos estéricos que generen interferencia frente al ligando por errores de conformación. La fusión de los modelos 1 y 2 obtenidos por I-TASSER, se contempló debido a la presencia de estructuras definidas tales como hélices  $\alpha$  (en mayor proporción), y que una de las hojas  $\beta$  se encontraba la CYS 227, la cual enlaza con la CYS 148, manteniendo el puente disulfuro. Esta fusión fue evaluada mediante gráficos de Ramachandran, obteniéndose 85,7% de los aminoácidos en regiones favorables, evidenciando mejora conformacional con respecto al modelo 1 o modelo preliminar (**Figura Supl E**). El refinamiento de los bucles se llevó mediante software MODELLER ModLoop, por método de restricciones espaciales en un número máximo de 12 aminoácidos por bucle, corrigiendo todos los impedimentos estéricos y errores de conformación.

### Minimización de energía

Una vez ubicados todos los aminoácidos dentro de regiones permitidas del gráfico Ramachandran, se realiza minimización de energía utilizando el campo de fuerza OPLS\_2005, el cual está definido por parametrizaciones



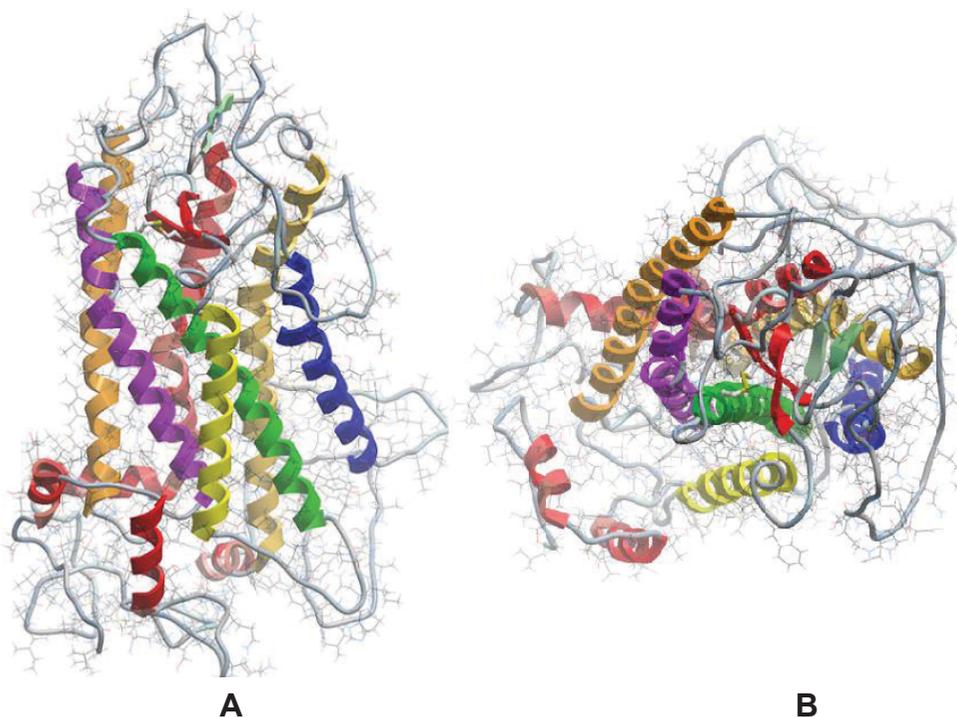
**Figura 3.** Conservación puente disulfuro en 2,05 Å, después de minimizada energía modelo final. Modelo final obtenido 5HT<sub>2A</sub> de Homo sapiens

de interacciones no enlazantes por simulaciones de Monte Carlo, debido a que en el proceso de refinamiento se producen movimientos en los bucles y hélices transmembranales generando cambios conformacionales representados en energía libre alrededor del modelo. En las minimizaciones de energía no es recomendable hacer uso excesivo del campo de fuerza ya que éste conlleva a las estructuras a un reordenamiento y posterior empaquetamiento, esto se ve representando en las posiciones tomadas por los aminoácidos en el gráfico de Ramachandran, encontrándose por fuera de las regiones favorables. La elección del campo de fuerza se atribuye a los arreglos geométricos leves de forma que no inducen al modelo construido al empaquetamiento, ya que el uso de otros campos de fuerza tales como AMBER y CHARMM modifican significativamente la estructura del modelo obtenido rompiendo el puente disulfuro generando así cambios conformacionales sobre el gráfico de Ramachandran, induciendo al modelo a un empaquetamiento; en los estudios experimentales estos rompimientos no se presentan (excepto cuando se ligan los agonistas), es decir, que si en los modelos construidos no existe el puente disulfuro en el estado inactivo de la proteína se impide hacer uso del modelo propuesto en ensayos de acoplamiento (docking)

Durante los diversos procesos de modelaje tales como, la generación de estructuras por el servidor I-TASSER, la fusión de los modelos 1 – 2, el refinamiento y minimización de energía de la estructura, todos fueron sometidos a evaluación de geometría conformacional por medio de gráficos de Ramachandran como indicador mediante la herramienta MOLPROBITY.

Del modelo final minimizado se obtuvo en la validación 91,7 % de los aminoácidos ubicados en regiones favorables de su geometría conformacional (**Figura supl F**). EL RMSD del modelo final minimizado, comparado con estructuras cristalinas relacionadas con el receptor 5HT<sub>2A</sub> fue en promedio 0,95 Å. Otro aspecto relevante en la validación, durante toda la construcción del modelo se enfatizó en la conservación del sitio ortoesteroico (puente disulfuro) en una distancia inferior a 2,8 Å que es la distancia requerida para un enlace S-S, ya que éste brinda características estructurales y funcionales a esta clase de proteínas, después de minimizada la energía se verificó la conservación del puente disulfuro de las CYS 148- CYS 227 (**Figura 3**). Estos resultados demuestran que la estructura terciaria del modelo molecular teórico del receptor 5HT<sub>2A</sub> de *Homo Sapiens* fue construido y mantiene las características propias de este tipo de receptores acoplados a proteína G (**Figura 4**).

## Validación del modelo



**Figura 4.** Modelo final del Receptor 5HT<sub>2A</sub> de Humano, con 7 hélices transmembranales, protonado y minimizado con campo de fuerza OPLS\_2005, (A: visualización lateral, B: visualización Posterior)

## Conclusiones

En el presente trabajo se ha seguido el protocolo planteado por Ballesteros y Weinstein en el año de 1995 establecido para la construcción de modelos teóricos de receptores acoplados a proteína G, enfocados en: aminoácidos específicos que son conservados en cada una de las hélices que atraviesan la membrana; perfiles de hidrofobicidad que demuestran las características químicas de los aminoácidos dando posibilidades de los límites transmembranales; asegurar la correcta conformación de la estructura por medio de la evaluación geométrica (Gráfico Ramachandran). Cumplir con los parámetros planteados, el uso de algoritmos, estructuras  $\beta$  adrenérgicas obtenidas experimentalmente y el uso de un campo de fuerza más preciso y adecuado para este tipo de proteínas, permitió construir el modelo del receptor 5HT<sub>2A</sub> de *Homo sapiens*, demostrando la vigencia de dicho protocolo utilizándolo con planteamientos teóricos de última generación aplicados a herramientas computacionales que permiten generar un modelo teórico 3D propuesto para el estudio de acoplamiento de ligando a receptor.

## Agradecimientos

Los autores agradecen al Departamento de Física de la Facultad de Ciencias de la Pontificia Universidad Javeriana y al Grupo de Biofísica y Bioquímica Estructural por haber permitido el uso de los equipos y el tiempo dedicado al desarrollo de este trabajo.

## Financiación

Este trabajo fue financiado por los autores, como parte del proyecto de grado para la obtención del título de químico farmacéutico de la Universidad de ciencias aplicadas y ambientales (U.D.C.A.).

## Conflicto de intereses

Los autores manifiestan no tener conflicto de intereses.

## Referencias

1. Florez J, Armijo J, Mediavilla A. *Farmacología Humana*, Quinta Edición, Elsevier – Masson 2008; 7-12 p.
2. Arboleda D. *Jerarquía estructural de las proteínas*. Editorial club universitario 2011; 164 – 171p
3. Research Collaboratory for Structural Bioinformatics (RCSB), *PROTEIN DATA BANK* [http://www.pdb.org/pdb/static.do?p=general\\_information/about\\_pdb/index.html](http://www.pdb.org/pdb/static.do?p=general_information/about_pdb/index.html) Consultado el 01 de febrero de 2012.
4. Nayeem A, Stikoff D, Krystek S. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science* 2006; **15**:808-824
5. McRobb FM, Capuano B, Crosby IT, Chalmers DK, Yuriev E. Homology modeling and docking evaluation of aminergic G protein-coupled receptors. *Journal of Chemical Information and Modeling* 2010; **50** (4): 626-637.
6. Ballesteros J A, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled Receptors. *Methods in Neurosciences* 1995; **25** 366-428
7. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. *Oxford Journals Bioinformatics* 1997; **13** (4): 425-430.
8. Leach A. *Molecular Modelling: principles and applications*, second edition, 2001 Pearson Education UK. Cap: 10 509-562 p
9. Roth B, Allen J, Yadav P. insights into the regulation of 5-HT<sub>2A</sub> receptors by scaffolding proteins and kinases. *Neuropharmacology* 2008; **55** (6): 961–968
10. Kasper S, Lerman M, McQuade R, Saha A, Carson W, Ali M, Archibald D, Ingenito G, Marcus R, Pigott T. Efficacy and safety of aripiprazole vs. haloperidol for long-term maintenance treatment following acute relapse of schizophrenia. *The International Journal Neuropsychopharmacology* 2003;**6**(4):325–337
11. Dos Santos A, Rosso A, Borges M. The use of an antagonist 5-HT<sub>2A/C</sub> for depression and motor function in Parkinson's disease. *Arquivos de Neuro-psiquiatria* 2009; **67**(2-B):407-412.
12. Reggio PH. Computational methods in drug design: modeling g protein-coupled receptor monomers, dimers, and oligomers. *The AAPS Journal* 2006; **8** (2): E322-E336
13. Huang E. Construction of a sequence motif characteristic of aminergic G protein-coupled receptors. *Protein Science* 2003; **12**:1360-1367
14. Weinstein H. Hallucinogen actions on 5HT receptors reveal distinct mechanism of activation and signalling by g protein-coupled receptors. *The AAPS Journal* 2006; **7** (4): E871-E884.
15. Ambrish Roy, Alper K, Yang Z. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 2010; **5**: 725-738
16. Yang Z. I-TASSER server for protein 3D structure prediction. *Biomed Central Bioinformatics* 2008; **9**: 40 1-8.
17. Sitao W, Yang Z. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* 2007; **35** (10): 3375–3382

18. Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2010 <http://www.schrodinger.com/citations/> Consultado el 02 de Febrero de 2012.
19. Fiser A, Sali A. ModLoop: automated modeling of bucles in protein structures. *Bioinformatics* 2003; **19** (18): 2500-2501.
20. Storz C, Johnson G, French A, Csonka G. Comparison of different force fields for the study of disaccharides. *Carbohydrate Research* 2009; **344**: 2217-2228
21. Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica section D: Biological Crystallography* 2010; **66**: 12-21
22. Mehler E, Periole S, Weinstein H. Key issues in the computational simulation of GPCR function: representation of loop domains. *Journal of Computer-Aided Molecular Design* 2002 **16**: 841-853

Rank	PDB Hit	20	40	60
		CCS	CCCC	HHHCCCCCCCCSSSSCCCCCCCCCCCCCCCCCCCCCCCCCCCC
		MDILCEENTSL	SSTTNSLMQLNDD	TRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGC
1	<a href="#">2ks9A</a>	-----	-----	DNVLPVDSDLSPNISTNTSEPN
2	<a href="#">3pb1A</a>	-----	-----	-----
3	<a href="#">2y00A</a>	-----	-----	-----
4	<a href="#">3em1A</a>	IM-----	-----	-----
5	<a href="#">2ks9A</a>	-----	-----	DNVLPVDSDLSPNISTM
6	<a href="#">2rh1A</a>	-----	-----	-----
7	<a href="#">3d4sA</a>	-----	-----	-----
8	<a href="#">2rh1A</a>	-----	-----	-----
9	<a href="#">3ny8A</a>	-----	-----	-----
10	<a href="#">2vt4B</a>	-----	-----	-----

Rank	PDB Hit	80	100	120
		CCCCCCCC	HHHHHHHHHHHHHHHHHHHHHHHHSSSSSSSS	CCCCC
		LSPSCLSL	LHLQEKNSALLTAVV	ILTIAGNIVIMAVSLEKKLQ
		NATNYFLMSL	AIAD	
1	<a href="#">2ks9A</a>	Q-----	FVQPAWQIVLWAAAYTV	IVVTSVVG
2	<a href="#">3pb1A</a>	-----	YALSYCALILAIV	FGNGLVCM
3	<a href="#">2y00A</a>	-----	AELLSQWEAGMSLL	MALVVLLIVAG
4	<a href="#">3em1A</a>	-----	GSSVYITVELAIAV	LAILGNVLCW
5	<a href="#">2ks9A</a>	TSEPNQFVQPAWQIVLWAAAYTV	IVVTSVVG	NVVVMWII
6	<a href="#">2rh1A</a>	-----	DEVWVVG	MGIIVMSLIVL
7	<a href="#">3d4sA</a>	-----	WVVG	MGIIVMSLIVL
8	<a href="#">2rh1A</a>	-----	DEVWVVG	MGIIVMSLIVL
9	<a href="#">3ny8A</a>	-----	WVVG	MGIIVMSLIVL
10	<a href="#">2vt4B</a>	-----	WEAGMSLL	MALVVLLIVAG





Rank	TM-score	RMSD <sup>a</sup>	IDEN <sup>a</sup>	Cov.	PDB Hit	Structural alignment using TM-align
1	0.7153	0.64	0.19	0.72	<a href="#">2peda</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
2	0.7147	0.68	0.19	0.72	<a href="#">2g87A</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
3	0.7146	0.68	0.19	0.72	<a href="#">1u19A</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
4	0.7139	0.72	0.19	0.72	<a href="#">2hpyA</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
5	0.7052	1.85	0.17	0.74	<a href="#">2ks9A</a> Model1	-----DNVL-----FVDSLSPNISINTSEPN-QFVQPANQIVLWAAAYTVIVVTS MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
6	0.6939	0.91	0.19	0.70	<a href="#">119hA</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
7	0.6848	1.29	0.20	0.70	<a href="#">1f88A</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
8	0.6603	1.51	0.20	0.68	<a href="#">1gzmA</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
9	0.6603	1.51	0.20	0.68	<a href="#">3c91A</a> Model1	-----MNGTEGPNFYVVPFSNKTGVVRSPPFEAPQ--YYLAEPVQFSLAAYMFLLIIMLG MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSALLTAVVVIILT
10	0.6278	3.52	0.19	0.70	<a href="#">2ziyA</a> Model1	DLRDNETW-----WYNPSIIVHPHURE-----F-----DQVPDAVYYSLGI -----MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCSLLLHLOEKNWSAL

PDB Hit	HTM1	HTM2	HTM3	HTM4
<a href="#">2peda</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">2g87A</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">1u19A</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">2hpyA</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">2ks9A</a> Model1	VVGNVVMWII LANHKRRTVTNYFLVNLAFASMAAFNTVNFVYAVHNE-WYGLFYCKFHNFPIAAVFAIYSNTAVAFDRYIAIIP-P-LQPRLSATATKVVICVIVWLAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">119hA</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">1f88A</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">1gzmA</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">3c91A</a> Model1	FFINFLTLTYVTVQHKLRTPLNLYILLNLAVADLFMVFGGFTTTLTSLHGY-FVFGPTGCNLEGGFATLGGEIALWSLVVLAIERYVVKPKPM-SNFRFGENHAIHGVAFTVWMAL IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			
<a href="#">2ziyA</a> Model1	FIGIGGIICGGNGIVLYLFTKTKSLQTPAIFFIINLAFSDFTFSLVNGPPLHTISCFKKWIFG-FAACKVYGFIGGIFGHSIHTHAMISIDRYNWIGRPMASAKKMSHRRAFI LAVVVIILT IAGNVLVIMAVSLEKKLQONATNYFLMSLAIAIDMLLGLVMPVSMILTILYGYRWPLPSKLCVWYIYLDVLFSTASIMHLCATSLDRYVAIQNP IHSRFRNSRTKAFKIIAVWTISV			

PDB Hit	HTM5
<a href="#">2pedA</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAAQQQES----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">2g87A</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAAQQQES----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">1u19A</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAAQQQES----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">2hpyA</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAAQQQES----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">2ks9A</a> Model1	LLAFPGQ-YYSTT-ETMPSRVVCMIEWPEHPNKIYKQVYHICVTVLIYFLIPLLVIGYAYTVVG-----ITLWASEIPGDSDDR----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">119hA</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAA----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">1f88A</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKEAAA-S----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">1gzmA</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKE-----AAAQ----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">3c91A</a> Model1	ACAAPPLVG-WSRYIPEGHCQSCGIDYYTPHEETNNEFVYIMFVVHFIIPLIVIFFCYGQLV-----FTVKE-----AAAQ----- GISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIHREPG
<a href="#">2ziyA</a> Model1	MIIFVVLWSVLAIGPI-FGWGAYTLEGVLCNCSFDYISRDSTTRSNILCIPLILGFFGPIILIIFFCYFNIV-----MSVSN-HE----KEMAAAKRLNAK--- KIIAVWTISVSGISHPPIPVFGLQDDSKVFKEGSCLLADDN-----F--VLIGSFVSPFIIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQ-----SSL

PDB Hit	HTM 6	HTM 7
<a href="#">2pedA</a> Model1	-----ATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGNPLGDDEASTT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">2g87A</a> Model1	-----ATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGNPLGDDEASTT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">1u19A</a> Model1	-----ATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGNPLGDDEASTT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">2hpyA</a> Model1	-----ATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGNPLGDDEASTT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">2ks9A</a> Model1	-----YHEQVSAKRKRVKMMIVVCTFAICWLPFHIFLLPYINPDLYLKKFIQOVLAINWULAMSSTYNPPIYCCLNDRFRLGFKHAFRCPPFISAGDYEGLE SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLI-	
<a href="#">119hA</a> Model1	-----ATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGNPLGD--TT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">1f88A</a> Model1	-----A-TTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLCGKGN--P---S--TT SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">1gzmA</a> Model1	QQ-E--SATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLC--GKN----- SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">3c91A</a> Model1	QQ-E--SATTQKAEKEVTRNVIIMVIAFLICWLPYAGVAFYIFTHQGS---DFGPIFMTIPAFFAKTSAVYNPVIYIMHNKQFRNCMVTTLC--GKN----- SYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCYKENKKPLQLIL	
<a href="#">2ziyA</a> Model1	-----ELRKAQAGANAEMRLAKISIVIVSQFLLSWSPYAVVALLAQFG--PL-EWVTPYAAQLPVMFAKASAIHNPHIYSVSHPKFREAISQTF SSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYI	

PDB Hit	
<a href="#">2pedA</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">2g87A</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">1u19A</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">2hpyA</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">2ks9A</a>	MKSTRYLQTQGSVYKVSRL-----LETTISTVVG-A-----
Model1	L----VNTIPALA-YKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">119hA</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">1f88A</a>	VSKTETSQVAP-----A-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">1gzmA</a>	DD-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">3c91A</a>	DD-----
Model1	VNTIPALAYKSSQLQMGQKKNKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
<a href="#">2ziyA</a>	PVVLTC--QFDKETEDDK-----DAETEIPAGESSDAAPSADAA-QMKE-----
Model1	QC----QYKENKK-PLQLILVNTIPALAYKSSQLQMGQKKN-----SKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV

**Figura suplementaria B.** Comparación del modelo Generado por I-TASSER. Con estructuras cristalinas obtenidas experimentalmente, relacionadas con la secuencia.

- (a) pares de residuos que son estructuralmente alineado por el programa de TM-align se resaltan en color. El régimen de colores se basa en la propiedad de los aminoácidos, donde polares son de colores brillantes, mientras que los residuos no polares son de color oscuro.
- b) El ordenamiento de las proteínas se basa en el TM-score del ensamble estructural entre modelo 1 y las estructuras de la base de datos PDB.
- c) RMSD es la desviación estándar entre los residuos del modelo que son estructuralmente alineado por TM-align y las estructuras obtenidas experimentalmente.
- d) IDEN<sup>a</sup> es el porcentaje de identidad de la secuencia con estructuras obtenidas experimentalmente de la base de datos PDB.
- e) Cov. Representa la cobertura del alineamiento realizado por TM-align y es igual a la cantidad de residuos alineados estructuralmente dividida por la longitud del modelo.

```

+gi|10835175|MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSSGEANTSDAFNWTVDSE|
model2_A_ssa
model2_A MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSSGEANTSDAFNWTVDSE
modell1_A_ssa
modell1_A MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSSGEANTSDAFNWTVDSE
gi|10835175|ref|NP_000612.1|pred_ssa
qi|10835175|MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSSGEANTSDAFNWTVDSE

+gi|10835175|NRTNLSCEGCLSPSCLSLHHLQEKNSALLTAVVILTIAGNILVIMAVS|
model2_A_ssa
model2_A NRTNLSCEGCLSPSCLSLHHLQEKNSALLTAVVILTIAGNILVIMAVS
modell1_A_ssa
modell1_A NRTNLSCEGCLSPSCLSLHHLQEKNSALLTAVVILTIAGNILVIMAVS
gi|10835175|ref|NP_000612.1|pre
qi|10835175|NRTNLSCEGCLSPSCLSLHHLQEKNSALLTAVVILTIAGNILVIMAVS

+gi|10835175|LEKKLQATNYFLMSLAIADMLLGLVMPVSMILTILYGYRWPLPSKLCV|
model2_A_ssa
model2_A LEKKLQATNYFLMSLAIADMLLGLVMPVSMILTILYGYRWPLPSKLCV
modell1_A_ssa
modell1_A LEKKLQATNYFLMSLAIADMLLGLVMPVSMILTILYGYRWPLPSKLCV
gi|10835175|ref|NP_000612.1|pre
qi|10835175|LEKKLQATNYFLMSLAIADMLLGLVMPVSMILTILYGYRWPLPSKLCV

+gi|10835175|WIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAPLKI IAVW|
model2_A_ssa
model2_A WIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAPLKI IAVW
modell1_A_ssa
modell1_A WIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAPLKI IAVW
gi|10835175|
qi|10835175|WIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAPLKI IAVW

```

1

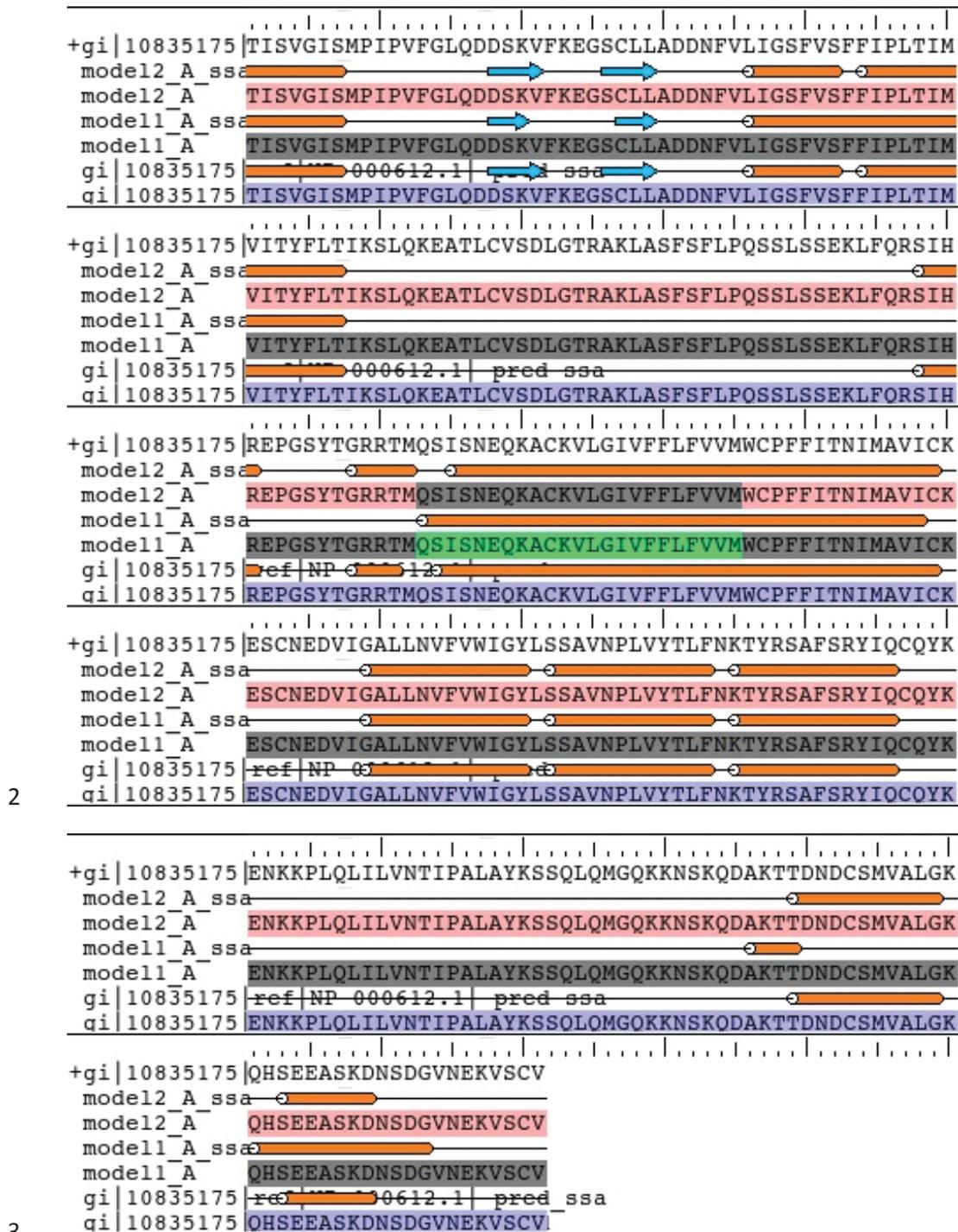


Figura suplementaria C. fusión modelos 1 y 2 regiones en verde fluorescente son las seleccionadas del modelo 1, regiones en rosado son las seleccionadas del modelo 2. Morado es la secuencia 5HT2A, con la imagen que finalmente asumió de los modelos 1 y 2.

MolProbity Ramachandran analysis

model1.pdb, model 1

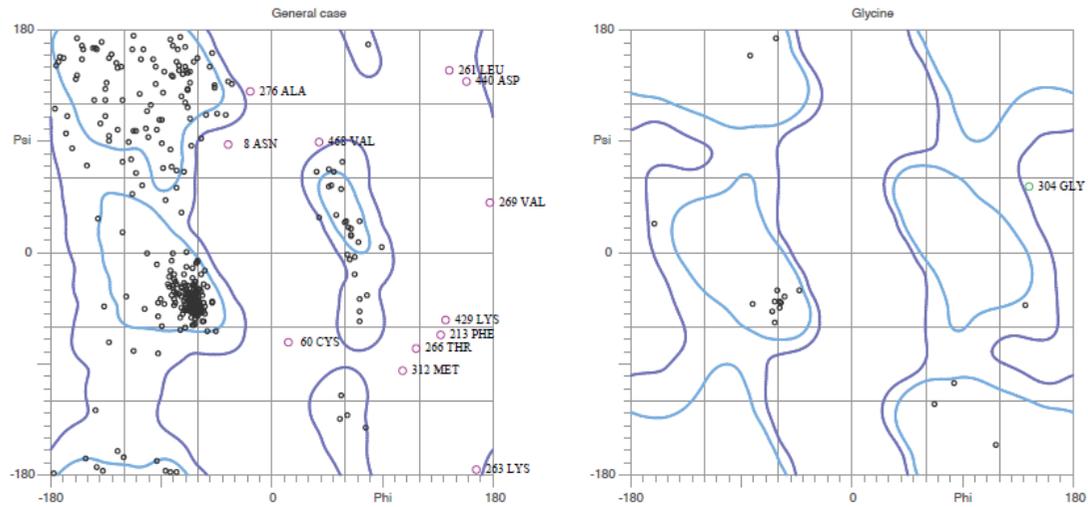
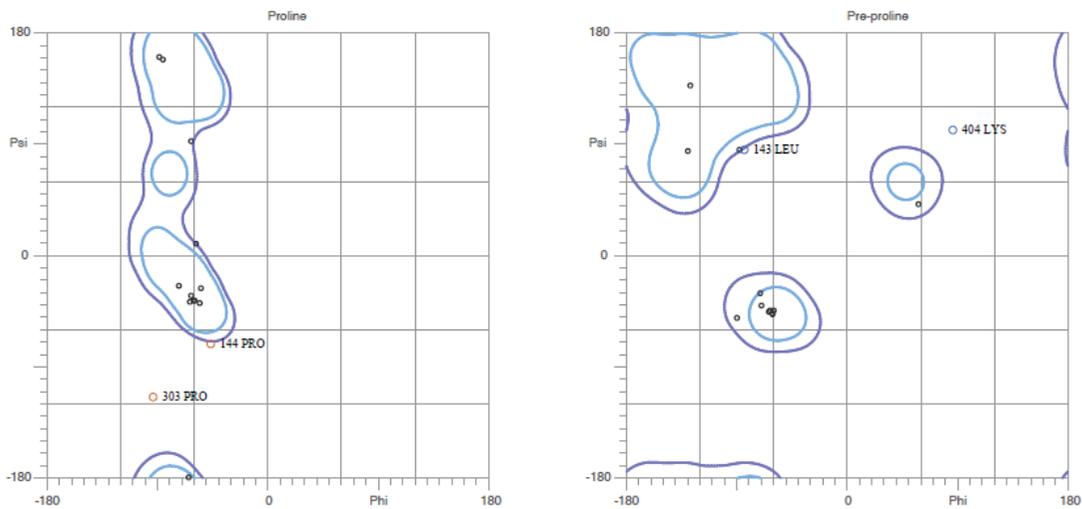


Gráfico de regiones favorecidas en el diagrama Ramachandran



83.4% (391/469) of all residues were in favored (98%) regions.  
96.4% (452/469) of all residues were in allowed (>99.8%) regions.

There were 17 outliers (phi, psi):

- 8 ASN (-36.2, 88.7)
- 60 CYS (13.2, -72.4)
- 143 LEU (-84.9, 86.2)
- 144 PRO (-47.6, -71.5)
- 213 PHE (137.8, -66.3)
- 261 LEU (144.8, 148.8)

- 263 LYS (166.9, -175.7)
- 266 THR (117.9, -77.9)
- 269 VAL (177.5, 41.7)
- 276 ALA (-18.5, 131.1)
- 303 PRO (-94.3, -114.8)
- 304 GLY (144.9, 54.6)
- 312 MET (106.6, -95.5)
- 404 LYS (87.0, 102.3)
- 429 LYS (141.6, -54.7)
- 440 ASP (158.4, 139.1)
- 468 VAL (38.6, 90.8)

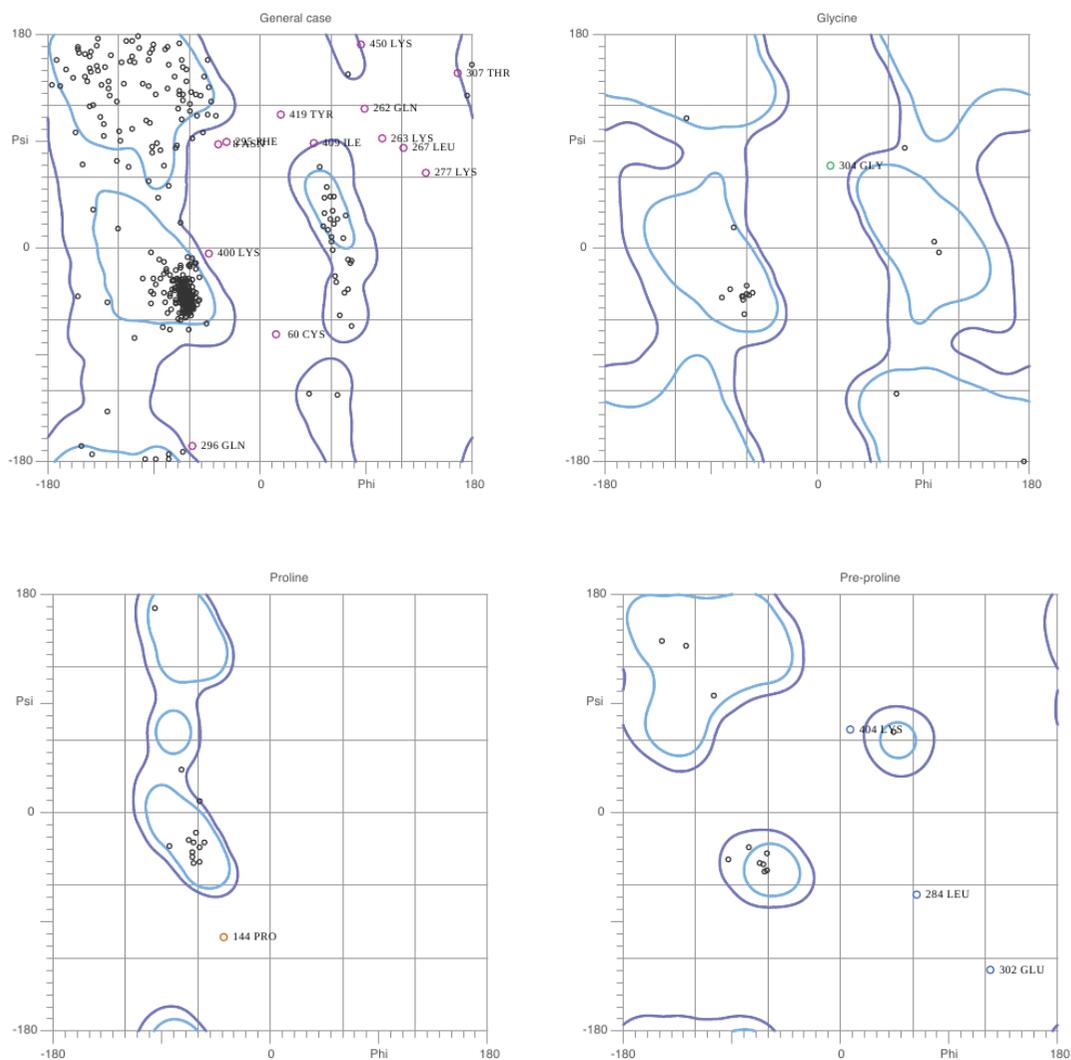
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figura suplementaria D. Gráfico Ramachandran modelo preliminar I-TASSER 83.4% de aminoácidos en regiones permitidas.

### MolProbity Ramachandran analysis

5ht2aoutput1FH.pdb, model 1



85.7% (400/467) of all residues were in favored (98%) regions.  
 96.1% (449/467) of all residues were in allowed (>99.8%) regions.

There were 18 outliers (phi, psi):  
 8 ASN (-36.2, 88.7)  
 60 CYS (13.2, -72.4)  
 144 PRO (-39.5, -102.9)  
 262 GLN (88.2, 118.8)  
 263 LYS (103.8, 93.7)  
 267 LEU (121.6, 85.8)  
 277 LYS (140.0, 64.9)

284 LEU (63.0, -67.9)  
 295 PHE (-29.2, 90.5)  
 296 GLN (-58.2, -166.5)  
 302 GLU (124.8, -129.9)  
 304 GLY (11.1, 70.2)  
 307 THR (167.7, 148.7)  
 400 LYS (-44.1, -4.4)  
 404 LYS (8.4, 69.8)  
 409 ILE (45.8, 89.3)  
 419 TYR (17.3, 113.1)  
 450 LYS (85.0, 172.8)

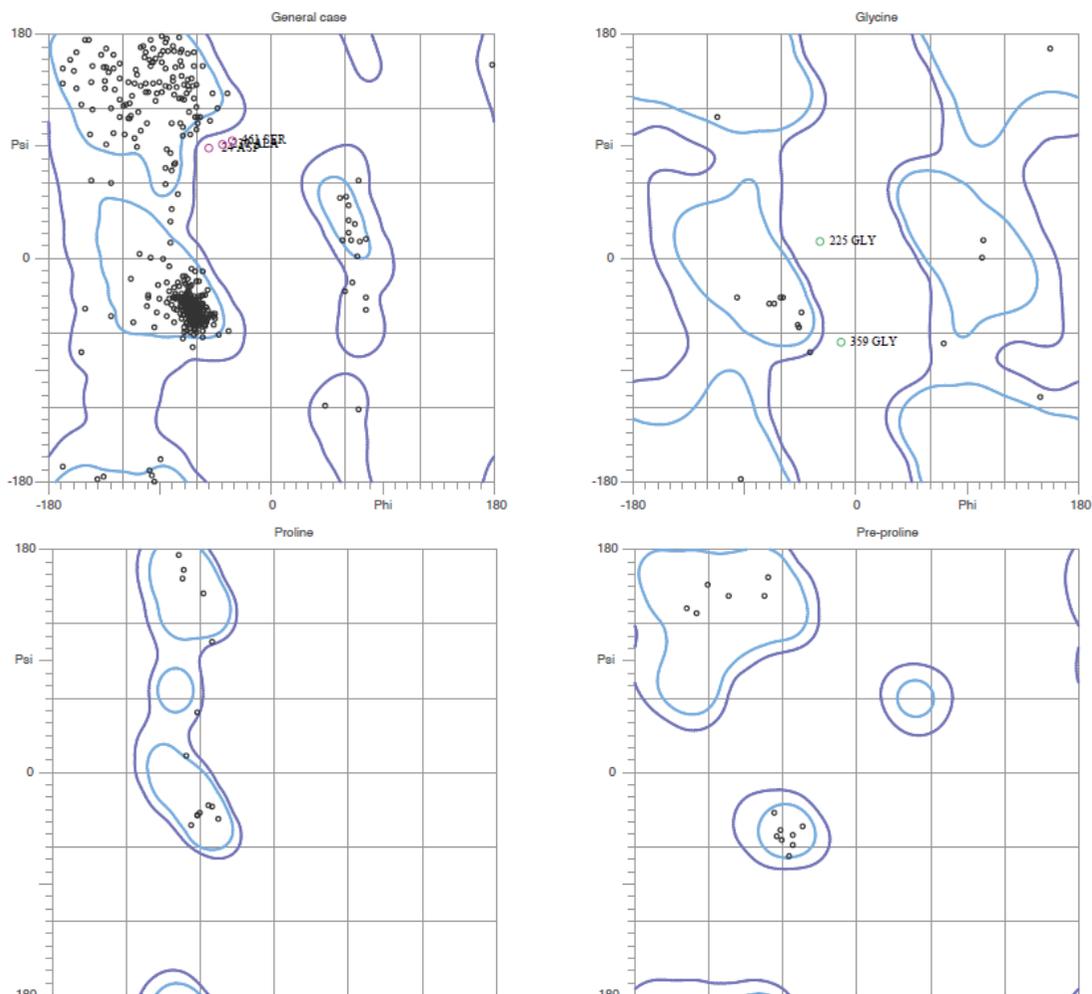
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

**Figura suplementaria E.** Gráfico Ramachandran modelo 1 Y 2 fusionado 85.7 % de aminoácidos en regiones permitidas.

**MolProbity Ramachandran analysis**

5ht2a\_15\_1FHprewiz\_5th2aoutput15\_1FH.1\_pdbv3.pdb, model 1



91.7% (430/469) of all residues were in favored (98%) regions.  
 98.9% (464/469) of all residues were in allowed (>99.8%) regions.

There were 5 outliers (phi, psi):

24 ASP (-51.3, 89.5)  
 225 GLY (-29.2, 14.5)  
 230 ALA (-40.5, 92.6)  
 359 GLY (-12.3, -67.2)  
 461 SER (-32.5, 95.1)

<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

**Figura suplementaria F.** Gráfico Ramachandran modelo 1 Y 2 fusionado luego de refinamiento y minimización de energía con campo de fuerza OPLS\_2005 91.7% de aminoácidos en regiones permitidas.