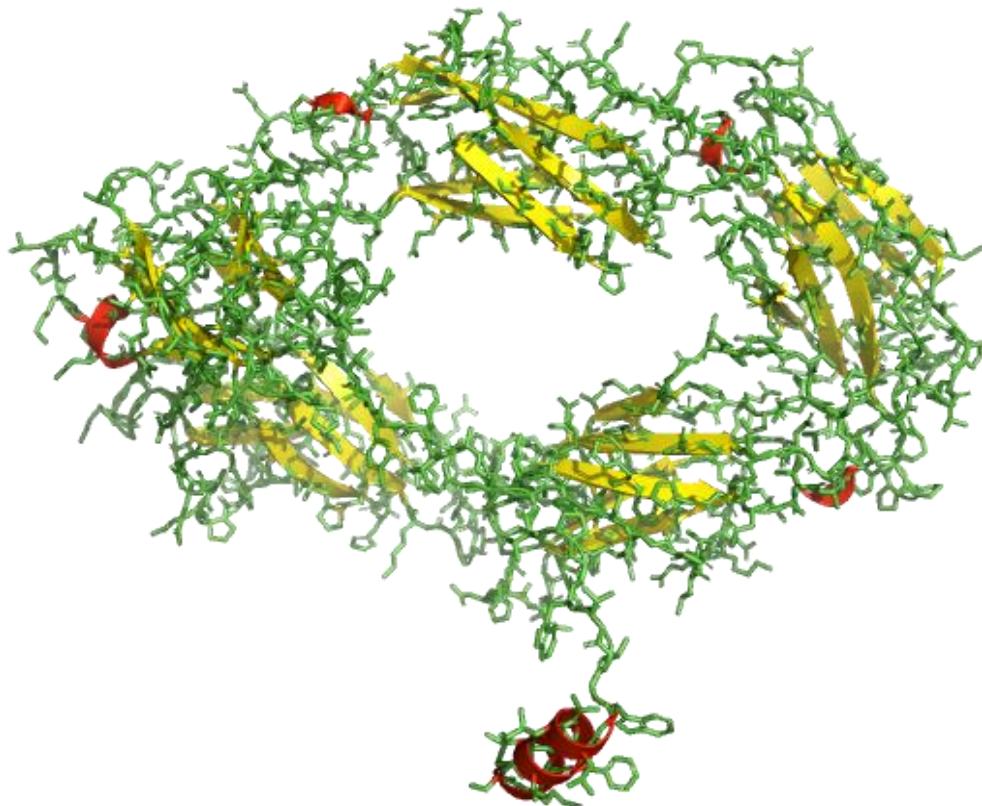


# Modelado comparativo de la estructura de la glicoproteína alfa 1B en *Homo sapiens*



Universitat  
Oberta  
de Catalunya



UNIVERSITAT DE  
BARCELONA

**Anna Pozuelo Montfort**

Máster Universitario en Bioinformática y Bioestadística  
Área de trabajo final: Diseño de fármacos y biología estructural

**Nombre Tutor/a de TF**

Emmanuel Fajardo

**Profesor/a responsable de la asignatura**

Nuria Pérez Álvarez

**Fecha Entrega**

15 de enero 2023



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](#)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Modelado comparativo de la estructura de la glicoproteína Alpha 1B en Homo sapiens</i>
<b>Nombre del autor:</b>	<i>Anna Pozuelo Montfort</i>
<b>Nombre del consultor/a:</b>	<i>Emmanuel Fajardo</i>
<b>Nombre del PRA:</b>	<i>Nuria Pérez Álvarez</i>
<b>Fecha de entrega (mm/aaaa):</b>	<i>01/2023</i>
<b>Titulación o programa:</b>	<i>Máster de Bioinformática y Bioestadística</i>
<b>Área del Trabajo Final:</b>	<i>Diseño de fármacos y biología estructural</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Modelado comparativo, MODELLER, A1BG</i>

## Resumen del Trabajo

El modelado comparativo nos permite generar predicciones sobre la estructura tridimensional de una proteína determinada a partir del alineamiento con una o más proteínas de estructura conocida. La predicción de la estructura de las proteínas permite contribuir al estudio de sus funciones, entre otras aplicaciones.

En este trabajo se describe como se predice la estructura tridimensional de la proteína A1BG utilizando el programa *MODELLER*. La proteína A1BG es una glicoproteína plasmática de función desconocida, aunque algunos estudios revelan su presencia en procesos tumorales <sup>(9)</sup> <sup>(10)</sup>. Se generan 3 modelos distintos: el primero con 9 moldes del PDB, el segundo con un molde de *AlphaFold* y el último, generado con el servidor *I-TASSER*.

La evaluación de la calidad de los modelos, mediante el sistema de validación *MolProbit*, muestra resultados favorables para el segundo modelo. El primer modelo no muestra suficiente calidad, la baja identidad de las secuencias afecta al proceso de modelado de la proteína. La comparación de los 3 modelos concluye que el modelo generado con la estructura de *AlphaFold* es de mejor calidad.

Un estudio describió la unión de la proteína A1BG y CRISP3 formando un complejo <sup>(7)</sup> en el plasma humano. Se ha comprobado, mediante el servidor *PyDockWEB*, el acoplamiento de la proteína A1BG con la proteína secretora CRISP-3, con resultados positivos para los dos modelos.

Los resultados concluyen que se podría tratar de una proteína asociada a la membrana. La fuerte unión del complejo entre las dos proteínas en plasma podría asociarse a una función de transporte.

## **Abstract**

Comparative modeling allows us to generate predictions about the three-dimensional structure of a given protein from the alignment with one or more proteins of known structure. The prediction of the structure of proteins allows us to contribute to the study of their functions, among other applications.

In this work we describe how the three-dimensional structure of the A1BG protein is predicted using the MODELLER program. The A1BG protein is a plasma glycoprotein of unknown function, although some studies reveal its presence in tumor processes<sup>(9)(10)</sup>. Three different models are generated: the first one with 9 PDB casts, the second one with an AlphaFold cast and the last one, generated with the I-TASSER server.

The evaluation of the quality of the models, using the MolProbity validation system, shows favorable results for the second model. The first model does not show sufficient quality, the low sequence identity affects the protein modeling process. Comparison of the 3 models concludes that the model generated with the AlphaFold structure is of better quality.

A study described the binding of A1BG protein and CRISP3 forming a complex<sup>(7)</sup> in human plasma. The docking of the A1BG protein with the CRISP-3 secretory protein has been tested using the PyDockWEB server, with positive results for both models.

The results conclude that it could be a membrane-associated protein. The strong binding of the complex between the two proteins in plasma could be associated with a transport function.

# Índice

<b>1. Introducción .....</b>	<b>1</b>
<b>1.1. Contexto y justificación del Trabajo.....</b>	<b>1</b>
<b>1.2. Objetivos del Trabajo.....</b>	<b>2</b>
<b>1.3. Impacto en sostenibilidad, ético-social y de diversidad.....</b>	<b>3</b>
<b>1.4. Enfoque y método seguido .....</b>	<b>3</b>
<b>1.5. Planificación del Trabajo .....</b>	<b>4</b>
<b>1.6. Breve sumario de productos obtenidos.....</b>	<b>7</b>
<b>1.7. Breve descripción de los otros capítulos de la memoria .....</b>	<b>7</b>
<b>2. Estado del arte .....</b>	<b>9</b>
<b>3. Materiales y métodos.....</b>	<b>10</b>
<b>4. Resultados.....</b>	<b>25</b>
<b>5. Conclusiones.....</b>	<b>35</b>
<b>6. Glosario.....</b>	<b>37</b>
<b>7. Bibliografía .....</b>	<b>38</b>
<b>8. Anexos .....</b>	<b>40</b>

# **Lista de figuras**

Figura 1. Archivo “target.fa” con la secuencia de aminoácidos de la proteína A1BG.....	10
Figura 2. Visualización esquemática de los resultados de la búsqueda de homólogos de la proteína A1BG con PSI-BLAST. ....	12
Figura 3. Alineamiento con mejor puntuación en los resultados de la búsqueda de estructuras de referencia con PSI-BLAST para la proteína A1BG. ....	13
Figura 4. Archivo “9templates.pir”. .....	19
Figura 5. Archivo “1template.pir”. .....	19
Figura 6. Archivo “9templates.py”.....	20
Figura 7. Visualización de la proteína CRISP-3 con PyMOL.....	23
Figura 8. Resultados del análisis de la estructura secundaria de la proteína A1BG con DISOPRED3. ....	25
Figura 9. Resultados del análisis de la estructura secundaria de la proteína A1BG con PSIPRED 4.0. ....	25
Figura 10. Resultados del análisis de los dominios de la proteína A1BG con InterPro. ....	26
Figura 11. Modelo 1 de la estructura de la proteína A1BG, obtenido a partir de 9 moldes del Protein Data Bank. ....	27
Figura 12. Modelo 2 de la estructura de la proteína A1BG, obtenido a partir de 1 molde de AlphaFold. ....	28
Figura 13. Modelo 3 de la estructura de la proteína A1BG, generado con I-TASSER.....	28
Figura 14. Visualización con PyMOL de la mejor configuración (Conf. 9415) del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	30
Figura 15. Visualización con PyMOL del enlace correspondiente a la primera zona de contacto entre el Modelo 1 y la proteína CRISP-3. Unión entre la fenilalanina (posición 7) de la proteína receptora y una treonina (posición 29) del ligando. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3. ....	31
Figura 16. Visualización con PyMOL del enlace correspondiente a la primera zona de contacto entre el Modelo 1 y la proteína CRISP-3. Unión entre la	

isoleucina (posición 23) de la proteína receptora, y los aminoácidos 16 y 17 (prolina y serina) del ligando. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.....	31
Figura 17. Visualización con PyMOL de la zona de interacción entre el tercer dominio del Modelo 1 (verde) y la proteína CRISP-3 (azul). .....	32
Figura 18. Visualización con PyMOL de la mejor configuración (Conf. 595) del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	33
Figura 19. Visualización con PyMOL de los enlaces correspondientes a la primera zona de contacto entre las dos moléculas. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.....	33
Figura 20. Visualización con PyMOL del enlace correspondiente a la segunda zona de contacto entre las dos moléculas. En la parte derecha de la imagen, el enlace entre el ácido aspártico (posición 303) de la proteína receptora y la tirosina (posición 222) del ligando. En el centro, se muestran las distancias entre los dominios 3 y 4 de A1BG y CRISP-3. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.....	34
Figura 21. Visualización con PyMOL de la configuración 7310 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	40
Figura 22. Visualización con PyMOL de la configuración 59 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	40
Figura 23. Visualización con PyMOL de la configuración 4776 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	40
Figura 24. Visualización con PyMOL de la configuración 2992 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	41
Figura 25. Visualización con PyMOL de la configuración 8865 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	41
Figura 26. Visualización con PyMOL de la configuración 7297 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	41
Figura 27. Visualización con PyMOL de la configuración 4612 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). .....	42

Figura 28. Visualización con PyMOL de la configuración 4772 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	42
Figura 29. Visualización con PyMOL de la configuración 9414 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	42
Figura 30. Visualización con PyMOL de la configuración 6007 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	43
Figura 31. Visualización con PyMOL de la configuración 8249 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	43
Figura 32. Visualización con PyMOL de la configuración 3163 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	43
Figura 33. Visualización con PyMOL de la configuración 2449 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	44
Figura 34. Visualización con PyMOL de la configuración 7748 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	44
Figura 35. Visualización con PyMOL de la configuración 6404 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	44
Figura 36. Visualización con PyMOL de la configuración 3853 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	45
Figura 37. Visualización con PyMOL de la configuración 6384 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	45
Figura 38. Visualización con PyMOL de la configuración 787 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul). ....	45

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

El modelado comparativo de proteínas es un método que permite predecir, mediante aproximaciones, la estructura atómica de una cadena polipeptídica en base a estructuras conocidas. Las estructuras conocidas se denominan moldes o “templates”, y la elección de estos moldes se basa en métodos de homología en base a la secuencia problema o diana. En el caso que no exista una secuencia homóloga con estructura conocida, hay la alternativa del modelado “de novo” <sup>(1)</sup>.

En la actualidad, *Protein Data Bank* <sup>(2)</sup> contiene casi 200.000 estructuras de proteínas determinadas experimentalmente. Además, existen más de 1 millón de Modelos de Estructura Computada en las bases de datos *AlphaFoldDB* <sup>(3)</sup> y *RoseTTAFold* <sup>(4)</sup>. El conocimiento de la estructura de las proteínas hace posible el estudio de sus funciones e implicaciones en el organismo. En función de la calidad del modelo obtenido existen diferentes aplicaciones: el análisis del efecto de las mutaciones, la predicción de ligandos, o las actividades catalíticas, entre otras.

Mediante las técnicas actuales de modelado comparativo de proteínas, se pretende predecir la estructura de la proteína A1BG. La glucoproteína alfa 1B fue descubierta en 1963 <sup>(5)</sup>, aunque hoy día aún no se ha identificado su función biológica. El análisis de la secuencia de aminoácidos muestra cinco dominios repetidos, que contienen unos 95 aminoácidos y un puente disulfuro cada uno. Varios dominios, especialmente el tercero, muestran un alto grado de homología con las regiones variables de algunas proteínas miembros de la superfamilia de las inmunoglobulinas <sup>(6)</sup>.

Un estudio realizado en 2004 observó la unión, mediante fuertes fuerzas electrostáticas, de la proteína secretora CRISP-3 con una alta afinidad a A1BG <sup>(7)</sup>. Este complejo ha sido identificado en plasma, aunque se desconoce la función. Aunque no tenemos datos respecto a las funciones asociadas al complejo formado, se sabe que la proteína CRISP-3 está implicada en los procesos de regulación de la fecundación en mamíferos. Además, en ciertos tipos de cáncer de próstata existe una regulación positiva del gen que codifica para la proteína CRISP-3.

Una investigación publicada en 2015 <sup>(8)</sup>, basada en las alteraciones de las proteínas séricas en pacientes con endometriosis, identificó la proteína A1BG como posible marcador eficaz para el diagnóstico de las etapas II, III y IV de la endometriosis. El estudio compara el proteoma sérico de pacientes con endometriosis y sujetos sanos para identificar el conjunto de proteínas más importantes para la predicción de la enfermedad. El análisis de los resultados reveló 3 proteínas (HP, IGKC y A1BG) que podrían identificarse como

marcadores eficaces para el diagnóstico de endometriosis en las fases II, III, y IV. Para el diagnóstico en la fase I solo podrían considerarse como marcadores las proteínas IGKC (región constante de la cadena kappa de inmunoglobulinas) y HP (haptoglobina). El estudio concluye que la proteína A1BG podría contribuir al diagnóstico de la endometriosis a partir de la segunda fase de la enfermedad.

Recientemente, un estudio realizado en un modelo canino de cáncer de mama detectó la sobreexpresión de la proteína A1BG en poblaciones de células tumorales <sup>(9)</sup>. También se han realizado estudios en pacientes con cáncer de pulmón <sup>(10)</sup>, con resultados parecidos. Los resultados indican una relación significativa entre la proteína estudiada y la fase metastásica del cáncer. Esto indica que la proteína A1BG es un posible biomarcador que puede contribuir a reconocer estados avanzados de la enfermedad.

Hoy día, existen diferentes aplicaciones orientadas al estudio de las funciones e interacciones de las proteínas. El conocimiento de la estructura tridimensional de la proteína A1BG puede contribuir al estudio de sus funciones e implicaciones en el organismo, desconocidas hasta el momento.

## 1.2. Objetivos del Trabajo

A continuación, se detallan los objetivos generales y específicos del trabajo:

1. Buscar e identificar las estructuras molde o “templates”.
  - 1.1. Definir los dominios de la proteína A1BG a partir de la secuencia obtenida del navegador genómico de UCSC con la herramienta del servidor *InterPro* <sup>(11)</sup>.
  - 1.2. Buscar estructuras de referencia (“templates”). Encontrar los posibles moldes o secuencias con estructuras conocidas mediante alineamientos utilizando las herramientas *HHpred 4.0* y *PSI-BLAST* <sup>(12)</sup>.
  - 1.3. Seleccionar los moldes y editar los alineamientos. Preparar el archivo con las secuencias alineadas en formato PIR y el archivo Python con las órdenes de ejecución para *MODELLER* <sup>(13)</sup>.
2. Predecir la estructura 3D de la proteína A1BG.
  - 2.1. Modelar la estructura 3D de la secuencia de la proteína A1BG con el programa *MODELLER*.
  - 2.2. Visualizar la estructura resultante con *PyMOL* <sup>(14)</sup>.
3. Evaluar el modelo y las posibles aplicaciones.
  - 3.1. Evaluar la calidad de los modelos obtenidos con *MolProbity* <sup>(15)</sup>.

3.2. Comparar el mejor modelo obtenido con el que genera *I-TASSER*  
<sup>(16)</sup>.

3.3. Simular el acoplamiento molecular (“docking”) con la proteína CRISP-3.

### 1.3. Impacto en sostenibilidad, ético-social y de diversidad

En este TFM se pueden identificar impactos positivos en dos de las tres dimensiones de los ODS (Objetivos de Desarrollo Sostenible 2030, ONU).

En la dimensión **Sostenibilidad**, los resultados de este trabajo pueden contribuir a futuras investigaciones. La proteína A1BG no tiene funciones asociadas hoy día. El conocimiento de la estructura tridimensional de la proteína facilita el estudio de sus funciones e interacciones con otras moléculas. Por esta razón, se considera que los resultados tienen un impacto positivo en el fomento de la innovación (**ODS 9 – “Industry, innovation and infrastructure”**).

En la dimensión **Comportamiento ético y responsabilidad social**, las posibles investigaciones a partir de la estructura obtenida en los resultados de este TFM, podrían contribuir a la creación de empleo. Por lo tanto, en este caso se podría considerar que tienen un impacto positivo en el crecimiento económico, el empleo y el trabajo (**ODS 8 – “Decent work and economic growth”**).

Respecto a la dimensión **Diversidad, género y derechos humanos**, no se considera ningún impacto ni positivo ni negativo en ninguno de los aspectos. El desarrollo, metodología y resultados del TFM incluyen los aspectos técnicos necesarios para el modelado comparativo y posterior acoplamiento molecular.

### 1.4. Enfoque y método seguido

Existen varios programas de software libre para el modelado comparativo de proteínas. A continuación, se describen a grandes rasgos los posibles programas indicados para este tipo de estudio:

- **3D-JIGSAW:** Permite elegir dominios y moldes, y editar los alineamientos.
- **BIOINFO.PL: META:** Es un meta servidor que envía la secuencia a un grupo de servidores públicos, y devuelve una página web con los alineamientos y algunos modelos. Tiene un sistema de evaluación de modelos, pero no todos los modelos incluyen cadenas laterales.
- **ROBETTA:** Permite construir modelos en ausencia de moldes obvios.
- **SWISS-MODEL:** Tiene una excelente evaluación de la calidad del modelo. Su interfaz complementaria *DeepView* permite la manipulación de las estructuras y modelos.

- **MODELLER:** Construye modelos completos basados en alineamientos propios o proporcionados por el usuario.

La intención de este trabajo es establecer la posible estructura 3D de la proteína A1BG basándonos en la construcción de los mejores alineamientos. *MODELLER* es el software libre más indicado para este tipo de trabajos. Además, dispone de varios tutoriales, tanto oficiales como de sus usuarios, que facilitan el uso de esta herramienta.

Una vez obtenido el modelo de la proteína, se envía la secuencia de aminoácidos al servidor *I-TASSER* para comparar los resultados obtenidos con *MODELLER*. El servidor *I-TASSER* genera predicciones de estructura y función de las proteínas a partir de la búsqueda de plantillas estructurales en el *Protein Data Bank*, utiliza algoritmos de última generación para proporcionar las predicciones más precisas.

## 1.5. Planificación del Trabajo

### 1.5.1. Tareas

Enumeración de las tareas para realizar el trabajo:

1. Búsqueda e identificación de las estructuras molde o “templates”.
  - 1.1. Definición de los dominios de la proteína A1BG.
  - 1.2. Búsqueda de estructuras de referencia (“templates”).
  - 1.3. Selección de los moldes y edición de los alineamientos.
2. Predicción de la estructura 3D de la proteína A1BG.
  - 2.1. Modelado de la estructura 3D.
  - 2.2. Visualización de la estructura.
3. Evaluación del modelo y posibles aplicaciones.
  - 3.1. Evaluación de la calidad del modelo.
  - 3.2. Comparar el mejor modelo obtenido con el que genera I-TASSER.
  - 3.3. Acoplamiento molecular a ciegas con CRISP-3.

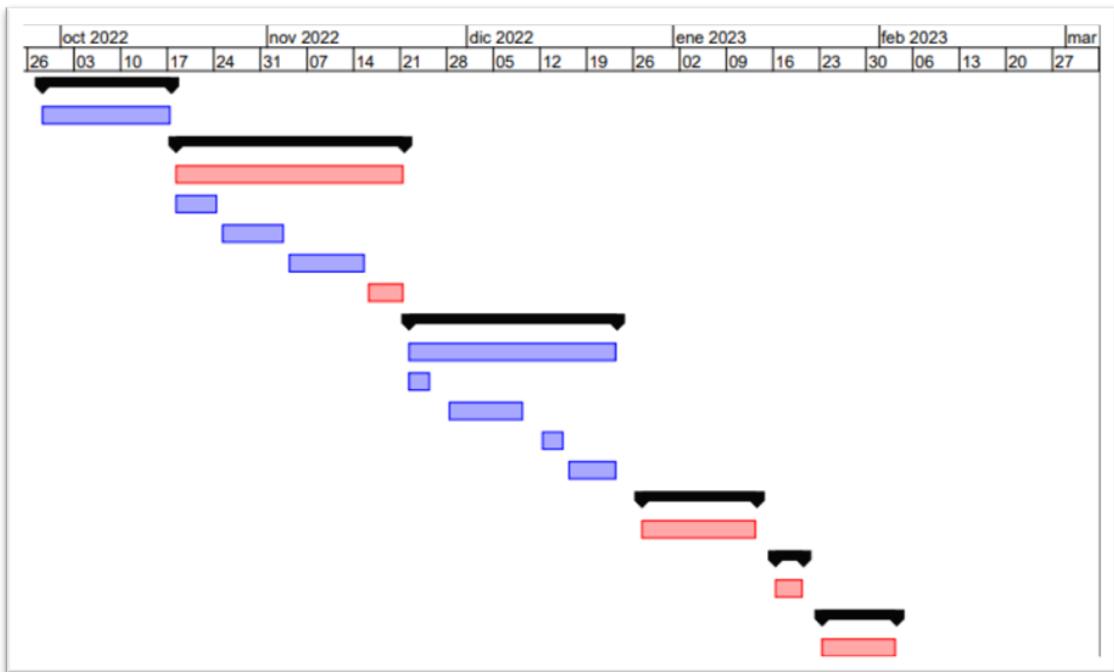
### 1.5.2. Calendario

A continuación, se muestra el calendario al detalle de las tareas del TFM. La planificación se ajusta a los hitos marcados por el Plan Docente de la asignatura, el cálculo de las horas asignadas para cada objetivo puede variar en función de los resultados obtenidos en cada apartado.

Nombre	Tipo	Duración	Inicio	Fin
<b>PEC – 1</b>				
<b>Descripción y plan de trabajo</b>	Hito	14 días	28/09/22	17/10/22
<b>PEC – 2</b>				
<b>Desarrollo del trabajo (fase 1)</b>	Hito	25 días	18/10/22	21/11/22
1.1. Definición de los dominios de la proteína A1BG	Objetivo	3 días	18/10/22	20/10/22
1.2. Búsqueda de estructuras de referencia (“templates”)	Objetivo	10 días	21/10/22	03/11/22
1.3. Selección de los moldes y edición de los alineamientos	Objetivo	8 días	04/11/22	15/11/22
2.1. Modelado de la estructura 3D	Objetivo	4 días	15/11/22	21/11/22
2.2. Visualización de la estructura 3D	Objetivo	4 días	15/11/22	21/11/22
<b>PEC – 3</b>				
<b>Desarrollo del trabajo (fase 2)</b>	Hito	24 días	22/11/22	24/12/22
3.1. Evaluación de la calidad del modelo	Objetivo	7 días	22/11/22	30/11/22
3.2. Comparar el mejor modelo obtenido con I-TASSER	Objetivo	6 días	01/12/22	12/12/22
3.3. Acoplamiento molecular con CRISP3	Objetivo	9 días	13/12/22	24/12/22
<b>PEC – 4</b>				
<b>Cierre de la memoria y presentación</b>	Hito	14 días	27/12/22	15/01/23
<b>Preparación de la defensa</b>	Objetivo	5 días	16/01/23	20/01/23
<b>PEC – 5</b>				
<b>Defensa pública</b>	Hito	10 días	23/01/23	03/02/23
<b>Total</b>		<b>92 días</b>		

**Tabla 1.** Planificación del TFM

Los días de trabajo han sido contabilizados teniendo en consideración todos los días laborables del calendario, de manera que están excluidos los festivos y fines de semana. Esto no significa que, de manera puntual, se pueden utilizar estas horas para realizar alguna tarea de las descritas. Para llegar al mínimo de 300 horas de dedicación, se contempla que cada día se dedicaran un total de 3,5 horas a la realización del trabajo.



**Tabla 2.** Diagrama de Ganntt

### 1.5.3. Hitos

La siguiente tabla muestra los hitos, marcados por el calendario del Plan Docente de la asignatura:

<b>PEC 1 - Descripción y plan de trabajo</b>	28/09/2022	17/10/2022
<b>PEC 2 - Desarrollo del trabajo (fase 1)</b>	18/10/2022	21/11/2022
<b>PEC 3 - Desarrollo del trabajo (fase 2)</b>	22/11/2022	24/12/2022
<b>PEC 4 - Memoria y presentación</b>	27/12/2022	15/01/2023
<b>PEC 5 - Defensa pública</b>	23/01/2023	03/02/2023

**Tabla 3.** Hitos del trabajo

Tal como se ha descrito en el calendario, se han fijado los hitos correspondientes en función de las fechas de entrega determinadas en el Plan Docente:

- **PEC 1 - Descripción y plan de trabajo.** Definición del contexto y ámbito de trabajo y planificación de las tareas.
- **PEC 2 - Desarrollo del trabajo (fase 1).** Definición de los dominios de la proteína, búsqueda de moldes o “templates”, y edición de los archivos para iniciar el modelado de la proteína.

- **PEC 3 - Desarrollo del trabajo (fase 2).** Visualización y evaluación del modelo obtenido. En función del modelo obtenido, la estructura podrá ser utilizada para diferentes aplicaciones.
- **PEC 4 - Memoria y presentación del trabajo final.** Escritura de la memoria final del trabajo, con toda la información, los resultados y las conclusiones extraídas. A partir de la memoria del trabajo, realización de la presentación resumiendo los puntos a destacar del trabajo.
- **PEC 5 - Defensa pública.** Defensa del trabajo ante un tribunal.

#### **1.5.4. Análisis de riesgo**

Los siguientes factores han podido repercutir de manera negativa en el cumplimiento de la temporalización planteada:

- Dificultad para encontrar secuencias con un alto grado de homología. Si algún dominio de la proteína no tiene ninguna secuencia homóloga de estructura conocida, será necesario editar “de novo” el molde.
- Modelos de baja calidad. Una vez se obtiene el modelo, debe evaluarse la calidad de este. Si el modelo no tiene suficiente calidad no será válido para futuras aplicaciones.
- El tiempo de procesado de datos de las aplicaciones requeridas para el trabajo. Muchas de las aplicaciones utilizadas en el trabajo realizan búsquedas en bases de datos de gran dimensión y requieren de un tiempo prolongado para procesar la respuesta.

#### **1.6. Breve sumario de productos obtenidos**

En este trabajo se han obtenido 3 predicciones distintas de la estructura de la proteína A1BG, desconocida hoy día. Se incluyen imágenes de los modelos obtenidos, además de diversos gráficos y tablas para la evaluación de las estructuras. Las estructuras han sido visualizadas con *PyMOL* y extraídas en formato PNG para ser incluidas en el texto.

También se incluyen las imágenes de los resultados del acoplamiento con *PyDockWEB*<sup>(17)</sup> entre los modelos generados con MODELLER y la proteína secretora CRISP-3.

Junto con la memoria, se incluye el archivo PDB del modelo computacional de la estructura tridimensional de la proteína A1BG.

#### **1.7. Breve descripción de los otros capítulos de la memoria**

A continuación, se encuentran detallados los capítulos que contiene esta memoria:

1. Introducción: se introduce el contexto del TFM, además incluye la planificación del trabajo junto con los objetivos previstos.
2. Estado del arte: se describen las hipótesis de este TFM.
3. Materiales y métodos: se describen los diversos pasos para el modelado comparativo y el acoplamiento molecular a ciegas entre el modelo obtenido y la proteína CRISP-3.
4. Resultados: se detallan los resultados obtenidos de los diferentes procesos y análisis descritos en el apartado de métodos.
5. Conclusiones: análisis crítico de los resultados.

## 2. Estado del arte

La predicción de la estructura tridimensional de las proteínas tiene distintas aplicaciones, que dependen directamente de la calidad de la estructura obtenida. Algunas de sus aplicaciones son el análisis del efecto de mutaciones, la predicción de ligandos o el estudio de la actividad catalítica de las moléculas.

La hipótesis principal de este TFM es obtener un modelo computacional de la estructura tridimensional de la proteína A1BG con suficiente calidad para poder contribuir al estudio de esta proteína.

La proteína seleccionada para el modelado comparativo (A1BG) no tiene funciones definidas hoy día. Con el modelo obtenido en este trabajo se pretende contribuir al conocimiento de las funciones asociadas.

Se ha descrito, en estudios previos, la unión de la proteína A1BG con la proteína secretora CRISP-3. Este complejo se encuentra unido por fuertes fuerzas electrostáticas en el plasma, entre las que intervienen las fuerzas de Van der Waals.

Con el modelo obtenido a partir del proceso de modelado con MODELLER se pretende simular el acoplamiento entre las dos proteínas. El éxito en la formación del complejo, entre el modelo obtenido y la proteína secretora CRISP-3, nos puede ayudar a contrastar la predicción planteada. Si, por el contrario, no se realiza ningún acoplamiento entre las moléculas, podríamos rechazar el modelo obtenido, ya que no se ajusta a los datos experimentales observados.

### 3. Materiales y métodos

En este apartado se explican los materiales y métodos utilizados para el modelado comparativo y el acoplamiento molecular:

#### 3.1. Pasos previos al modelado de la proteína A1BG

##### Predicción de la estructura secundaria

Para la realización del TFM se ha obtenido la secuencia de la proteína diana en formato FASTA. Para ello, se ha descargado la secuencia de aminoácidos desde el navegador genómico del UCSC (<https://genome.ucsc.edu/>)<sup>(18)</sup> y se ha almacenado en un archivo de texto con el nombre “target.fa” (Figura 1). El genoma de referencia utilizado ha sido *Homo sapiens*, en su versión más reciente (hg38).

```
>ENST00000263100.8 (A1BG) length=495
MSMLVVFLLLWGVTVGPVTEAAIFYETQPSLWAESSESLLKPLANVTLTCQAHLETPDFQL
FKNGVAQEPMVLDSPAIIKHQFLLTGDTQGRYRCRSGLSTGWTQLSKLLELTGPKSLPAPW
LSMAPVSWITPGLKTTAVCRGVLRGVTFLLREGDHEFLEVPEAQEDVEATFPVHQPGNY
SCSYRTDGEALSEPSATVTIEELAAPPPVLMHHGESQVLHPGNKVTLTCVAPLSGVD
FQLRRGEKELLVPRSSTSPDRIFFHLNAVALGDGGHYTCRYRLHDNQNGWSGDSAPVELI
LSDETLPAPPEFSPEPESGRALRLRCLAPLEGARFALVREDRGRRVHRFQSPAGTEALFE
LHNISVADSANYSCVYVDLKKPPFGGSAPSERLELHVDPGPPRPQLRATWSGAVALGRDAV
LRCEGPIPDVTFELLREGETKAVKTVRTPGAAANLELIFVGPQHAGNYRCRYRSWPHTF
ESELSDPVELLVAES
```

Figura 1. Archivo “target.fa” con la secuencia de aminoácidos de la proteína A1BG.

En primer lugar, se ha realizado un análisis de la estructura secundaria de la proteína con el servidor *PSIPRED* (<http://bioinf.cs.ucl.ac.uk/psipred/>)<sup>(19)</sup>. Este servidor proporciona varios métodos de predicción de la estructura de las proteínas. A partir de la secuencia de aminoácidos de la proteína en formato FASTA, se realiza el análisis incluyendo 2 métodos de predicción: *PSIPRED 4.0*<sup>(20)</sup> y *DISOPRED3*<sup>(21)</sup>. El análisis de *PSIPRED 4.0* permite predecir la estructura secundaria de la proteína, incluyendo las hélices, láminas beta y bucles a partir de la secuencia de aminoácidos. En el caso de *DISOPRED3*, nos permite detectar regiones intrínsecamente desordenadas en proteínas, además permite la predecir las hélices transmembrana.

Los análisis de la predicción de la estructura secundaria han sido realizados a través del navegador web de manera interactiva. *PSIPRED* devuelve los resultados en un archivo comprimido en formato ZIP. Los resultados se pueden descargar de manera local o ser visualizados en el propio navegador. Los archivos contienen representaciones gráficas y muy visuales de la estructura secundaria de la proteína.

##### Definición de los dominios

El siguiente paso es la definición de los dominios de la proteína. Se ha enviado la secuencia al servidor *InterPro* (<https://www.ebi.ac.uk/interpro/>), que integra

13 bases de datos distintas de familias de proteínas. *InterPro* proporciona un análisis funcional de las proteínas, clasificándolas en familias y prediciendo los dominios a partir de las bases de datos existentes. Este sitio web está diseñado de manera muy intuitiva, los resultados se muestran en el propio navegador con ilustraciones que muestran la posición de los dominios en función del esquema global de la proteína. A partir del análisis de los dominios, podemos obtener información sobre las familias de proteínas relacionadas, dominios, sitios de interés, realizar búsquedas dentro de la secuencia o navegar a través de las anotaciones de *InterPro*.

### Búsqueda de estructuras de referencia (“templates”)

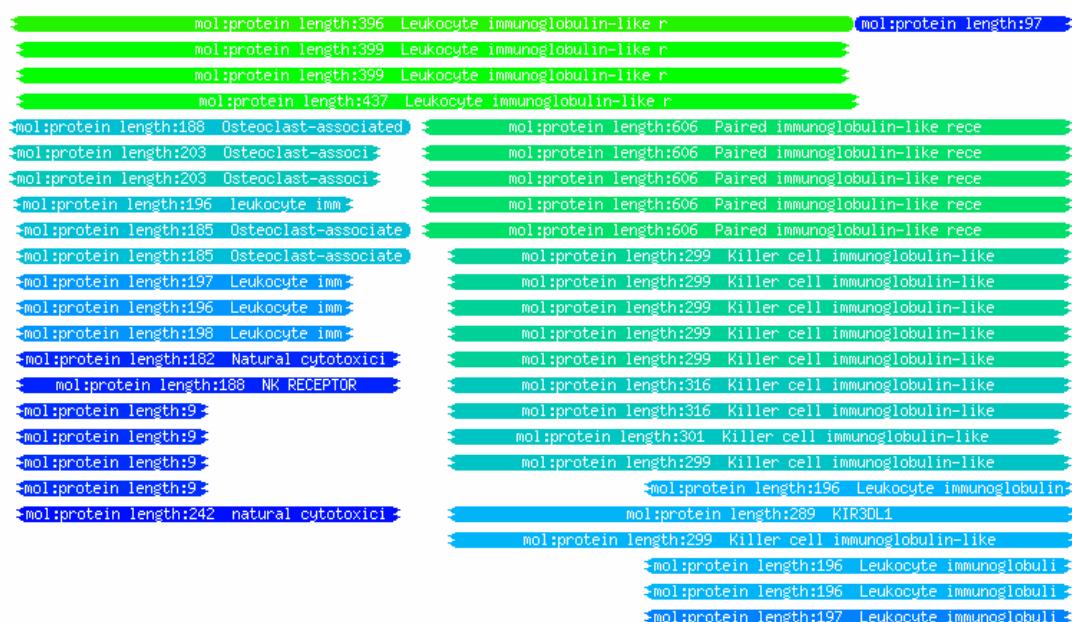
Para seleccionar los moldes necesarios para generar la predicción de la estructura tridimensional de la proteína A1BG se ha utilizado el kit de herramientas de bioinformática del servidor *HHPred* (<https://toolkit.tuebingen.mpg.de/>). Este servidor contiene varias herramientas para el análisis de secuencias de aminoácidos, desde la búsqueda de estructuras homólogas y el análisis de la estructura secundaria, hasta el moldeado de la proteína con *MODELLER*. Existen dos herramientas muy útiles para la búsqueda de estructuras homólogas, *HHPred 4.0* y *PSI-BLAST*.

En primer lugar, se ha seleccionado la búsqueda con la herramienta *ProtBLAST/PSI-BLAST* del servidor *HHPred*, que nos permite comparar la secuencia con las bases de datos de estructuras conocidas. La herramienta, por defecto, tiene la búsqueda seleccionada para la base de datos *AlphaFold*, que contiene todas predicciones de las estructuras de las proteínas obtenidas mediante métodos computacionales y publicadas hasta el momento.

Se han realizado 2 búsquedas distintas con *PSI-BLAST*, la primera seleccionando la base de datos del *Protein Data Bank* (PDB) y la segunda con *AlphaFold*.

Como los valores predeterminados de la aplicación realizan la búsqueda en la base de datos *AlphaFold* (“alphafold\_uniprot50”), para la búsqueda de estructuras homólogas en el *Protein Data Bank* se ha seleccionado la opción “*pdb\_nr\_27\_Aug*”. Los resultados generados por la aplicación se muestran en 3 formatos distintos: un esquema con la localización de las secuencias; una tabla resumen con todos los resultados estadísticos; y finalmente, los alineamientos respectivos con cada secuencia.

La búsqueda de homólogos en el *Protein Data Bank* ha obtenido como resultado un total de 92 secuencias con un E-valor inferior a 1e-3. El examen visual del esquema con la localización de las secuencias (*Figura 2*) revela la ausencia de estructuras homólogas para la secuencia inicial de la proteína estudio. El color de las secuencias varía en función de la puntuación obtenida en el análisis, en rojo los mejores alineamientos, seguido por el verde, hasta el azul oscuro que corresponde a las secuencias con una menor puntuación. Las secuencias con mejor puntuación tienen una baja cobertura (75%), esto implica que algunas estructuras no podrán ser predichas si moldeamos la estructuras a partir de un único molde.



**Figura 2.** Visualización esquemática de los resultados de la búsqueda de homólogos de la proteína A1BG con PSI-BLAST.

Los resultados estadísticos de la búsqueda de homólogos en el *Protein Data Bank* (*Tabla 5*) están ordenados en función de la puntuación total de cada secuencia (“*Bitscore*”). La puntuación del alineamiento varía en función del “*E-value*”, que está relacionado directamente con la cobertura e identidad entre las secuencias. El número de acceso de cada estructura del PDB (“*Accession ID*”) contiene un enlace a la página web del *Protein Data Bank*, de esta manera facilita la descarga del archivo PDB asociado a la estructura.

Nr	Accession ID	Title	E-value	Bitscore	Length	Aligned Positions
1	6aed_A	mol:protein length:396 Leukocyte immunoglobulin-like receptor subfamily B member 2	7.98e-20	94.7449	396	402
2	6aee_H	mol:protein length:399 Leukocyte immunoglobulin-like receptor subfamily B member 1	5.06e-16	83.1889	399	399
3	6aee_G	mol:protein length:399 Leukocyte immunoglobulin-like receptor subfamily B member 1	5.06e-16	83.1889	399	399
4	6zdx_B	mol:protein length:437 Leukocyte immunoglobulin-like receptor subfamily B member 1	1.21e-15	82.4185	437	403
5	6grt_B	mol:protein length:606 Paired immunoglobulin-like receptor B	2.29e-14	79.337	606	302

**Tabla 5.** Resultados de la búsqueda de estructuras de referencia en el PDB con PSI-BLAST para la proteína A1BG.

En general, las secuencias obtenidas en la búsqueda mediante *PSI-BLAST* tienen una baja identidad con la proteína A1BG. El alineamiento de la estructura con mejor puntuación (Figura 3) tiene una identidad del 26% con la secuencia problema.

**1. 6aed\_A mol:protein length:396 Leukocyte immunoglobulin-like receptor subfamily B member 2**

E-value: 7.98e-20, Length: 396, Score: 94.7449 bits (234), Identities: 104/402 (26%), Positives: 175/402 (44%), Gaps: 38/402 (9%)

```

Q  26   ETQPSLWAESSELKPLANVTLTCQAHLETPDFQLFKNGVAQEWPVHLDSPA--KHQF---LLTGDGTQGRYRCRSGLSTGWTQLSKLLEL  110
+ +P+LWAE +S++ + VTL+CQ LE +++L++ + + P + QF +T + GRY C+ W++LS L L
T  3    QPKPTLWAEPDSVITQGSPVTLSCQGSLEAQEYRLREKKSASWITRIRPELVKNGQFHIPSITWEHTGRYGCQYYRSARIWSELSDPLVL  92

Q  111  TGPKSLPAPWLMSMAPVSITPGLKTTAVCRGVLRGVTFLLREGDHFLEV---PEAQEDVEATFPV----HQPGRNYC-SYRTDGE  190
+ P P LS P + T G + T C + F+L +EG+ E + P A+ A F V ++ ++ C Y +
T  93   VMTGAYPKPTLQAQPSPVVTSQQGRVTLQCESQVAFFGFGILCKEGEEEHPQCLNSQPHARGSSRAIFSVGPVSPNRRSHRCYGYDLNSPY  182

Q  191  ALSEPSATVTIEELAAPPPPVLMHHGESSIONQVHLPGNKVTLCVAPLSGVDFQL-RRGEKEL----LVPRSSSTSPDRIFFHNAVALGDG  274
S PS + + P L + V+ PG +T CV+ + F L + GE++L P++ S + F L V+ G
T  183  VNSSPSDNLLELLVPGVSKKPSL--SVQPGPVMAPGESLTLCVSDVGYDRFVLYKEGERDLRQLPGRQPQAGLS--QANFTLGPVSRSYG  268

Q  275  GHYTCRYRLHDNQNQNGWSGDSAPVELILSDETLPAPEFSPEP---ESGRALRLRCLAPLEGARFALVREDRGG----RRVHRFQSPAGT  355
G Y C Y H+ + S S P++++++ + P S +P SG + L C + + F L + R +H +
T  269  GQYRC-YGAHNLSSECAPSAPSDPLDILITGQIRGTPFISVQPGPTVASGENVTLLCQSWRQFHTFLTKAGAADAPLRLRSIHEYPK---Y  354

Q  356  EALFELHNISVADSANYSCVYVDLKPFFGGSAPSERLELHVD  397
+A F + ++ A + Y C P+ S PSE LEL V
T  355  QAEFPMSPVSAHAGTYRCYGSLSNDPYLLSHPSEPLEVVS  396

```

**Figura 3.** Alineamiento con mejor puntuación en los resultados de la búsqueda de estructuras de referencia con *PSI-BLAST* para la proteína A1BG.

Se realiza una segunda búsqueda con *PSI-BLAST*, esta vez en la base de datos *AlphaFold*. Para ello se mantienen los valores predeterminados por la aplicación. El resultado de la búsqueda (*Tabla 6*), en este caso, tiene mejor puntuación respecto a la anterior. La estructura con mejor puntuación pertenece a la proteína A1BG de la especie *Macaca mulatta*, con código de acceso [UniProt F7FJQ0](#). La estructura tiene una identidad del 89% respecto a la proteína estudio y una cobertura del 100% de la secuencia.

Nr	Accession ID	Title	E-value	Bitscore	Length	Aligned Positions
1	F7FJQ0-F1	Alpha-1-B glycoprotein UA=F7FJQ0 UI=F7FJQ0_MACMU OS=Macaca mulatta OX=9544 GN=A1BG	0.00e+00	887.1	549	501
2	A0A1U8C678-F1	alpha-1B-glycoprotein UA=A0A1U8C678 UI=A0A1U8C678_MESAU OS=Mesocricetus auratus OX=1	2.44e-144	431.024	513	518
3	A0A3Q7ME96-F1	alpha-1B-glycoprotein-like isoform X1 UA=A0A3Q7ME96 UI=A0A3Q7ME96_CALUR OS=Callorhinus	2.53e-141	417.927	364	362
4	A0A6P6C2T2-F1	alpha-1B-glycoprotein UA=A0A6P6C2T2 UI=A0A6P6C2T2_PTEVA OS=Pteropus vampyrus OX=1329	1.63e-88	281.567	323	325
5	A0A6P5j8X0-F1	venom metalloproteinase inhibitor DM43-like isoform X1 UA=A0A6P5j8X0 UI=A0A6P5j8X0_P	1.87e-87	284.648	506	486

**Tabla 6.** Resultados de la búsqueda de estructuras de referencia en AlphaFold con *PSI-BLAST* para la proteína A1BG.

En segundo lugar, se realiza la búsqueda de moldes con la herramienta *HHpred 4.0*. Esta aplicación realiza la búsqueda de estructuras en las bases de datos de secuencias y genera la predicción de la estructura con la aplicación *MODELLER* en el mismo servidor. Esta herramienta es mucho más sensible que *BLAST* o *PSI-BLAST* para encontrar homólogos remotos. Se realiza la búsqueda de homólogos de la secuencia con los valores predeterminados (por defecto, se encuentra seleccionada la base de datos del *Protein Data Bank*) para la proteína A1BG. Los resultados de la aplicación se pueden visualizar de manera esquemática, en tablas resumen o bien podemos acceder a cada alineamiento generado, tal y como se ha descrito en el análisis con *PSI-BLAST*. En la siguiente tabla se muestran los 5 primeros resultados, ordenados en función de la puntuación “E-value”:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
□ 1	6GRQ_A	Paired immunoglobulin-like receptor B; neuronal growth inhibition, B cell down regulation, IMMUNE SYSTEM; HET: NAG; 3.3A	100	4.6e-34	301.49	47.9	464	606
□ 2	6GRQ_A	Paired immunoglobulin-like receptor B; neuronal growth inhibition, B cell down regulation, IMMUNE SYSTEM; HET: NAG; 3.3A	100	1.8e-33	297	47.1	459	606
□ 3	6ZDX_B	Leukocyte immunoglobulin-like receptor subfamily B member 1; RIFIN, LILRB1, inhibitory immune receptor, malaria, infecte	99.97	1.8e-27	238.73	37.4	382	437
□ 4	6GRQ_A	Paired immunoglobulin-like receptor B; neuronal growth inhibition, B cell down regulation, IMMUNE SYSTEM; HET: NAG; 3.3A	99.96	1.4e-26	244.65	36.8	372	606
□ 5	6GRQ_A	Paired immunoglobulin-like receptor B; neuronal growth inhibition, B cell down regulation, IMMUNE SYSTEM; HET: NAG; 3.3A	99.95	4.1e-25	233.47	36.8	375	606

**Tabla 7.** Resultados de la búsqueda de estructuras de referencia con *HHpred* para la proteína A1BG.

El resultado del análisis muestra distintas estructuras respecto a la búsqueda con *PSI-BLAST*. Los alineamientos resultantes incluyen el porcentaje de similitud entre las secuencias, siendo 0 el menos similar y 1 para secuencias idénticas. La estructura con mejor puntuación tiene una similitud de 0.31 respecto a la secuencia problema. La identidad entre las secuencias sigue siendo muy baja (23%) aunque existe una mayor cobertura (94%). En este caso, tampoco se encuentran estructuras de referencia para la secuencia inicial de la proteína, que corresponde con los primeros 25 aminoácidos.

### Selección de los moldes o “templates”

Las estructuras homólogas que pertenecen a *Protein Data Bank* encontradas con las herramientas *HHpred* y *PSI-BLAST* no tienen suficiente identidad para realizar una buena predicción basándonos en un solo molde. Para seleccionar los mejores moldes disponibles, se envía la secuencia de la proteína al programa *GenTHREADER* <sup>(22)</sup>. Este programa inicia el reconocimiento de

pliegues de la proteína estudiada con las proteínas disponibles en el PDB para establecer una puntuación que nos permitiría seleccionar las estructuras que mejor predicen el modelo. La evaluación de los pliegues de la proteína A1BG en el programa *GenTHREADER* (*Tabla 2*) devuelve 15 resultados significativos de alineamientos con 13 proteínas diferentes.

Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Length	Target Len	Query Len	Fold
CERT	92.635	2e-08	-321.0	-13.9	386.0	467	536	495	6grqA0
CERT	83.164	2e-07	-238.8	-5.4	358.0	456	578	495	6grsA0
CERT	71.775	3e-06	-184.1	-17.7	271.0	432	691	495	5oj2A0
CERT	69.665	5e-06	-299.9	-5.4	263.0	374	389	495	6aedA0
CERT	66.846	9e-06	-257.8	-10.7	243.0	373	475	495	5k6wA0
CERT	64.862	1e-05	-128.0	-7.5	297.0	181	198	495	4IllaA0
CERT	63.013	2e-05	-244.5	-7.4	256.0	186	198	495	4IllaA0
CERT	62.855	2e-05	-295.0	-4.3	222.0	357	369	495	6aeeeG0
CERT	61.075	3e-05	-300.1	-6.9	216.0	276	281	495	3vh8G0
CERT	60.014	4e-05	-235.1	-6.4	241.0	177	179	495	2d3vA0
CERT	59.776	5e-05	-108.0	-16.6	212.0	405	815	495	6iaaA0
CERT	58.927	6e-05	-130.0	-13.5	230.0	263	292	495	5xeqb0
CERT	58.342	7e-05	-111.4	-7.3	255.0	185	194	495	4II9A0
CERT	57.484	8e-05	-191.5	-7.8	230.0	184	194	495	4II9A0
CERT	56.872	9e-05	-112.9	-9.2	228.0	265	375	495	5k6zA0

Showing 1 to 15 of 15 entries (filtered from 70 total entries)

**Tabla 8.** Extracto de los resultados de GenTHREADER proporcionados por el servidor PSIPRED.

Se han seleccionado un total de 9 estructuras (con resultados significativos en la evaluación de los pliegues de la proteína con *GenTHREADER*) para el modelado de la proteína:

Código	Descripción	Organismo
4Illa	Estructura cristalina del dominio D3D4 de la molécula LILRB2. Sistema inmune.	<i>Homo sapiens</i>
3vh8	KIR3DL1 en complejo con HLA-B. Sistema inmune.	<i>Homo sapiens</i>
5oj2	Estructura cristalina del ectodominio MGDA1. Adhesión celular.	<i>Gallus gallus</i>
6aed	Estructura cristalina de los cuatro dominios similares a las inmunoglobulinas de LILRB2 (LIR2/ILT4/CD85d). Sistema inmune.	<i>Homo sapiens</i>
6grq	Receptor B similar a la inmunoglobulina emparejada (PirB) o miembro 3 de la subfamilia B del receptor similar a inmunoglobulina leucocitaria (LILRB3) dominio extracelular completo. Sistema inmune.	<i>Mus musculus</i>
6grs	Receptor B similar a la inmunoglobulina emparejada (PirB) o miembro 3 de la subfamilia B del receptor similar a inmunoglobulina leucocitaria (LILRB3) dominio extracelular completo. Sistema inmune.	<i>Mus musculus</i>
6aee	Estructura cristalina de los cuatro dominios similares a las inmunoglobulinas de LILRB1 en complejo con HLA-G. Sistema inmune.	<i>Homo sapiens</i>
5k6w	Dominios de inmunoglobulina 1-5. Adhesión celular.	<i>Mus musculus</i>
6iaa	Ectodominio hRobo2. Proteína de señalización.	<i>Homo sapiens</i>

**Tabla 9.** Selección de estructuras del PDB para el moldeado.

Las estructuras homólogas encontradas en la base de datos *AlphaFold* mantienen una alta identidad y cobertura con la secuencia, por lo tanto, seleccionamos la estructura de código **F7FJQ0** para generar un segundo modelo.

Una vez seleccionados los moldes, se han descargado localmente los archivos PDB de las estructuras correspondientes desde la base de datos del *Protein Data Bank* y *AlphaFold*, respectivamente.

### Edición de los alineamientos

Para generar los alineamientos es recomendable utilizar un visor, en este caso se ha usado *JalView*<sup>(23)</sup>, que permite el alineamiento de las secuencias con distintos métodos, además de la modificación de este.

Se han seguido los siguientes pasos:

1. Extraer la secuencia de aminoácidos en formato FASTA a partir del archivo PDB descargado en el apartado anterior. Para ello, se ha utilizado el programa [pdb2fasta](#) disponible online.
2. Abrir la aplicación de *JalView*, instalada previamente en el equipo. Se ha seleccionado la entrada de datos para el alineamiento desde un “área de texto”; primero se introduce el “target” y seguidamente los “templates” escogidos en formato FASTA. Una vez cargadas las secuencias en el programa, se procede al alineamiento. En este trabajo se han generado 2 archivos de alineamientos, el primero con las estructuras obtenidas del *Protein Dada Bank* (9 moldes) y el segundo con la estructura que hemos obtenido de la base de datos *AlphaFold* (1 molde).
  - Los alineamientos de las secuencias descargadas desde la base de datos del *PDB* se han alineado manualmente basándose en los alineamientos obtenidos mediante las herramientas *PSI-BLAST* y *HHpred*.
  - La secuencia descargada desde la base de datos de *AlphaFold* se ha alineado mediante la herramienta *Clustal* incorporada en la misma aplicación.
3. En el caso que existan partes de la secuencia sin homólogos, se debe valorar eliminar los aminoácidos correspondientes para ajustar el alineamiento. En el alineamiento de las secuencias del *PDB*, la región inicial de la proteína no contiene homólogos; aun así, como no se conocen funciones asociadas de la proteína hasta el momento, se decide conservar la secuencia.
4. Una vez alineadas las secuencias, se extrae el archivo en formato PIR. Este archivo contiene las secuencias ya alineadas en un formato compatible con *MODELLER*, aunque para su utilización se deben modificar algunos datos.

5. Preparar el archivo PIR del alineamiento (*Figuras 4 y 5*) para que *MODELLER* pueda generar el modelado. Para ello, se han seguido los pasos del tutorial de [modelado avanzado](#) basado en múltiples plantillas.

```
>P1;6grq
structureX:6grq:123 :A:610 :A:undefined:undefined:-1.00:-1.00
-----TPSLLAQASPVVTSGGYVTLQCES-W-HNDHKFILTVEG----P
QKLSWTQDSQ-----YNYSTRKY--H-ALF---SVGPVTPNQ-----RWICRC--YSY--D
RN-RPYVWS-PPSESVELLVSGNLQKPTIKAEPGSVITSKRAMTIWC--QGNLDAEVYFL-----
-HNEG---SQKT---QST---QTLQQPGNKGKFFI---PSMTRQH--AGQYRC-YCYGSAGW
SQPSDTLELVVTGIYEH---YKPRLSVLPS--PVVTAGGNMTLHCAS--D-F--HY--DKFI-L-TK
-EDKK---F---G---NSLDEHISSSRQYRALFIIGPTTPHTGTFRCYGYFKNAPQLWSVP
SDLQQILISGLS-K---KPSLLTHQ---GHILDPE-GMTLQCYS---D-----INYD-RFA
LHKV---GGAD---IMQHSSQQT---DTGF--SVANFTLGYVSSSTGGQYRCYGA-HNLSS-EWSAS
SE-PLDILITG-QLPLTPS-LSV-----DTFILSKEG--SAQQPL
R-----RCY-GAQ--NSSFYLLSSAS-----
*
>P1;6grs
structureX:6grs:27 :B:315 :B:undefined:undefined:-1.00:-1.00
-----PKPILRVQPDS--VVSRTWKVTFCEE--T-I--GA--NEYR-L-YK
DGKLY---K---T---VT---KNNKAESLSNVDSLNAQQECSY--STQYKSSGY
SDPLKLVVTGHYWTPSLLAQASPVV---TS-----GGYVTLQCES-----W-----HNDH-KFI
LTVE---GP---QKLSW-TQDSQYNYSTRK--YHALFSVGPVTPNQRWICRCYSY-DRNRPYVWSPP
SE-SVELLVSG-NLQ-KPT-IKAEPGS--VITSKRAMTIWC--Q---G-N-LDAEVYFLHNEGSQKTQSTQ
T---L---Q-Q---PGN---KGKFFIPSMTROHAGQYRCYCYGSAGW-----SQPSDTLELVV-T---
*
>P1;5oj2
structureX:5oj2:152 :A:610 :A:ad_celular:gallus_gallus:-1.00:-1.00
-----TVFLRCTVNS-NPPARFIWKRGA----E
TL-----NGVDIYEPLYTQGE--T-KVL---KLKNLRPQD-----YASYTC--QVSVRN
VC-S---IP-DKSITFQLNTTAPPALKLSVNETLVNPNGDNVTMQC--SLTGGDPQPEV-----
-LWSH---SPGP---LPP---NS-LVQGGNLNTIWR---I---RVEDSGYYNC-TAI-NNVG
NPAKTVNLLVRSMKNATFQITPDVIKESETIQLGQDLKLSCHVDAVPQEKKV-Y---SW---YK--NGKP
ARFSDRLLI---TRN-----DPELPPVTCSELIIDLRFSDYGTLYLCVATFQG---APIPDLSE
VNISSETVP---PTISVPKG---QSTITVREGSRAELQCEVRGKPKP---P-----IIWS-RVD
KET-----PMPSG--TM-----TVETYDGKL--RLESVSRDMSGTYKCQTARYN-G-FNIRPREALVQ
LN-VQFPPV-V-EPAFQDV--RQGMG--RSVTLRCTMLKG--SPMKVA-T-SVWRFNG--TLLA-----
-QQDY-S---ELK---VDSVSRETSGSYECS-----
*
>P1;6aed
structureX:6aed:3 :A:397:A:undefined:undefined:-1.00:-1.00
-----QPKPTLWAEPDSVITQGSPVTLSCQG-S-LEAQEYRLYREK----K
SAS-----W---ITRIRPELVKN-GQFHIPSI---TWEH-----TGRYGC--QYY--S
RA-R---WS-ELSDPLVLVMTGAYPKPTLSAQPSPVTSGGRTVLQC--ESQVAFGGFIL-----
-CKE-----QCLNSQPH-ARGSSRAIFSVGPVSPN---R--RWSHRCYGYDLNSPY
VWSSPSDLLELLVPGVSKKPSL--SVQPGPVMAPGESLTLQCVS---DVGYDR-F--VL--YKE--GERD
-----LRQLPGRQP-----QAGLS--QANFTLGPVSRSYGGQYRC-YGAHNLSSSECAPSPLD
ILITGQIRG---TPFISVQPGPTVAS-----GENVTLLCQS-WRQFH---T-----FLLT-KAG
AADA--PLRLRSIHE--YP---K---YQAEF--PMSPVTSAHAGTYRCYGSLNSDP-YLLSHPSEPLE
LV-VS-----
*
>P1;5k6w
structureX:5k6w:22 :A:459:A:undefined:undefined:-1.00:-1.00
```



```

GE-A---VS-RNASLEVALLRD-----DFRQNPTDVVA--AG--EPAILECQPPRGHPEPTIY
WKKDKVRIDDKEERISIRGGKLM-----SN-TRKS-----D----A---GMYTC-VGTN-----
--MVGERDSDPAELTVF-----ERPT-FLRRPINQWLEEEAVEFRCQV---QGD---PQ---PTVR-W-KK
-DDAD-----L----P----RG----RYDIKDDYTL-RIKKTMSTDEGTYMCIAE----NRVGKM
EASATLT-----RA---PPQFVVRP---RD-QIVAQGRTVTFCETKGNPQPAVFHQ----KEGS-QNL
LFPN---QPQQ----PNSRC--SV----SPTG---D-L-TITNIQRSDAGYYICQALTVA----GSIL
AKAQLEVTDVL-TDR-PPP-IIHQGPANQTLAVDGTALLKCKAT---GDP-LPVISW--LKEG--FTFPGR
D----P----RAT----IQE---QGTLQIKNLRISDTGTYTCV-ATS-----
*
>P1;a1bg
sequence:a1bg: : : : :: 0.00: 0.00
MSMLVVFLWGVTVWGPVTEAAIFYETQPSLWAESESLLKPLANVLTCA-H-LETPDFQLFKNG----V
AQE-----P-----VHLDSPAI--K-HQF---LL--TGDT-----QGRYRC--RSG--L
ST-G---WT-QLSKLLELTGPKSLPAPWLSMAPVSWITPGLKTTAVC--RGVLRGVTFL-----RREG
-----DHEF-----LEV---PE-AQEDVEATFPV----H----Q--PGNYSC-SYRTDGE
ALSEPSATVTIEELAAP----PPPVLMMHGESSQVLHPGNKVTLCVA---P-L---SG---VDFQ-L-RR
-GEKE-----L----L----VP----RSSTSPDRIFFHLNAVALGDGGHYTCYRLHDNQNGWSGD
SAPVELLSDETLP---APEFSPEP---ES-----GRALRLRCLA-----P-----LEGA-RFA
LVRE---DRGG----RRVHR--FQ----SPAG--TEALFELHNISVADSANYSCVYV-DLKPPFGGSAP
SE-RLELHVVG-PPP-RPQ-LRATWSG--AVLAGRDAVLRC--E---G-P-IPDVTFELLREG--ETKAVK
T---V---R-T---PGA---AANLELIFVGPQHAGNYRCR-YRS--WPHTFESELSDPVELLV-AES
*
```

**Figura 4.** Archivo “9templates.pir”.

```

>P1;template
structureX:template:49:A:549:A:alpha_A1BG:mr:1.90: 0.19
MSMLVAFLFWGLPMGPVTEAAIFYETQPSLWAESESLLKPSANVMLTCQARLGTPDFQLFKNGVAQEPVHL
DSPAIKHQFLLTGDTQGRYCRSGLFTGWTQLSKLLELTGPKSLPPPWLSPMTVSWITPGLNTTAVCRGGLR
GVTFLLKREGDHEFLEMPEAQEGVEAAFPVHQAGNYSCSYRTHVAGALSEPSATVIIELAAPPPVLMHHG
ESAQVLRPGNKVTACLAAPLSGVHFQLWRGEKEELLVPRSSSTSPDRIFFHLNAVALGDGGHYTCYRLHDNEN
AWSGDSAPVELLSDETLPAPEFTAEPASNPEPGTRVRLRCLAPLEGARFALVREDGGGHRVHRLQSPAGT
EALFELHNISVADSANYSCVYVDLKPFFGGSAPSERLELRVDGPPPRPQLRAAWSGAVLAGRDAVLRC
EGPIPDVTFELLREGETKAVDTVHTPEAAAANFELTFVGPQHAGSYRCYRSSWGPNTFESELSDPVELLVAES*
>P1;a1bg
sequence:a1bg:49:A:549:A:glicoproteina:hs:2.00:-1.00
MSMLVVFLWGVTVWGPVTEAAIFYETQPSLWAESESLLKPLANVLTCAHLETPDFQLFKNGVAQEPVHL
DSPAIKHQFLLTGDTQGRYCRSGLSTGWTQLSKLLELTGPKSLPAPWLSMAPVSWITPGLKTTAVCRGVLR
GVTFLLRREGDHEFLEVPEAQEDVEATFPVHQPGNYSCSYRTDGEGALSEPSATVIIELAAPPPVLMHHG
ESSQVLHPGNKVTLCVAPLSGVDFQLRRGEKEELLVPRSSSTSPDRIFFHLNAVALGDGGHYTCYRLHDNQ
GWGDSAPVELLSDETLPAPEFSPEP----ESGRALRLRCLAPLEGARFALVREDGGGRRVHRFQSPAGT
EALFELHNISVADSANYSCVYVDLKPFFGGSAPSERLELRVDGPPPRPQLRATWSGAVLAGRDAVLRC
EGPIPDVTFELLREGETKAVKTVRTPGAAAANLELIFVGPQHAGNYRCYR-SWPHTFESELSDPVELLVAES*
```

**Figura 5.** Archivo “1template.pir”.

## Archivo de comandos MODELLER

Finalmente, se debe preparar el script en *Python* que contiene las órdenes para que el programa MODELLER pueda generar el modelo. Para ello se ha utilizado un script más avanzado (*Figura 6*), en donde se ha incorporado la evaluación final de los modelos en base al potencial DOPE de las estructuras obtenidas en el proceso.

```

from modeller import *
from modeller.automodel import *

env = Environ()

env.schedule_scale = physical.Values(default=1.0,soft_sphere=0.7)
env.io.hetatm = True

a = AutoModel(env, alnfile='9templates.pir', knowns=('5k6w', '4lla', '5oj2',
'6aed', '6grq', '6grs', '3vh8', '6aee', '6iaa'), sequence='a1bg',
assess_methods=(assess.DOPE, assess.GA341))
a.starting_model = 1
a.ending_model = 50
a.deviation = 4.0

# Amount of randomization between models
a.generate_method = generate.transfer_xyz
a.rand_method = randomize.xyz

# Very thorough VTFM schedule
a.library_schedule = autosched.slow
a.max_var_iterations = 300

# Thorough MD optimization:
a.md_level = refine.very_slow

a.final_malign3d = True

a.make()

# Get a list of all successfully built models from a.outputs
ok_models = filter(lambda x: x['failure'] is None, a.outputs)

# Rank the models by DOPE score
key = 'DOPE score'
ok_models.sort(lambda a,b: cmp(a[key], b[key]))

# Get top model
m = ok_models[0]
print "Top model: %s (DOPE score %.3f)" % (m['name'], m[key])

```

**Figura 6.** Archivo “9templates.py”.

### 3.2. Modelado de la proteína A1BG con MODELLER

Para poder generar la predicción del modelo con MODELLER, primero se debe descargar la aplicación e instalarla de manera local. En este caso se ha utilizado una licencia destinada a la educación.

Se accede a la aplicación y mediante la función “**cd**” se indica la ruta de la carpeta que contiene los archivos necesarios para el modelado. En este trabajo, se ha creado una carpeta con nombre “Modeller” en el escritorio, de modo que la ruta para acceder a la carpeta es la siguiente: “**cd \Users\Anna\Desktop\Modeller**”.

Una vez indicada la ruta que contiene todos los archivos necesarios para ejecutar (**.pdb**, **.pir** y **.py**), se inicia el modelado con la función “**mod10.3 archivo.py**”. En función de la versión de *MODELLER* instalada en el equipo, esta función puede variar.

Este proceso puede tardar varios minutos. Si el número de moldes que se desea obtener es elevado, esto puede incrementar la duración del proceso. Además, el moldeado con varios “templates” también precisa de más tiempo para procesar los modelos. En este caso, para cada modelado, se han realizado 50 modelos distintos, y posteriormente evaluado para obtener el mejor puntuado. El tiempo medio para procesar los modelos es de unas 4 horas para cada modelado.

Al finalizar el proceso, los resultados con el mejor modelo se pueden visualizar en el archivo LOG generado por el mismo programa. En el final del archivo encontramos los resultados de la estructura con mejor puntuación DOPE.

La estructura con mejor puntuación seleccionada para cada proceso de modelado se visualiza con *PyMOL*. Este visor permite la visualización de la estructura tridimensional y modificación de los archivos PDB. En este trabajo, nos interesa observar la estructura secundaria de la proteína para poder compararla con la predicción de la aplicación *PSIPred*, para ello se ha seleccionado el color en función de las estructuras secundarias (“color > by ss”), a través de los comandos laterales. También se muestran los enlaces disulfuro que contienen las cadenas en forma de esferas (“show > disulfides > spheres”), de esta manera es más fácil distinguir las estructuras para poder compararlas con las predicciones obtenidas.

### **3.3. Modelado de la proteína A1BG con I-TASSER**

*I-TASSER* es una plataforma para predecir la estructura y función de las proteínas a partir de una secuencia de aminoácidos conocida. Identifica las plantillas a partir de la base de datos del *Protein Data Bank* para formar la estructura de la nueva proteína. Este servidor en línea permite generar una estructura a partir de los valores predeterminados de la aplicación, también se pueden añadir restricciones y plantillas adicionales, o excluir plantillas no deseadas.

La secuencia de aminoácidos de la proteína A1BG se envió al servidor *I-TASSER* para su predicción sin restricciones añadidas. El servidor envía los resultados en el correo electrónico al finalizar el proceso de modelado y análisis posterior de los modelos. Este proceso puede durar varias horas, incluso días.

Los resultados son proporcionados a través de un enlace que contiene: el análisis de la estructura secundaria de la secuencia; el análisis de los residuos; las 10 mejores plantillas utilizadas por *I-TASSER*; los 5 mejores modelos resultantes; además de las posibles anotaciones biológicas y términos GO asociados a la proteína. En este caso, solo nos interesa obtener el mejor modelo predicho con *I-TASSER*, el resto del análisis nos puede servir en el caso que tengamos que mejorar el modelo obtenido con *MODELLER*. *I-TASSER* clasifica los modelos en función de la puntuación C, es una

puntuación de confianza para estimar la calidad de los modelos predichos. Se ha descargado localmente el modelo con la mejor puntuación C, que corresponde al Modelo 1 (-1.03).

### 3.4. Evaluación de la calidad de los modelos

Se debe evaluar la calidad de los modelos obtenidos para seleccionar la mejor predicción. Para la validación de los modelos se utiliza la aplicación *MolProbity*, que se encuentra disponible en la página del laboratorio Richardson en la Universidad de Duke. Esta aplicación proporciona un análisis completo con las herramientas más tradicionales para la validación geométrica de las estructuras macromoleculares. *MolProbity* permite analizar con precisión la calidad de los modelos de estructura macromolecular tanto a nivel global como local. Es una herramienta, con una interfaz muy intuitiva, que nos proporcionara los datos necesarios para elegir las estructuras apropiadas. El mismo servidor proporciona tutoriales en línea que facilitan el proceso de evaluación, en este caso, se han seguido los pasos indicados en el tutorial de la [evaluación de la estructura de rayos X](#).

En las estructuras determinadas por cristalográfia casi nunca se incluyen los hidrógenos, y esto puede alterar los resultados de las interacciones de las moléculas. Por esto, *MolProbity* recomienda la adición de hidrógenos (fácilmente ejecutable desde la aplicación web) para mejorar la calidad de los análisis. Como en este caso, las estructuras generadas no tienen hidrógenos, los adicionamos para mejorar el contacto entre las moléculas.

El análisis de los contactos y la geometría de todos los átomos mediante *MolProbity* proporciona una fácil evaluación de los resultados de la calidad de las moléculas. Los resultados de los valores estadísticos aparecen en una tabla, coloreados en función del nivel de calidad de la estructura.

*Molprobity* otorga una puntuación global a las estructuras analizadas que permite definir la calidad de la molécula. Esta puntuación engloba los análisis de contactos y geometría de la estructura. Las estructuras con un percentil  $\geq 33$  se pueden considerar aceptables, a partir de  $\geq 66$  se puede decir que la molécula tiene una buena calidad.

El “*Clashscore*” indica el número de contactos o choques estéricos, a medida que aumenta, disminuye la calidad del modelo. El puntuaje ideal es 0, que correspondería con el percentil 100. El rango de aceptación de la calidad de las moléculas es el mismo que utiliza la puntuación otorgada por *Molprobity*.

La puntuación de “*Ramachandran outliers*” enumera el número de puntos en el gráfico Ramachandran que quedan fuera de la zona “ideal” de cada residuo. El número de valores atípicos aceptados para validar la calidad de la molécula es  $<0.5$ , los valores superiores se encuentran desfavorecidos, por lo tanto, aparecerán en rojo en la tabla. Para los residuos en zonas favorecidas (“*Ramachandran favored*”) se aceptan valores  $>95\%$ .

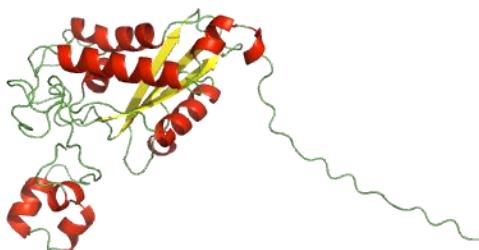
Los resultados de la evaluación de los 3 modelos se muestran en una tabla comparativa. Se han mantenido los colores definidos por la aplicación para facilitar la visualización de los resultados.

### 3.5. Acoplamiento molecular a ciegas con la proteína CRISP-3

El acoplamiento molecular a ciegas permite predecir la unión entre dos moléculas sin conocimiento previo de las zonas de interacción entre las mismas.

En estudios anteriores se ha observado la unión, mediante fuertes fuerzas electrostáticas, de la proteína A1BG con la proteína secretora CRISP-3. Para predecir la unión entre las dos proteínas, se ha utilizado el servidor *pyDockWEB*. Esta herramienta predice la estructura de las interacciones entre dos proteínas a partir de las coordenadas 3D.

En primer lugar, se debe descargar el archivo PDB que contiene la estructura tridimensional de la proteína CRISP-3. Para ello, iniciamos la búsqueda en el navegador genómico del UCSC. El navegador muestra como resultado la estructura de la proteína incluida en la base de datos *AlphaFold* (Figura 7), hoy día aún no se conoce la estructura determinada mediante métodos de cristalografía de rayos X. Se selecciona el enlace correspondiente a *AlphaFold* para descargar el archivo PDB de la proteína CRISP-3 con código *UniProt P54108*.



**Figura 7.** Visualización de la proteína CRISP-3 con PyMOL.

Una vez descargada la estructura, se envían los datos al servidor *pyDockWEB* para generar la predicción del acoplamiento entre las dos proteínas. Se realizan dos predicciones del acoplamiento de la proteína con CRISP-3, uno para cada modelo generado con *MODELLER*. Para generar la predicción del acoplamiento, se deben introducir los archivos PDB de la proteína receptora (Modelo 1 y 2) y el ligando (CRISP-3). Una vez introducidos los archivos, se deben seleccionar las cadenas de las dos proteínas, que en este caso tienen una única cadena. Por último, el servidor permite introducir restricciones de distancias experimentales. Como no se conocen estudios previos de la zona de interacción entre las dos estructuras, no se ha añadido ninguna restricción.

El servidor devuelve las 100 mejores conformaciones del acoplamiento de cuerpo rígido generadas por *FTDock*, puntuadas y clasificadas principalmente en función de la electrostática y la energía de solvatación. Los resultados

pueden tardar varias horas en procesarse debido a que se generan 10.000 configuraciones distintas con *FTDock* para su posterior evaluación.

Una vez finaliza el cálculo, *pyDockWEB* genera un archivo comprimido con GZIP que contiene los siguientes archivos (“[xxx]” es un indicador numérico):

- “**Proyecto[xxx].ftdock**”: datos de salida de la aplicación *FTDock*.
- “**Proyecto[xxx].rot**”: ángulos de rotación para cada una de las conformaciones generadas por *FTDock*.
- “**Proyecto[xxx]\_rec.pdb**”: archivo PDB de la proteína receptora analizado por *pyDock*.
- “**Proyecto[xxx]\_lig.pdb**”: archivo PDB de la proteína ligando analizado por *pyDock*.
- “**Proyecto[xxx]\_setup.log**”: archivo de registro del proceso de *pyDock* (análisis de archivos PDB).
- “**Proyecto[xxx].ene**”: tabla que contiene la lista de conformaciones generadas y clasificadas por *pyDock*.
- “**Proyecto[xxx]\_\* .pdb**”: una de las 100 conformaciones generadas por *pyDock* en formato PDB.

El archivo “Proyecto[xxx].ene” contiene la lista con las puntuaciones de energía de las conformaciones generadas por *pyDock*. Los resultados de la tabla contienen: el número de configuración asignado por *pyDock* (“Conf”), la energía de unión electrostática (“Ele”), la energía de solvatación (“Desolv”), las fuerzas de Van der Waals (“VDW”), la energía total del enlace (“Total”), y finalmente, la posición de la clasificación en función de la energía del enlace asignada por *pyDock* (“RANK”).

Después de analizar los resultados obtenidos, se descarga el archivo (“Proyecto[xxx]\_\* .pdb”) que contiene la configuración con mejor puntuación para su visualización con *PyMOL*. Para diferenciar las estructuras de las dos proteínas se han coloreado las cadenas (“color > by chain > by chain (element C)”).

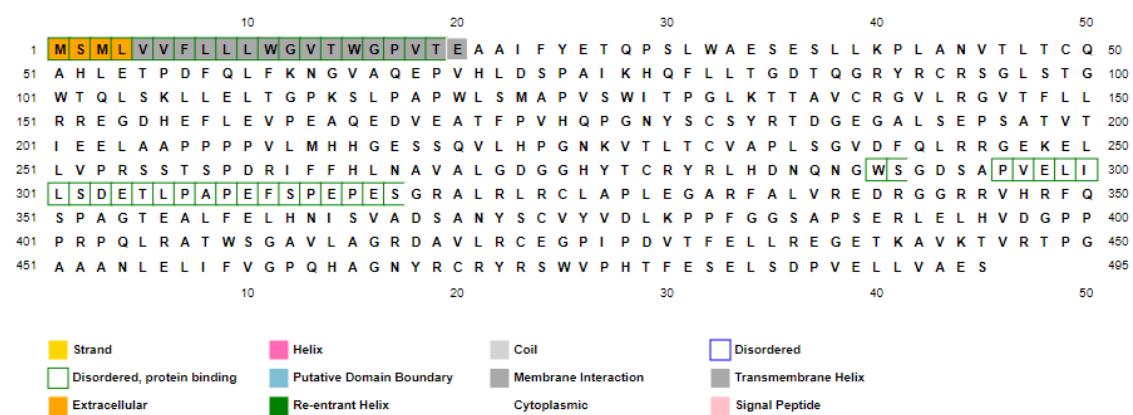
Las cadenas de las proteínas, por defecto, aparecen en forma de dibujo, representando las hélices y láminas que componen la estructura, esto dificulta la búsqueda de contactos entre las dos moléculas. Se ha modificado la visualización de la estructura a esferas (“show > as > spheres”), de esta manera se pueden observar mejor los contactos establecidos entre las proteínas.

La visualización de los enlaces formados entre las dos estructuras se muestra en forma de barras (“show > as > licorice > sticks”). De esta manera, si se amplían las zonas de contacto, se pueden observar las uniones entre los aminoácidos de las dos estructuras.

# 4. Resultados

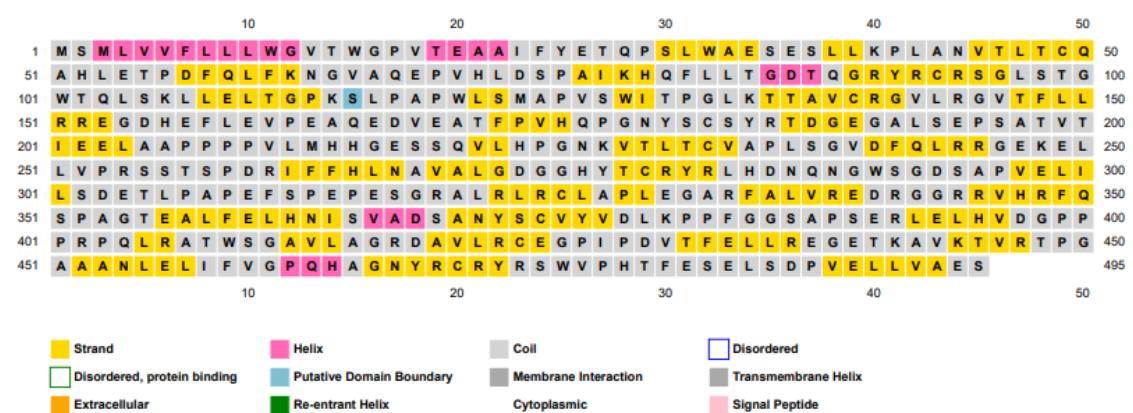
## 4.1. Predicción de la estructura secundaria de la proteína A1BG

El análisis con *DISOPRED3* (*Figura 8*) muestra dos regiones intrínsecamente desordenadas. La primera región corresponde con los 20 primeros aminoácidos. Existe una región corta al inicio de la secuencia que pertenece a una hélice transmembrana. También se observa, al inicio de la secuencia, una pequeña región extracelular. La segunda región intrínsecamente desordenada se sitúa entre los aminoácidos 295 y 317.



**Figura 8.** Resultados del análisis de la estructura secundaria de la proteína A1BG con DISOPRED3.

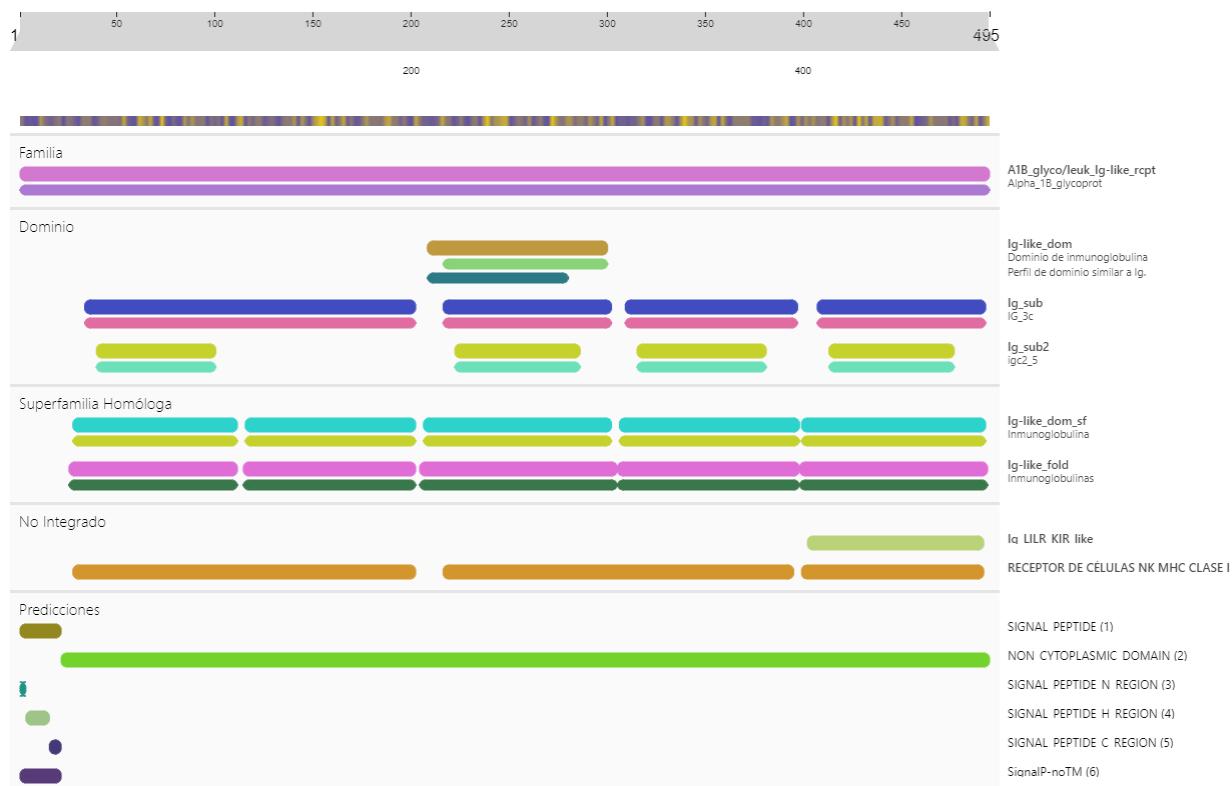
En el análisis con *PSIPRED 4.0* (*Figura 9*) se distinguen 5 hélices  $\alpha$  y varias láminas  $\beta$ . La primera hélice  $\alpha$  está compuesta por 10 aminoácidos y coincide con la hélice transmembrana de los resultados de *DISOPRED3*. La segunda hélice  $\alpha$  está compuesta por 4 aminoácidos, parte de esta secuencia podría tener interacción con la membrana. Las otras 3 hélices  $\alpha$  están compuestas por 3 aminoácidos y se encuentran distribuidas a lo largo de la secuencia.



**Figura 9.** Resultados del análisis de la estructura secundaria de la proteína A1BG con PSIPRED 4.0.

## 4.2. Definición de los dominios de la proteína A1BG

Los resultados de la identificación de los dominios con *InterPro* (*Figura 10*) muestran una estrecha relación con la superfamilia de las inmunoglobulinas.



**Figura 10.** Resultados del análisis de los dominios de la proteína A1BG con InterPro.

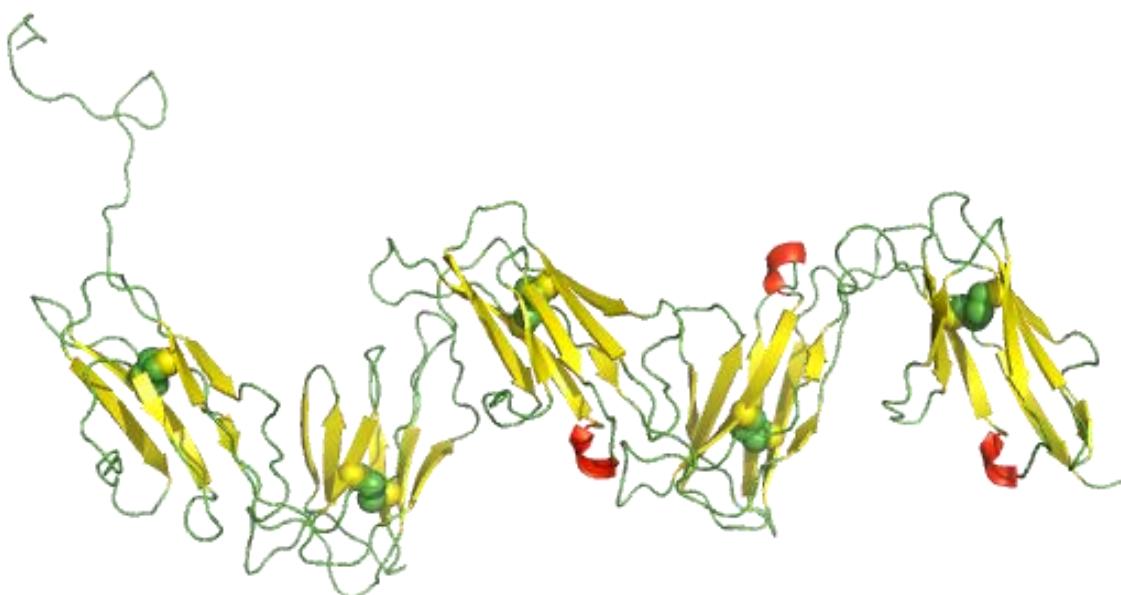
La predicción de los dominios de la proteína muestra 2 regiones:

- Un péptido señal, que contiene en su interior una región hidrofóbica (1-21). Esta predicción coincide con una zona intrínsecamente desordenada de la secuencia de aminoácidos.
- Una región de una proteína extracelular unida a la membrana (22-495). Esta región coincide aproximadamente con 5 dominios definidos de la superfamilia de las inmunoglobulinas. Basándonos en las homologías establecidas entre los dominios predichos de la superfamilia de las inmunoglobulinas, los límites de los dominios de la proteína estudio serían los siguientes:
  1. Dominio 1 (26-111)
  2. Dominio 2 (115-202)
  3. Dominio 3 (205-305)
  4. Dominio 4 (306-398)
  5. Dominio 5 (399-494)

#### 4.3. Modelos de la proteína A1BG generados con MODELLER

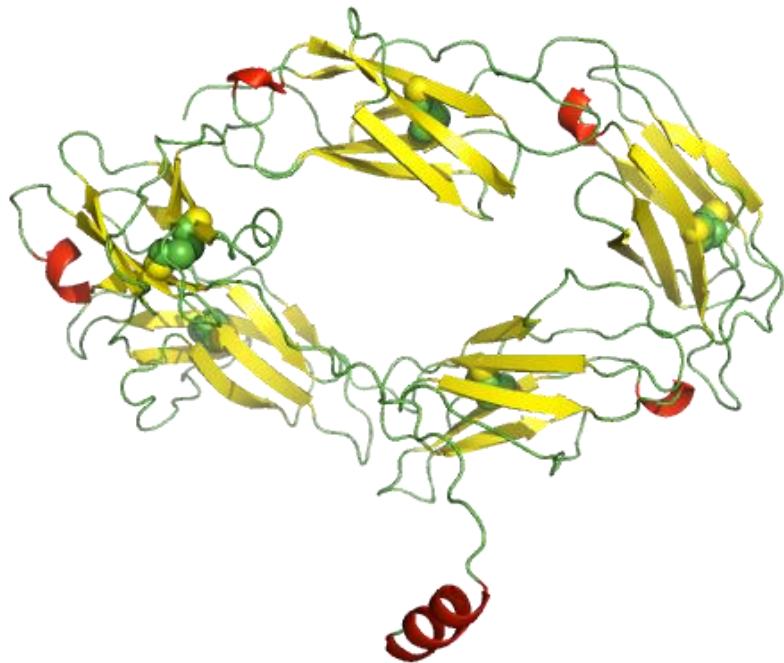
Los modelos generados con MODELLER, obtenidos a partir de las estructuras de las distintas bases de datos (*PDB* y *AlphaFold*), muestran diferencias en la distribución espacial de los dominios que conforman la proteína.

El Modelo 1 (*Figura 11*), obtenido a partir de 9 moldes procedentes del *Protein Data Bank*, tiene una distribución aparentemente lineal entre los dominios. La estructura está compuesta por 5 dominios, bien definidos, con un enlace disulfuro en cada uno de ellos. La secuencia inicial de la proteína (parte izquierda de la *Figura 11*) no contiene las hélices predichas en el análisis de la estructura secundaria de la secuencia de aminoácidos. Se observan 3 pequeñas hélices, compuestas por 3 aminoácidos cada una, distribuidas entre los dominios 3 (271-273), 4 (366-368) y 5 (462-464). Las dos últimas hélices α coinciden con la predicción de la estructura secundaria.



**Figura 11.** Modelo 1 de la estructura de la proteína A1BG, obtenido a partir de 9 moldes del Protein Data Bank.

El Modelo 2 (*Figura 12*), obtenido a partir de un molde procedente de la base de datos *AlphaFold*, tiene una distribución circular entre los dominios que conforman la proteína. La estructura está compuesta por 5 dominios con un enlace disulfuro en cada uno de ellos. En este caso, la secuencia inicial de la proteína (parte inferior de la *Figura 12*) contiene una sola hélice de las dos predichas en el análisis de la estructura secundaria de la secuencia de aminoácidos. Esta hélice está compuesta por 10 aminoácidos (2-11). Además de la hélice inicial, se observan 4 pequeñas hélices α distribuidas entre los dominios de la proteína: la primera se encuentra localizada en el dominio 1 (35-39) y está compuesta por 5 aminoácidos; las demás están compuestas por 3 aminoácidos cada una y se encuentran distribuidas entre los dominios 3 (271-273), 4 (366-368) y 5 (462-464).



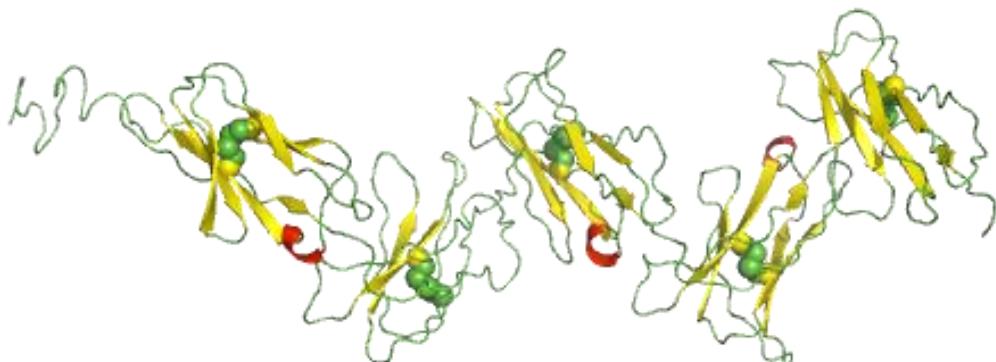
**Figura 12.** Modelo 2 de la estructura de la proteína A1BG, obtenido a partir de 1 molde de AlphaFold.

#### 4.4. Modelo de la proteína A1BG generado con I-TASSER

El resultado de la predicción del modelo de la proteína A1BG generado por *I-TASSER* (*Figura 13*) muestra una estructura similar al Modelo 1.

La estructura está compuesta por 5 dominios con un enlace disulfuro en cada uno de ellos. En este caso, tampoco se han encontrado estructuras homólogas para la secuencia inicial de la proteína, por lo tanto, el modelo carece de la hélice  $\alpha$  que contiene el péptido señal de la proteína.

Se observan 3 pequeñas hélices  $\alpha$ , compuestas por 3 aminoácidos cada una, distribuidas entre los dominios 1 (85-87), 3 (271-273) y 4 (366-368). En este caso, las hélices de los dominios 1 y 4 coinciden con las predicciones de la estructura secundaria de la proteína.



**Figura 13.** Modelo 3 de la estructura de la proteína A1BG, generado con I-TASSER.

#### 4.4. Evaluación de la calidad de los modelos

Los resultados del análisis de la calidad de los 3 modelos con la herramienta *MolProbity* se han englobado en una tabla comparativa (*Tabla 10*). Los resultados de las puntuaciones “*Clashscore*” y “*MolProbity score*” se muestran en percentiles.

		Modelo 1	Modelo 2	Modelo 3
Análisis de contactos	Clashscore	0 <sup>th</sup>	16 <sup>th</sup>	60 <sup>th</sup>
Geometría	Poor rotamers	1.92 %	1.2 %	16.35 %
	Favored rotamers	92.31 %	96.63 %	66.11 %
	Ramachandran outliers	2.84 %	1.22 %	13.79 %
	Ramachandran favored	90.67 %	97.97 %	68.97 %
	Rama distriburion Z-score	-1.09 ± 0.35	0.01 ± 0.34	-6.24 ± 0.29
	MolProbity score	18 <sup>th</sup>	75 <sup>th</sup>	11 <sup>th</sup>
	C $\beta$ deviations >0.25Å	1.97 %	0.44 %	9.65 %
	Bad bonds	0.43 %	0.25 %	0.03 %
	Bad angles	1.62 %	1.14 %	1.60 %

**Tabla 10.** Tabla comparativa de los resultados de la evaluación de calidad de los modelos, analizados con la herramienta MolProbity.

La evaluación de calidad del modelo obtenido a partir de 9 moldes del *Protein Data Bank* (**Modelo 1**) muestra una mala calidad de la estructura. Si nos fijamos en la puntuación “*MolProbity score*”, se obtienen resultados negativos, aunque los valores superan los percentiles obtenidos en la evaluación del Modelo 3. En el análisis de la geometría, la distribución de Ramachandran se mantiene en un rango aceptable.

En cambio, los resultados de la evaluación de la calidad de la estructura generada mediante un molde procedente de la base de datos *AlphaFold* (**Modelo 2**) indican que el modelo obtenido es de buena calidad. El análisis de todos los contactos sigue teniendo una puntuación negativa, aunque se observa una notable mejora respecto a las estructuras anteriores. Respecto al análisis de la geometría, la puntuación “*MolProbity score*” nos indica que la estructura es de buena calidad.

El análisis de la calidad de la estructura del modelo generado con *I-TASSER* (**Modelo 3**) muestra una puntuación desfavorable en términos de geometría. En este caso, el análisis de todos los contactos obtiene una puntuación aceptable, pero el análisis de la geometría en términos de Ramachandran obtiene una puntuación negativa. Además, si comparamos los resultados del análisis de la geometría de la estructura obtenida mediante *I-TASSER* con las demás, es la que obtiene la peor puntuación entre todos los modelos.

Por lo tanto, podemos concluir que la mejor predicción, si comparamos las 3 estructuras obtenidas, es la obtenida con el molde procedente de la base de datos *AlphaFold* (**Modelo 2**). Aunque la estructura obtenida mediante los 9 moldes del *Protein Data Bank* no tiene suficiente calidad, se incluye la predicción del acoplamiento del Modelo 2 con la proteína CRISP3.

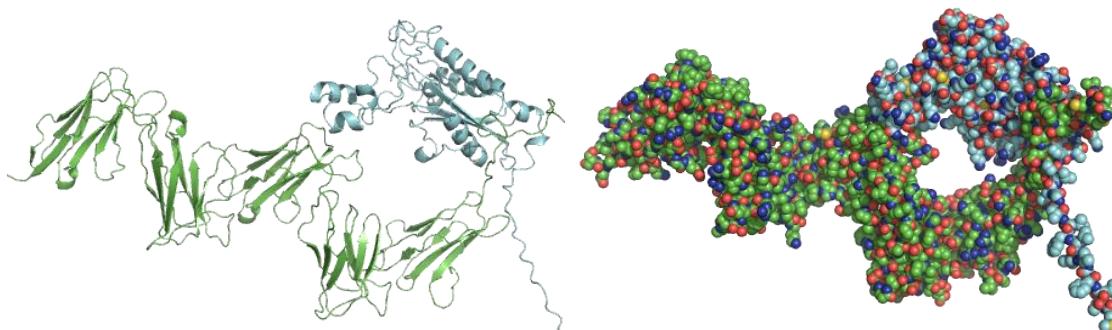
#### 4.5. Acoplamiento molecular a ciegas con la proteína CRISP-3

Los resultados de la simulación del acoplamiento molecular de cuerpo rígido entre el Modelo 1 de la proteína A1BG y la proteína secretora CRISP-3 (*Tabla 11*) son positivos. Las posiciones mejor puntuadas muestran una clara interacción entre las dos proteínas. El sitio de unión entre las dos moléculas es similar en 9 de las 10 conformaciones con mayor energía. La proteína CRISP-3 se une a la secuencia inicial del Modelo 1 de la proteína A1BG mediante fuertes fuerzas electrostáticas.

Conf	Electrostatics	Desolvation	VdW	Total	Rank
9415	-17.328	-46.996	71.697	-57.153	1
7310	-15.376	-47.987	77.375	-55.626	2
59	-9.104	-45.538	71.427	-47.499	3
4776	-11.467	-44.321	80.494	-47.738	4
2992	-10.884	-39.348	49.558	-45.276	5
8865	-9.045	-39.563	29.133	-45.695	6
7297	-10.183	-43.425	89.988	-44.610	7
4612	-15.907	-38.904	126.282	-42.183	8
4772	-7.217	-43.026	84.968	-41.747	9
9414	-6.747	-44.042	95.169	-41.272	10

**Tabla 11.** Resultado de las 10 mejores predicciones generadas por PyDockWEB del acoplamiento entre el Modelo 1 de la proteína A1BG y la proteína secretora CRISP-3.

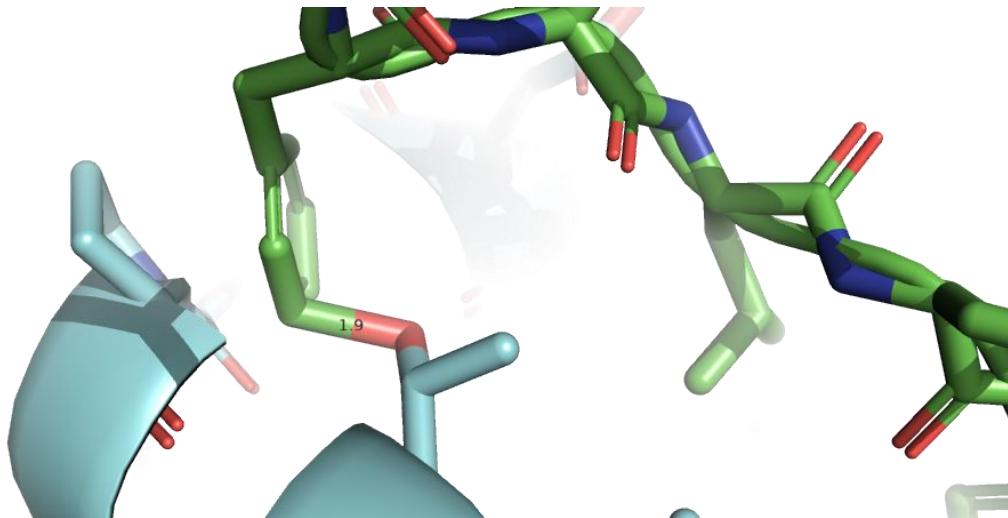
Se ha incluido en los resultados la visualización de la configuración mejor puntuada en cada acoplamiento, las demás predicciones han sido incluidas en los anexos para no cargar con excesivas imágenes el trabajo.



**Figura 14.** Visualización con PyMOL de la mejor configuración (Conf. 9415) del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).

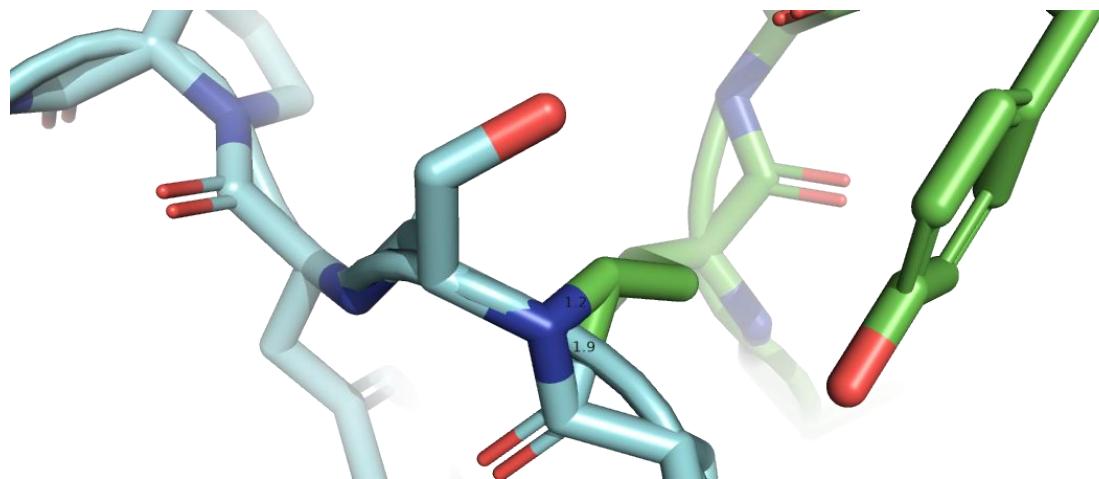
La mejor puntuación en función de la energía del enlace, obtenida con el servidor *pyDockWEB* (*Figura 14*), corresponde con la configuración **9415**. La zona de interacción entre las dos proteínas se sitúa al inicio de la secuencia de la proteína A1BG, entre los aminoácidos 6 y 23. En esta zona se forman dos enlaces entre las cadenas de las proteínas.

La fenilalanina de la posición 7 de la proteína A1BG genera un enlace con una treonina (posición 29) de la proteína CRISP-3 (*Figura 15*).



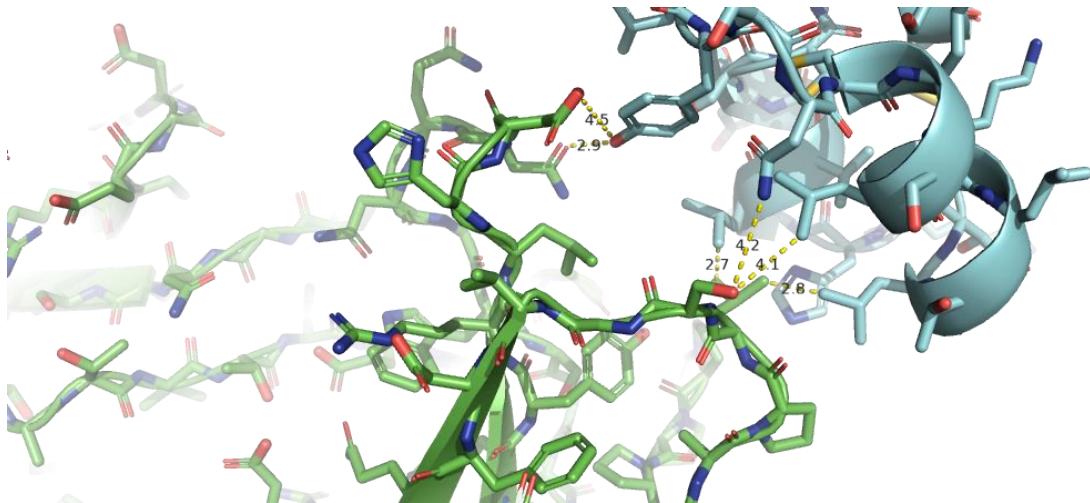
**Figura 15.** Visualización con PyMOL del enlace correspondiente a la primera zona de contacto entre el Modelo 1 y la proteína CRISP-3. Unión entre la fenilalanina (posición 7) de la proteína receptora y una treonina (posición 29) del ligando. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.

El segundo enlace se forma entre una isoleucina (posición 23) de la proteína A1BG y los aminoácidos de las posiciones 16 y 17 (prolina y serina) de la proteína CRISP-3 (*Figura 16*).



**Figura 16.** Visualización con PyMOL del enlace correspondiente a la primera zona de contacto entre el Modelo 1 y la proteína CRISP-3. Unión entre la isoleucina (posición 23) de la proteína receptora, y los aminoácidos 16 y 17 (prolina y serina) del ligando. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.

Existe una segunda zona de interacción entre las estructuras, situada en el tercer dominio de la proteína A1BG. Esta zona está situada entre los aminoácidos 206 y 286 de la proteína A1BG (*Figura 17*). En este caso las moléculas se aproximan, pero no se genera ningún enlace.



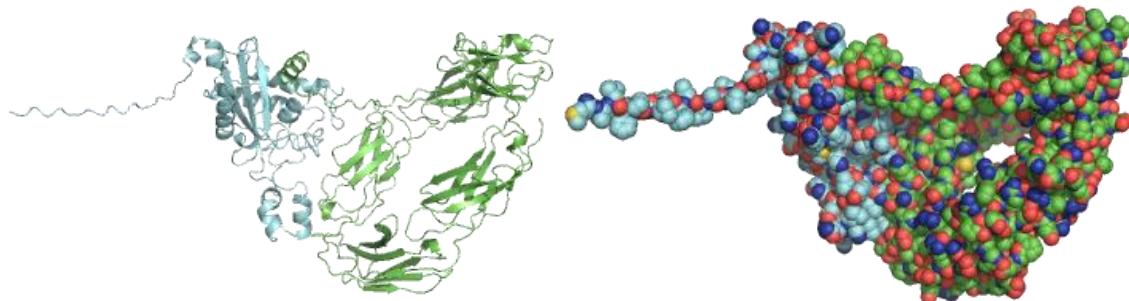
**Figura 17.** Visualización con PyMOL de la zona de interacción entre el tercer dominio del Modelo 1 (verde) y la proteína CRISP-3 (azul).

Los resultados de la simulación del acoplamiento entre el Modelo 2 de la estructura tridimensional de la proteína A1BG y la proteína secretora CRISP-3 (*Tabla 12*) también son positivos. Las 7 configuraciones mejor puntuadas, muestran una fuerte unión de la proteína CRISP-3 con la secuencia inicial de la proteína A1BG, que en este caso pertenece a una hélice transmembrana. La predicción con la mejor puntuación (configuración **595**) contiene un alto nivel de energía en el enlace. Aunque la puntuación de la energía total del enlace (-48.343) es inferior a los resultados del acoplamiento con el Modelo 1 (-57.153), se observa un incremento de las fuerzas de Van der Waals (135.109).

Conf	Electrostatics	Desolvation	VdW	Total	Rank
595	-15.499	-46.355	135.109	-48.343	1
6007	-10.741	-41.011	43.534	-47.398	2
8249	0.288	-53.039	75.312	-45.220	3
3163	-14.607	-42.214	131.263	-43.695	4
2449	-12.499	-32.853	35.803	-41.772	5
7748	-9.879	-36.224	44.516	-41.652	6
6404	-8.182	-36.963	50.466	-40.098	7
3853	-8.010	-38.048	70.259	-39.032	8
6384	-10.362	-34.703	56.473	-39.418	9
787	-11.718	-32.083	65.863	-37.215	10

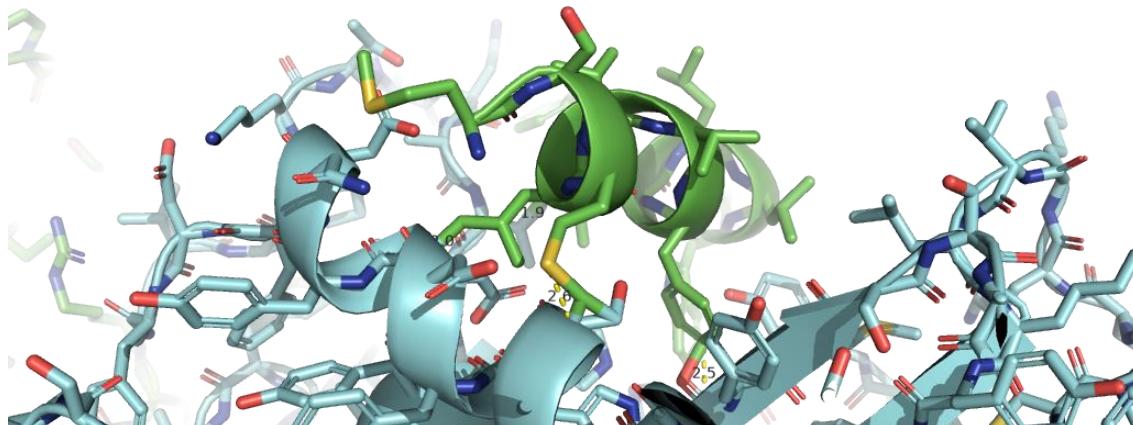
**Tabla 12.** Resultado de las 10 mejores predicciones generadas por PyDockWEB del acoplamiento entre el Modelo 2 de la proteína A1BG y la proteína secretora CRISP-3.

La visualización de la configuración 595 con *PyMOL* (*Figura 18*) muestra una amplia zona de contacto entre las dos moléculas. Se observan dos zonas de contacto entre las proteínas.



**Figura 18.** Visualización con PyMOL de la mejor configuración (Conf. 595) del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).

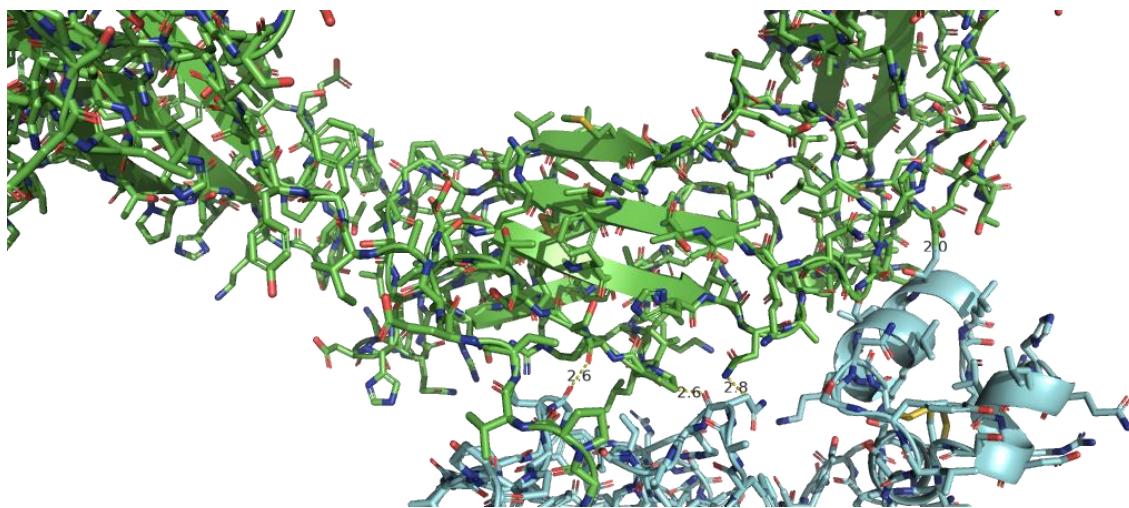
La primera zona se encuentra situada en la hélice transmembrana de la proteína A1BG (1-11), esta zona contiene enlaces entre algunos aminoácidos de las dos cadenas (*Figura 19*). Se observan 4 enlaces que corresponden con los aminoácidos 3, 4, 7 y 8 de la proteína A1BG.



**Figura 19.** Visualización con PyMOL de los enlaces correspondientes a la primera zona de contacto entre las dos moléculas. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.

La segunda zona está situada entre los dominios 3 y 4 de la proteína A1BG. En este punto, se observa la unión entre un ácido aspártico (posición 303) de la proteína receptora y una tirosina (posición 222) del ligando (*Figura 20*).

La interacción de la proteína CRISP-3 con los dominios 3 y 4 de la proteína A1BG tiene una amplia zona de contacto. Las distancias entre estos dominios son muy estrechas, aunque solo se observa un enlace.



**Figura 20.** Visualización con PyMOL del enlace correspondiente a la segunda zona de contacto entre las dos moléculas. En la parte derecha de la imagen, el enlace entre el ácido aspártico (posición 303) de la proteína receptora y la tirosina (posición 222) del ligando. En el centro, se muestran las distancias entre los dominios 3 y 4 de A1BG y CRISP-3. En verde, la proteína receptora A1BG, y en azul el ligando CRISP-3.

## 5. Conclusiones

El análisis de la estructura secundaria predice una región transmembrana al inicio de la secuencia de la proteína A1BG. El resto de la secuencia pertenece a una región citoplasmática, compuesta por 5 dominios. Los resultados indican que se podría tratar de una proteína asociada a la membrana.

El modelado comparativo de la proteína A1BG a partir de los moldes de las estructuras obtenidas en el *Protein Data Bank* (Modelo 1) no tiene suficiente identidad entre las secuencias. La baja identidad entre las secuencias afecta de manera negativa a la calidad de los modelos obtenidos. Aunque se han seleccionado los moldes en función de la similitud de los pliegues con *GenTHREADER*, la baja identidad entre las secuencias y la falta de estructuras homólogas conocidas genera unos modelos de muy mala calidad.

En cambio, se ha obtenido una estructura de buena calidad a partir del molde de *AlphaFold* (Modelo 2). Las secuencias tienen una alta identidad, además del 100% de cobertura, de esta manera es más fácil obtener una mejor calidad en las estructuras modeladas. Los resultados concluyen que el Modelo 2 tiene suficiente calidad en la estructura, por lo tanto, podría contribuir al estudio de las funciones o actividades catalíticas asociadas a la proteína A1BG.

El Modelo 2 tiene una estructura secundaria similar a la predicción generada por *PSIPred*. La hélice  $\alpha$  inicial coincide con la predicción de la región transmembrana, además de las hélices de los dominios 4 y 5. La forma en que se dispone la hélice  $\alpha$  inicial respecto al resto de la molécula presenta una cierta similitud con la estructura secundaria de algunas proteínas asociadas a la membrana.

El acoplamiento entre el modelo generado de la proteína A1BG y la proteína CRISP-3 respalda los resultados obtenidos hasta el momento. Las fuerzas de Van der Waals se ven aumentadas en el acoplamiento con el Modelo 2, tal como se ha observado a nivel experimental en el complejo formado. Además, 9 de los 10 acoplamientos con mejores puntuaciones tiene su punto de anclaje en la misma zona o próxima a la descrita en los resultados. Por esto, la predicción del acoplamiento entre el Modelo 2 y la proteína CRISP-3 respalda la hipótesis de que la estructura obtenida en el proceso de modelado presenta una cierta similitud con la proteína A1BG.

El conocimiento de la estructura de esta proteína puede implicar grandes avances en la investigación. Se ha demostrado la presencia de este complejo en el plasma de algunos mamíferos<sup>(24)</sup>, aunque se desconoce la función, la proteína CRISP-3 tiene algunas funciones asociadas. La proteína CRISP-3 es sintetizada en el interior de la célula y viaja a través de la membrana hacia el espacio extracelular. La unión entre las dos moléculas implica la participación de la hélice transmembrana de la proteína A1BG, esto podría indicar una posible función de transporte asociada a la proteína. En el caso que se tratara de una proteína asociada a la membrana, la unión con fuertes fuerzas electrostáticas con la proteína CRISP-3 provocaría la disociación de la proteína de membrana, esto podría explicar la presencia del complejo en el plasma.

El modelo de la estructura tridimensional de la proteína A1BG obtenido con la realización de este TFM puede tener distintas aplicaciones. La calidad del modelo obtenido indica que el estudio de las posibles interacciones con otras proteínas, o sus actividades catalíticas, podrían contribuir al conocimiento de sus funciones biológicas.

## **6. Glosario**

A1BG – Glicoproteína alfa 1-B

CRISP-3 – Proteína secretora rica en cisteína 3

IGKC – Región constante de la cadena kappa de la inmunoglobulina

HP – Haptoglobina

UCSC – del inglés “*University of California Santa Cruz*”

ODS – Objetivos de Desarrollo Sostenible

ONU – Organización de las naciones unidas

TFM – Trabajo de Final de Máster

PDB – del inglés “*Protein Data Bank*”

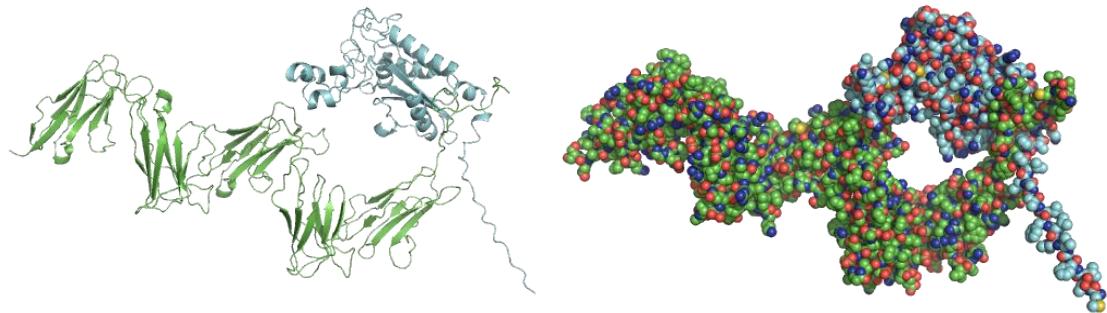
## 7. Bibliografía

1. Contreras-Moreira, Bruno. *Modelado comparativo de proteínas*. Digital CSIC, 2012 oct.
2. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. *The Protein Data Bank*. Nucleic Acids Research (2000), 28: 235-242.
3. Jumper, J., Evans, R., Pritzel, A. et al. *Highly accurate protein structure prediction with AlphaFold*. Nature 596, 583–589 (2021).
4. Minkyung Baek et al. *Accurate prediction of protein structures and interactions using a three-track neural network*. Science (2021), 373: 871-876.
5. H.E. Schultze, K. Heide, H. Haupt. *Isolation of an easily precipitable alpha1-glycoprotein of human serum*. Nature, 200 (1963), p. 1103
6. Ishioka N, Takahashi N, Putnam FW. *Amino acid sequence of human plasma alpha 1B-glycoprotein: homology to the immunoglobulin supergene family*. Proc Natl Acad Sci U S A. 1986 Apr;83(8):2363-7.
7. Udby L, Sørensen OE, Pass J, Johnsen AH, Behrendt N, Borregaard N, Kjeldsen L. *Cysteine-rich secretory protein 3 is a ligand of alpha1B-glycoprotein in human plasma*. Biochemistry. 2004 Oct 12;43(40):12877-86.
8. Dutta M, Subramani E, Taunk K, Gajbhiye A, Seal S, Pendharkar N, Dhali S, Ray CD, Lodh I, Chakravarty B, Dasgupta S, Rapole S, Chaudhury K. *Investigation of serum proteome alterations in human endometriosis*. J Proteomics. 2015 Jan 30;114:182-96.
9. Cordeiro YG, Mulder LM, van Zeijl RJM, Paskoski LB, van Veelen P, de Ru A, Strefezzi RF, Heijs B, Fukumasu H. *Proteomic Analysis Identifies FNDC1, A1BG, and Antigen Processing Proteins Associated with Tumor Heterogeneity and Malignancy in a Canine Model of Breast Cancer*. Cancers. 2021; 13(23):5901.
10. Liu Y, Luo X, Hu H, Wang R, Sun Y, Zeng R, Chen H. *Integrative proteomics and tissue microarray profiling indicate the association between overexpressed serum proteins and non-small cell lung cancer*. PLoS One. 2012;7(12):e51748.
11. Blum M, Chang H, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD. *The InterPro protein families and domains database: 20 years on*. Nucleic Acids Research, Nov 2020.

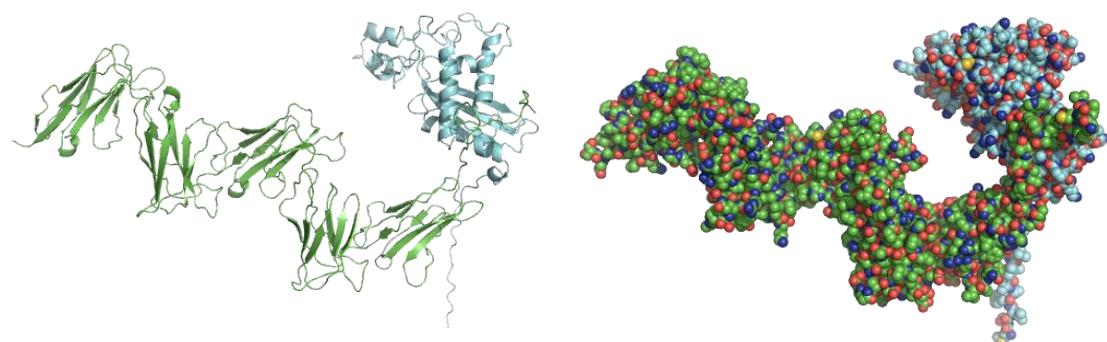
12. Söding J, Biegert A, Lupas AN. *The HHpred interactive server for protein homology detection and structure prediction*. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W244-8.
13. Webb, A. Sali. *Comparative Protein Structure Modeling Using Modeller*. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.
14. PyMOL. *El sistema de gráficos moleculares PyMOL*, versión 2.0 Schrödinger, LLC.
15. Williams et al. (2018). *MolProbity: More and better reference data for improved all-atom structure validation*. Protein Science 27: 293-315.
16. Jianyi Yang and Yang Zhang. *I-TASSER server: new development for protein structure and function predictions*. Nucleic Acids Research, 43: W174-W181 (2015).
17. Brian Jiménez-García, Carlos Pons and Juan Fernández-Recio. "pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring". Bioinformatics (2013). 29(13):1698-1699.
18. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. *The human genome browser at UCSC*. Genome Res. 2002 Jun;12(6):996-1006.
19. Buchan DWA, Jones DT. *The PSIPRED Protein Analysis Workbench: 20 years on*. Nucleic Acids Research (2019).
20. Jones DT. *Protein secondary structure prediction based on position-specific scoring matrices*. J. Mol. Biol. (1999). 292: 195-202.
21. Jones DT, Cozzetto D. *DISOPRED3: precise disordered region predictions with annotated protein-binding activity*. Bioinformatics. 2015 Mar 15;31(6):857-63.
22. McGuffin LJ & Jones DT. *Improvement of the GenTHREADER method for genomic fold recognition*. Bioinformatics (2003), 19, 874-881.
23. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. *Jalview Version 2 - a multiple sequence alignment editor and analysis workbench*. Bioinformatics (2009). 25 (9) 1189-1191.
24. Lene Uddy, Anders H. Johnsen, Niels Borregaard. *Human CRISP-3 binds serum α1B-glycoprotein across species*. Biochimica et Biophysica Acta (BBA) - General Subjects, Volume 1800, Issue 4, 2010, 481-485.

## 8. Anexos

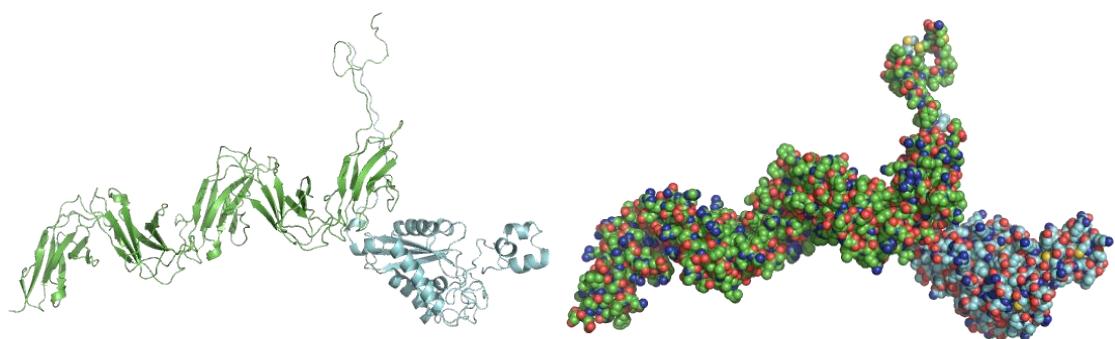
Para no sobrecargar con demasiadas imágenes el trabajo, se ha incluido la visualización de las 9 predicciones restantes del acoplamiento entre la proteína A1BG y CRISP-3 generado con *pyDockWEB*.



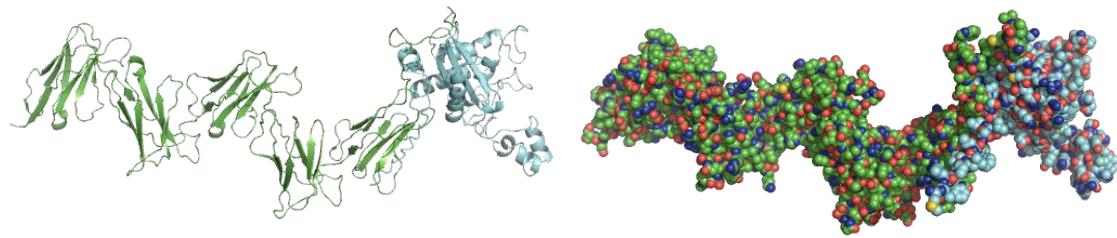
**Figura 21.** Visualización con PyMOL de la configuración 7310 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



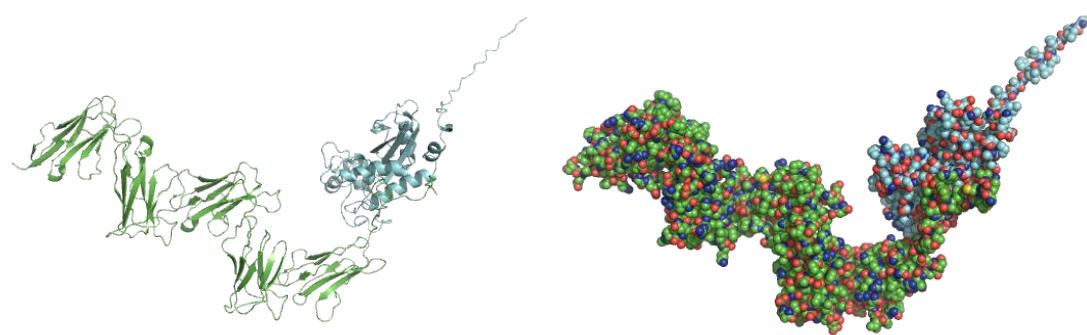
**Figura 22.** Visualización con PyMOL de la configuración 59 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



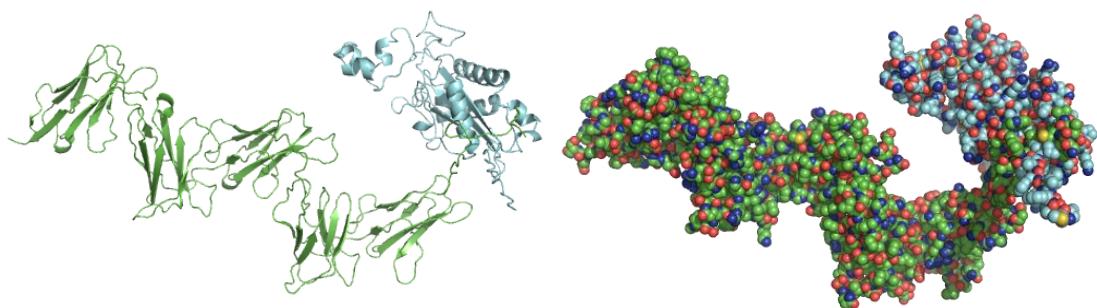
**Figura 23.** Visualización con PyMOL de la configuración 4776 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



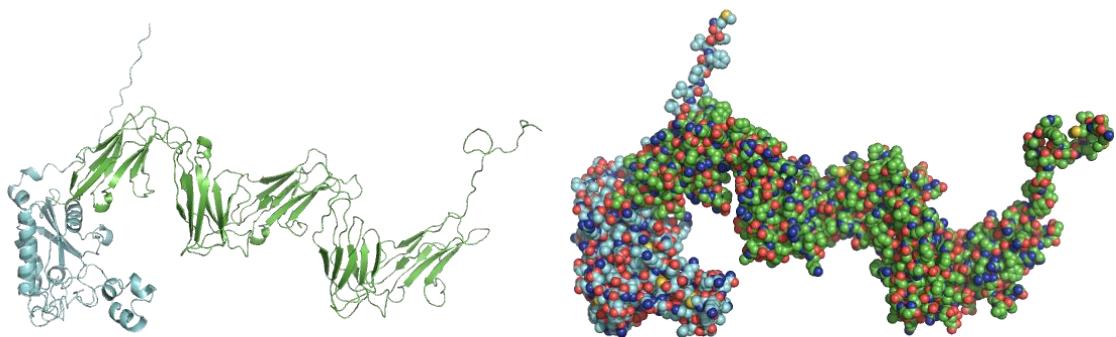
**Figura 24.** Visualización con PyMOL de la configuración 2992 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



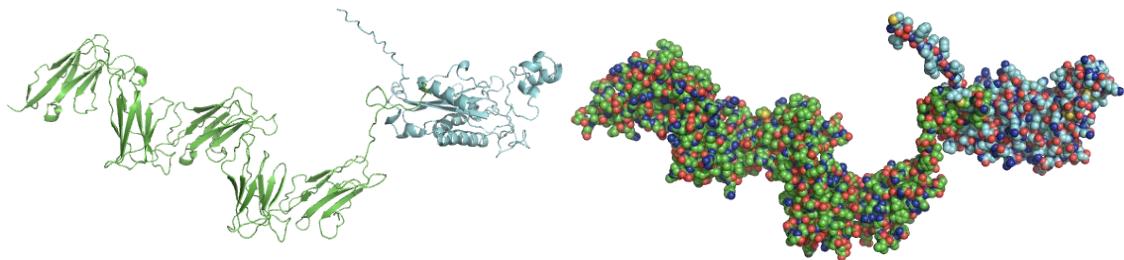
**Figura 25.** Visualización con PyMOL de la configuración 8865 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



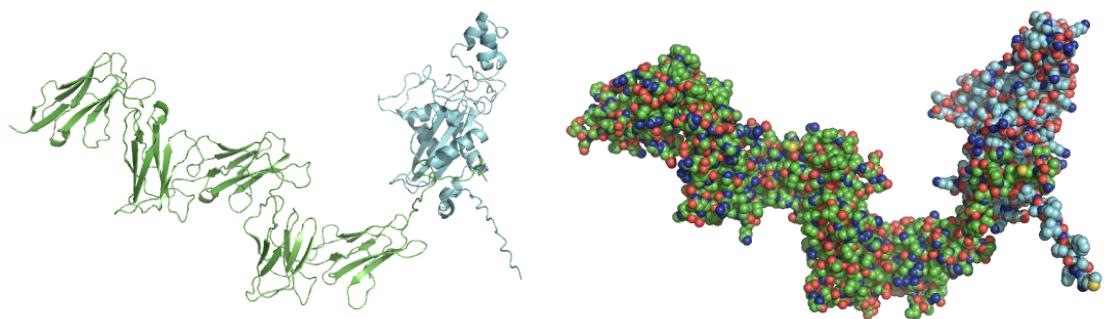
**Figura 26.** Visualización con PyMOL de la configuración 7297 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



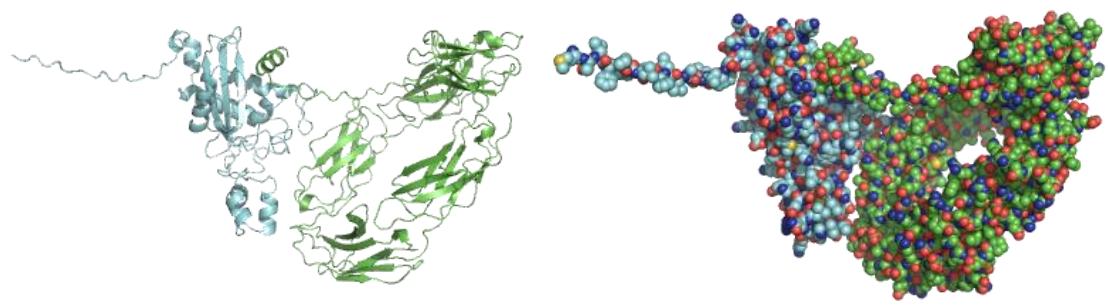
**Figura 27.** Visualización con PyMOL de la configuración 4612 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



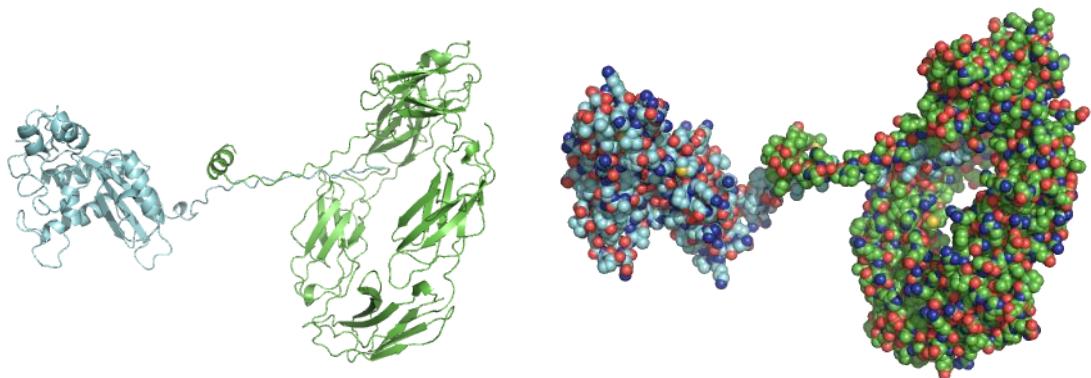
**Figura 28.** Visualización con PyMOL de la configuración 4772 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



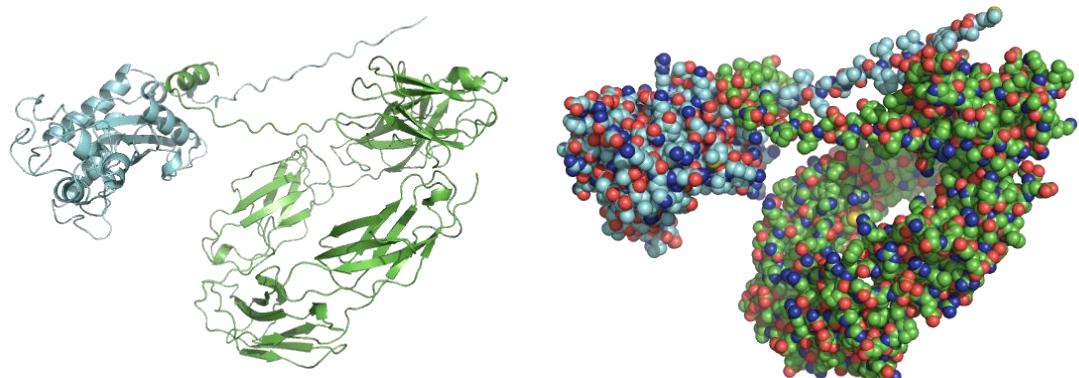
**Figura 29.** Visualización con PyMOL de la configuración 9414 del enlace entre el Modelo 1 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



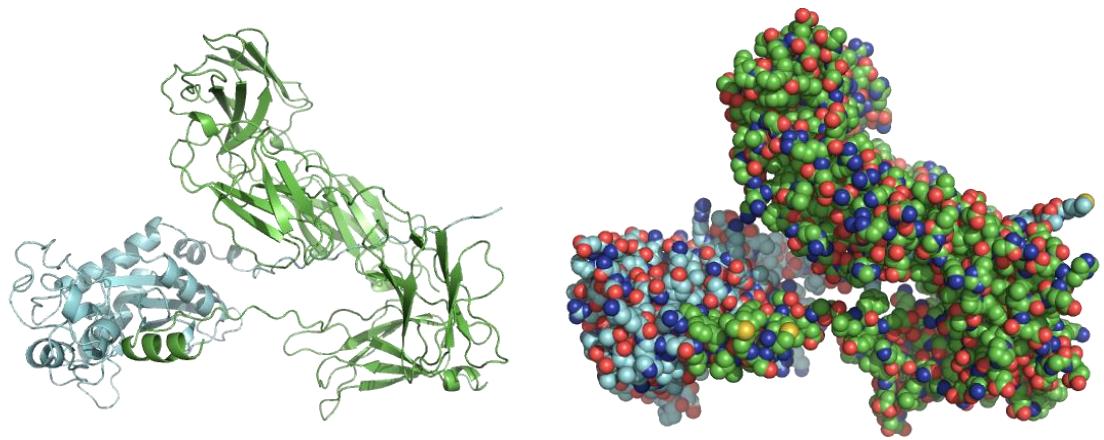
**Figura 30.** Visualización con PyMOL de la configuración 6007 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



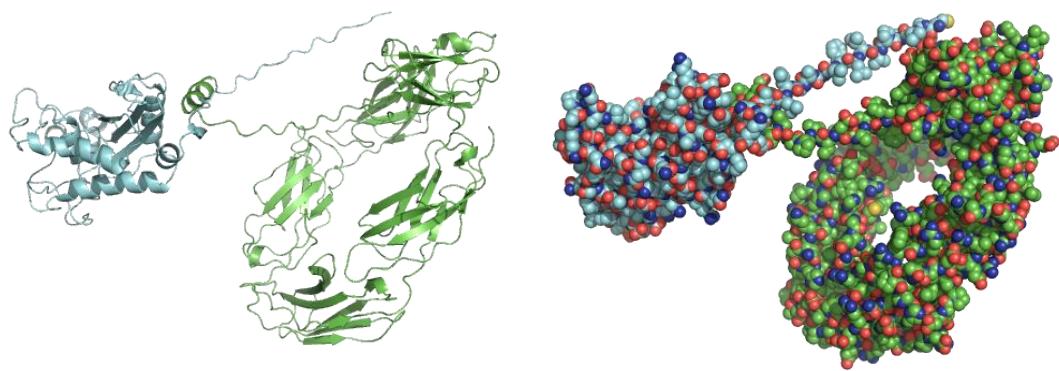
**Figura 31.** Visualización con PyMOL de la configuración 8249 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



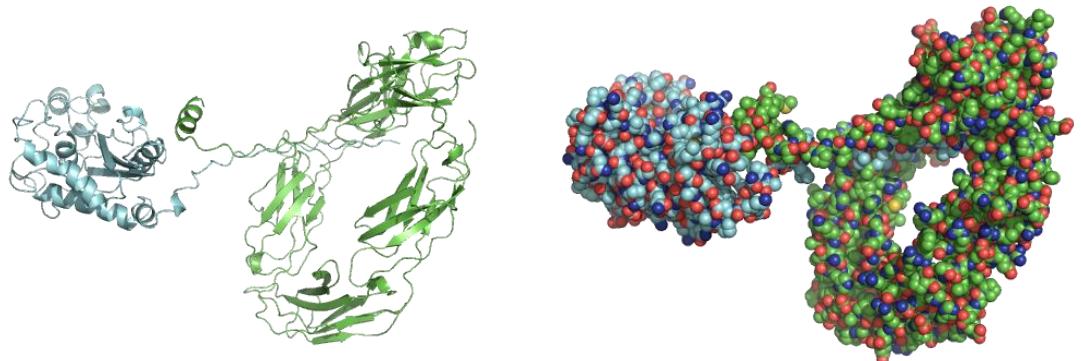
**Figura 32.** Visualización con PyMOL de la configuración 3163 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



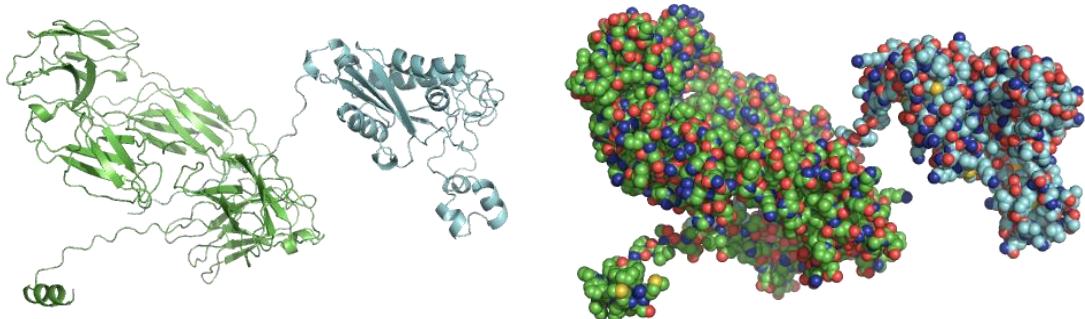
**Figura 33.** Visualización con PyMOL de la configuración 2449 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



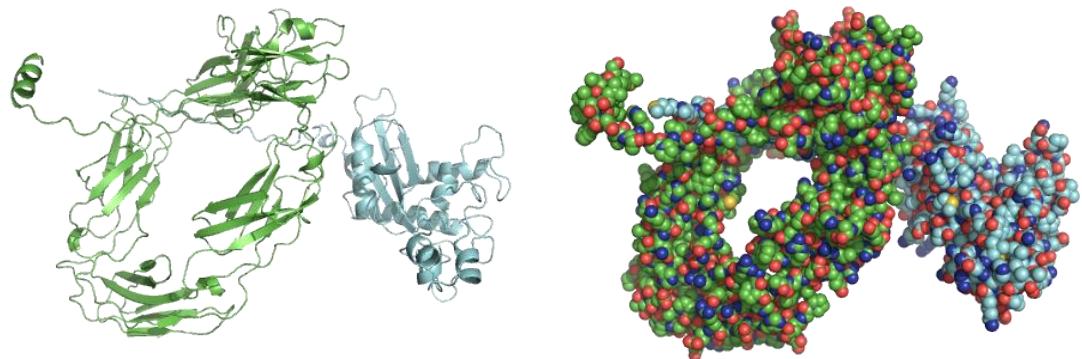
**Figura 34.** Visualización con PyMOL de la configuración 7748 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



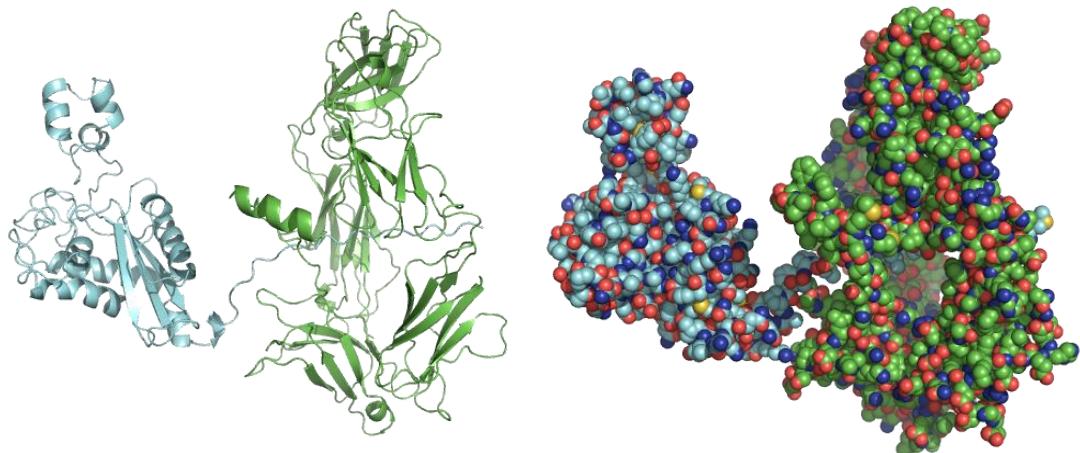
**Figura 35.** Visualización con PyMOL de la configuración 6404 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



**Figura 36.** Visualización con PyMOL de la configuración 3853 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



**Figura 37.** Visualización con PyMOL de la configuración 6384 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).



**Figura 38.** Visualización con PyMOL de la configuración 787 del enlace entre el Modelo 2 de la proteína A1BG (verde) y la proteína secretora CRISP-3 (azul).