

ArgosMol: Una plataforma web para el modelado de estructuras terciarias de proteínas

ALUMNO	PROGRAMA	CURSO
MSc. Vicente Enrique Machaca Arceda	Doctorado en Ciencias de la Computación	Tópicos en Computación Gráfica

1. Introducción

Las proteínas son moléculas complejas que cumplen un rol crítico en nuestro cuerpo, estas cumplen la mayoría de funciones en la células (Anderson and Anderson, 1998). Además, la función de una proteína depende de su estructura (Rangwala and Karypis, 2010) y ultimamente se ha descubierto que esta función también depende de la relación de una proteína con otras (Canzar and Ringeling, 2020). Mas aún, es importante saber, que la estructura de una proteína puede cambiar en el tiempo y su función también cambia en el tiempo.

Conocer la estructura de una proteína es de suma importancia para el análisis de su función, generación de medicamentos, etc. (Rangwala and Karypis, 2010). Además, lograr predecir y entender el funcionamiento de estas proteínas y la interacción de redes de proteínas es considerado el nuevo santo grial de la Bioinformática en estos tiempos (Srihari et al., 2017).

2. Marco teórico

En esta sección detallaremos algunos conceptos previos del área de Bioinformática/*Proteomics* para comprender el trabajo.

2.1. Estructura de las proteínas

Existen 4 tipos de estructuras de proteínas (Russell and Gordey, 2002):

1. **Estructura primaria:** Secuencia de aminoácidos (ver Figura 1 (a)).
2. **Estructura secundaria:** Pequeños patrones, los más comunes son las hélices α y hojas β (ver Figura 1 (b)).
3. **Estructura terciaria:** Representa la unión de los segmentos de la estructura secundaria (ver Figura 1 (c)). En este caso solo estamos considerando una cadena de aminoácidos (las proteínas a veces son conformadas por varias cadenas de aminoácidos).
4. **Estructura cuaternaria:** Unión de varias estructuras terciarias (varias cadenas de aminoácidos) (ver Figura 1 (d)).

2.2. Contact map

Representa la distancia de cada posible aminoácido, cuando forman proteínas. El *contact map*, es representado como un gráfico en 2D, y es el elegido por los modelos de machine learning en la predicción de las estructuras de proteínas. En la Figura 2, mostramos como es un *contact map*.

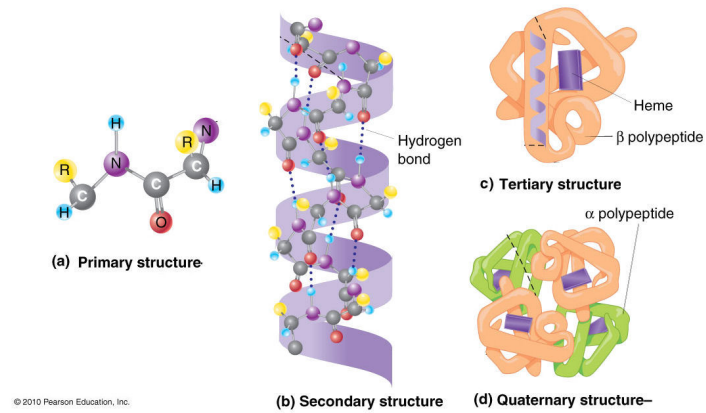


Figura 1: Ejemplo de las 4 estructuras de proteínas. Fuente: (Russell and Gordey, 2002)

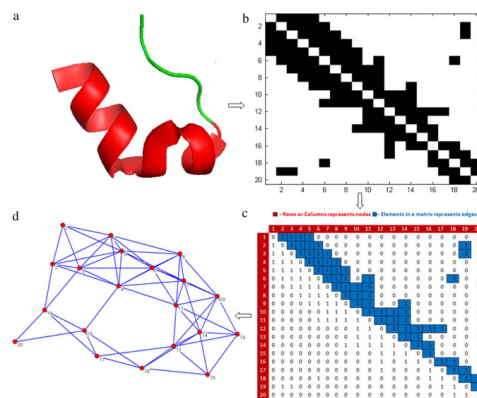


Figura 2: Ejemplo del contact map de una proteína.

3. Trabajos relacionados

Las herramientas mas utilizadas para la visualización de proteínas son herramientas de escritorio, entre estas tenemos: PyMol, VMD, Jmol y Chimera. Una de las principales desventajas de estas aplicaciones es que son faciles de utilizar y requieren ser instaladas. Debido a esta problemática, en los ultimos años se ha empezado el desarrollo de herramientas Web que permiten visualizar proteínas con el menor esfuerzo posible.

Una de las primeras propuestas es Jolecule, esta herramienta es desarrollada por Bosco Ho desde el 2011, aunque no tiene una publicación, es una herramienta con constante desarrollo hasta la actualidad. Dicha aplicación esta desarrollada sobre three.js y es de código libre. Jolecule, tiene algunos problemas de renderización, por momentos ciertas partes de la proteína no se visualizan, tambien, en su página no se permite cargar una archivo .pdb para visualizarlo al menos que levatemos la aplicación en nuestro servidor.

La herramienta NGL propuesta por Rose and Hildebrand (2015), tambien está desarrollada sobre three.js y es de código libre. Aunque, en ciertas ocaciones la aplicación presenta errores, es una aplicación facil de utilizar.

EzMol propuesta por Reynolds et al. (2018) es una aplicación desarrollada sobre 3dmol.js. Según los autores, el objetivo de EzMol, es ser la una herramietna de facil uso. EzMol, tiene un *wizard* que guía al usuario durante el proceso de visualización, este *wizard* es util para usuarios inexpertos pero le resta puntos de usabilidad a la aplicación.

Finalmente tenemos iCn3D, propuesta por Wang et al. (2020). Esta herramienta tambien ha sido desarrollada sobre three.js. Esta aplicación, tiene mas funcionalidades que las otras herramientas mencionadas anteriormente y tambien es de facil uso. Un problema con esta herramienta es que no permite ver la secuencia de aminoacidos en forma paralela al modelo 3D, tambien tiene algunos problemas con la posición de sus menus y botones.

4. Propuesta

ArgosMol surge como una alternativa a herramientas *desktop* como: PyMol, VMD, Jmol y Chimera, considerando que en estos ultimos años, se ha priorizado el desarrollo de herramientas Web que hacen uso de WebGL ¹, Three.js ² y 3DMol.js. ³; Además, ArgosMol ha tomado en cuenta los principios de desarrollo de toda aplicación de visualización de moléculas, presentado por Youkharibache (2017)

Para el desarrollo de ArgosMol se ha utilizado 3dmol.js y p5.js en el *frontend* y PHP en el *backend*. La herramienta es presentada en la Figura 3 y esta disponible en este enlace.

¹Librería para el desarrollo de modelos 3D. Enlace.

²Librería de alto nivel para el desarrollo de modelos 3D, hace uso de WebGL. Enlace.

³Librería para el modelado de átomos y moléculas, hace uso de WebGL. Enlace.

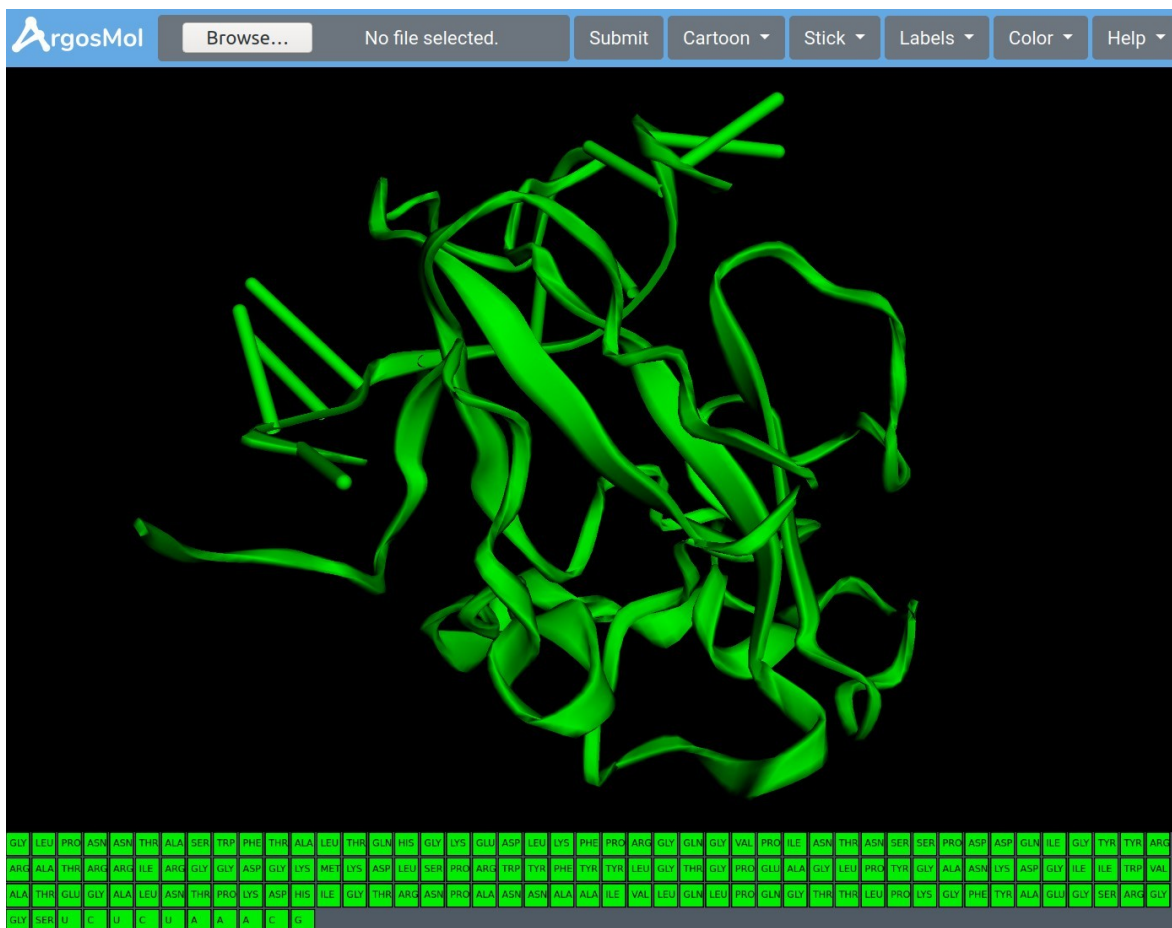


Figura 3: Ejemplo de visualización de proteínas con ArgosMol.

4.1. Estilo

ArgosMol permite ver la estructura de una proteína en tres modos distintos: *Cartoon*, *Stick* y *Trace*, y la combinación de estos. Por ejemplo para la proteína con el identificador: *7act* en el *Protein Data Bank*. En la Figura 4, mostramos las 3 formas de visualización de ArgosMol, la Figura 4(a)(b)(c)(d) representan la proteína en un estilo *cartoon*, *trace*, *stick*, *cartoon/stick* respectivamente.

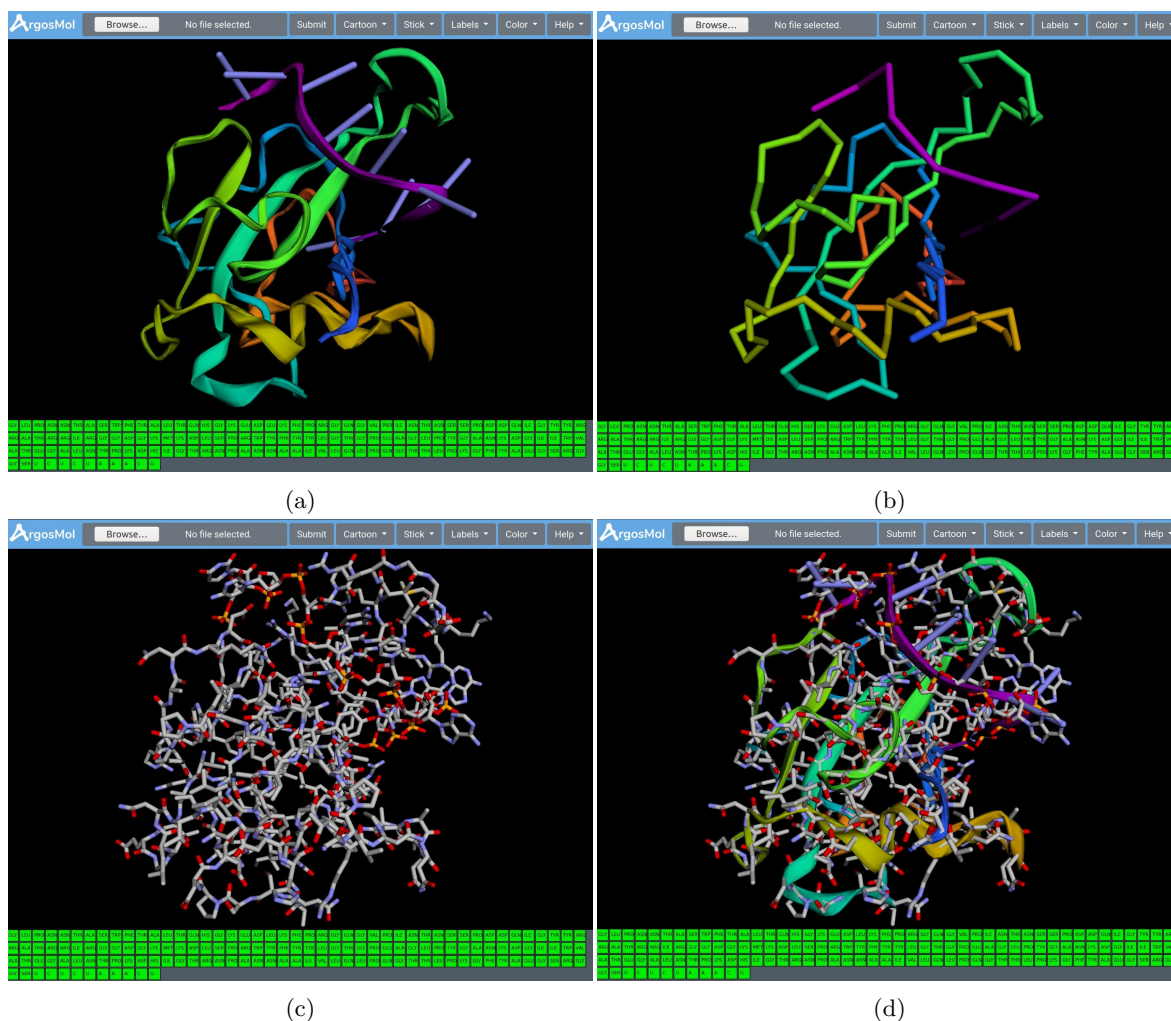


Figura 4: Formas de visualizar una proteína en ArgosMol. (a) *cartoon*, (b) *trace*, (c) *stick* y (d) *cartoon/stick*.

4.2. Esquema de colores

ArgolMol tiene siete esquemas de colores, la vista por defecto es el color verde y se presenta en la Figura 3. Los demas esquems de colors son descritos en la Tabla 1

Esquema	Descripción
<i>Spectrum</i>	Cada grupo de minoacidos es pintado aplicando una degradación de colores(Fig. 5.a).
<i>By chain</i>	Se asigna un color a cada cadena polipeptídica (Fig. 5.b).
<i>By structure</i>	Se asigna un color según el tipo de estructura secundaria (Fig. 5.c).
<i>Rasmol color</i>	Se utiliza un color según la polaridad de cada aminoacido (Fig. 5.d).
<i>Rasmol shape</i>	Se utiliza un color según la forma de cada aminoacido (Fig. 5.e).
<i>MAE color</i>	Se utiliza el mismo esquema de colores que la herramienta MAE (Fig. 5.f).

Tabla 1: Esquema de colores propuestos por ArgosMol.

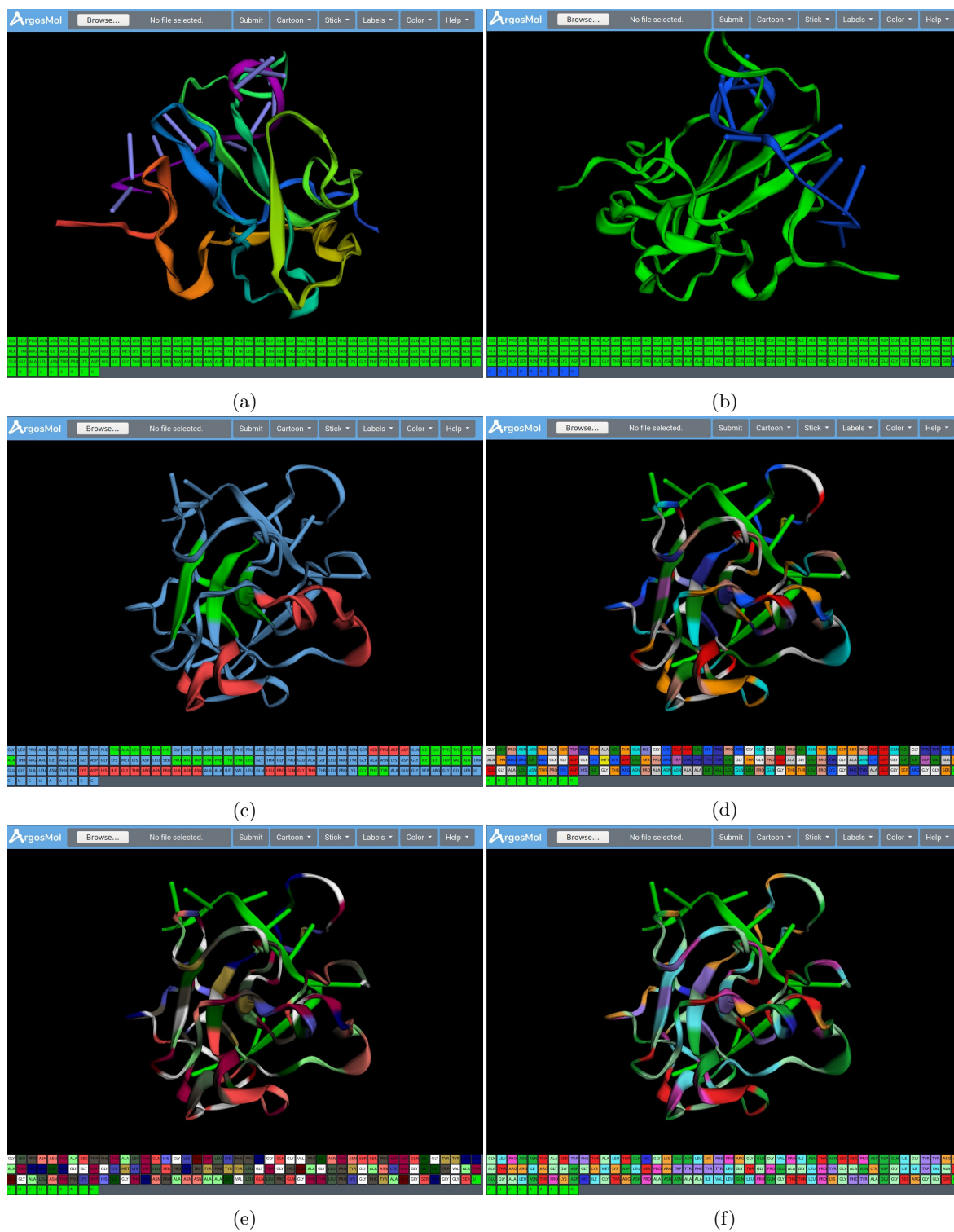


Figura 5: Esquema de colores utilizados por ArgosMol. (a) *spectrum*, (b) *by chain*, (c) *by structure* y (d) *rasmol color*, (e) *rasmol shape* y (f) *MAE color*.

4.3. Secuencia de Aminoácidos

Youkharibache (2017) recomienda que toda herramienta de visualización debe incluir una vista a la secuencia de aminoácidos y estas deben estar enlazadas, de manera tal que si selecciono un aminoácido, este debe estar resaltado en el modelo. ArgosMol, ha tomado en cuenta este requisito, por ejemplo en la Figura 6, mostramos como al hacer *click* izquierdo en algunos aminoácidos, estos son pintados de color verde tanto en el modelo; esta funcionalidad también puede ser utilizada de forma inversa, es decir: si hago *click* en el modelo, los aminoácidos en la secuencia serán resaltados.

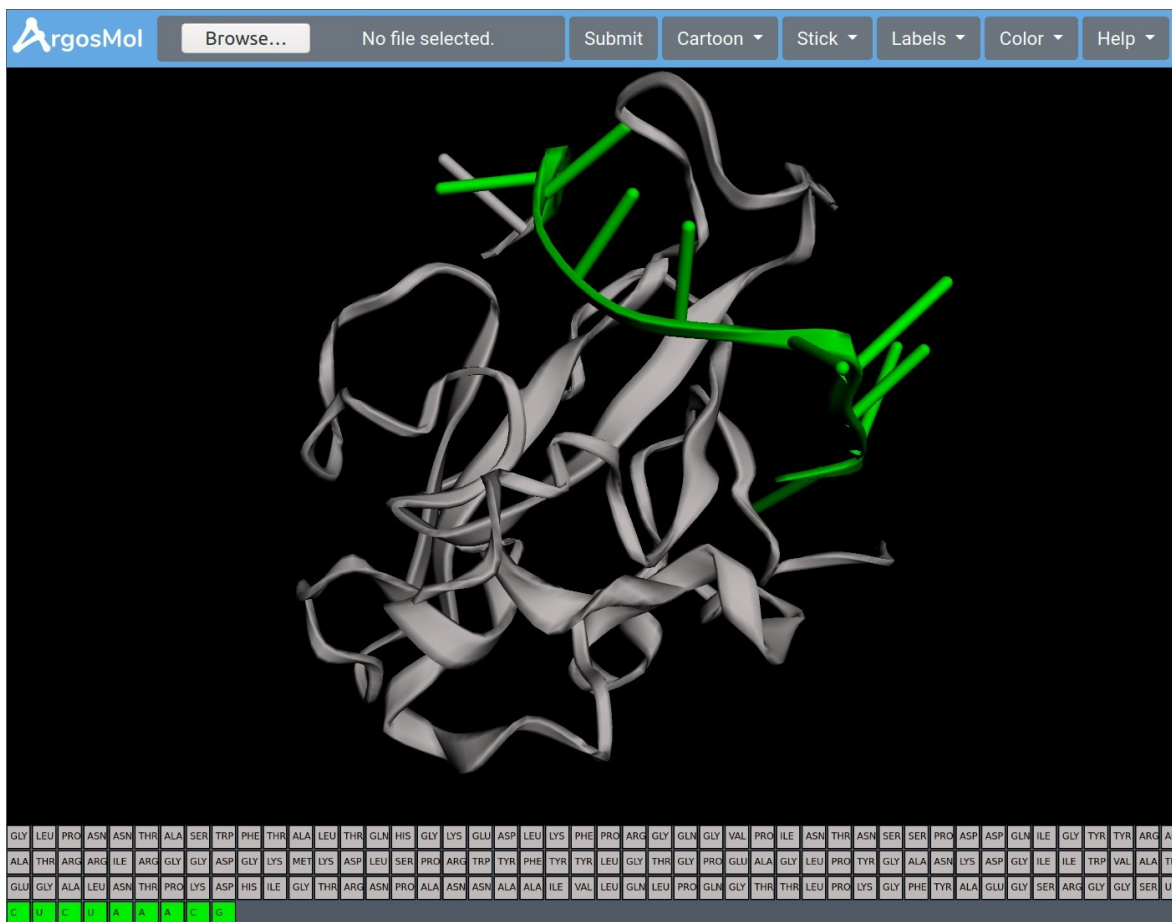


Figura 6: Ejemplo del enlace entre el modelo y la secuencia de aminoácidos mediante eventos del *mouse*.

4.3.1. Eventos por teclado

También se ha agregado eventos por teclado para facilitar el uso de la herramienta, en la Tabla de tallamos estos eventos.

Teclado	Descripción
c	Activamos la opción para ver el <i>cartoon</i> .
C	Eliminamos el <i>cartoon</i> del modelo.
t	Activamos la opción para ver en estilo <i>trace</i> .
s	Activamos la opción para ver el <i>stick</i> .
S	Eliminamos el <i>stick</i> del modelo.
l	Agregamos todas las etiquetas de los aminoácidos al modelo.
L	Eliminamos todas las etiquetas de los aminoácidos al modelo.
r	Reiniciamos el modelo a su estado inicial.
Ctrl + z	Regresamos a un estado anterior.

Tabla 2: Atajos de teclado utilizados en ArgosMol.

5. Trabajos futuros

ArgosMol esta en su versión 1.0 permite ver la estructura terciaria de las proteínas, para versiones futuras se tiene planificado agregar la funcionalidad de ver las proteínas a nivel de moléculas. También se tiene como objetivo a futuro, incorporar algoritmos de predicción de estructuras de proteínas, de manera tal, que tomando como entrada una secuencia de aminoácidos o un *contact map*, ArgosMol genere la posición de cada átomo en un archivo .pdb.

6. Conclusiones

En este proyecto se ha estudiado el área de predicción de estructuras de proteínas y su visualización. Al ser un trabajo complejo, se decidió enfocarnos en la visualización de proteínas.

Se propuso la herramienta Web ArgosMol de fácil uso, la cual es una alternativa a herramientas de escritorio que requieren instalación y tienen una curva de aprendizaje alto.

Entre las funcionalidades de ArgosMol se propuso 3 formas de visualización (*cartoon*, *trace*, *stick*), varios esquemas de colores, atajos de teclado y un enlace permanente entre la secuencia de aminoácidos y el modelo 3D.

Referencias

- Adhikari, B. and Cheng, J. (2018). Confold2: improved contact-driven ab initio protein structure modeling. *BMC bioinformatics*, 19(1):22.
- Anderson, N. L. and Anderson, N. G. (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, 19(11):1853–1861.
- Canzar, S. and Ringeling, F. R. (2020). Protein-protein interaction networks.
- Rangwala, H. and Karypis, G. (2010). Introduction to protein structure prediction. *Introduction to Protein Structure Prediction*, 58.
- Reynolds, C. R., Islam, S. A., and Sternberg, M. J. (2018). Ezmol: a web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *Journal of molecular biology*, 430(15):2244–2248.
- Rose, A. S. and Hildebrand, P. W. (2015). Ngl viewer: a web application for molecular visualization. *Nucleic acids research*, 43(W1):W576–W579.

- Russell, P. J. and Gordey, K. (2002). *IGenetics*. Number QH430 R87. Benjamin Cummings San Francisco.
- Srihari, S., Yong, C. H., and Wong, L. (2017). *Computational prediction of protein complexes from protein interaction networks*. Morgan & Claypool.
- Wang, J., Youkharibache, P., Zhang, D., Lanczycki, C. J., Geer, R. C., Madej, T., Phan, L., Ward, M., Lu, S., Marchler, G. H., et al. (2020). icn3d, a web-based 3d viewer for sharing 1d/2d/3d representations of biomolecular structures. *Bioinformatics*, 36(1):131–135.
- Youkharibache, P. (2017). Twelve elements of visualization and analysis for tertiary and quaternary structure of biological molecules. *BioRxiv*, page 153528.