

Avance N° 03

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

AVANCE	TEMA	FECHA
03	Predicción y modelado en 3D de estructuras terciarias de proteínas a partir del <i>contact map</i>	20-02-2021

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. Conceptos previos

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)).

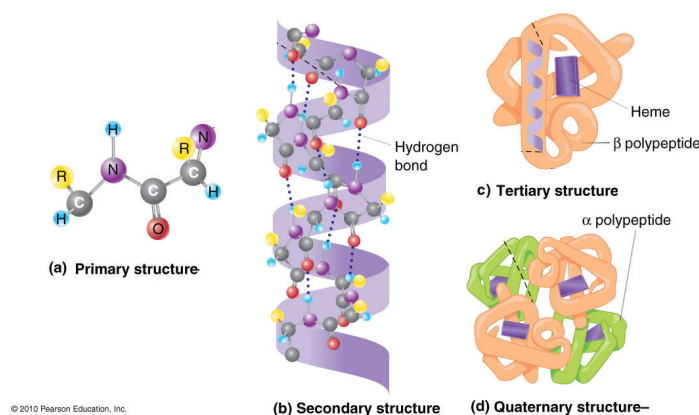


Figura 1: Ejemplo de las 4 estructuras de proteínas. Fuente: (Russell and Gordey, 2002)

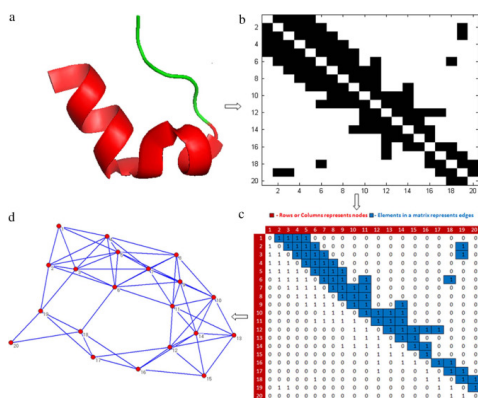


Figura 2: Ejemplo del contact map de una proteína.

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*.

3. Avances

En esta sección, detallaremos el estado anterior del proyecto y los avances realizados estas dos ultimas semanas.

3.1. Estado anterior

Se descargo las fuentes de CONFOLD2 Adhikari and Cheng (2018) y CNS para replicar los resultados de la predicción de estructuras de proteínas a partir del *contact map*. Dicha herramienta genera como salida un archivo .pdb que contiene la posición de cada atomo. Tambien se empezo el desarrollo de una herramienta que pueda visualizar los archivos .pdb, llamaremos a esta herramienta ArgosMol.

3.2. Avance y estado actual

Se inicio el desarrollo de la herramienta ArgosMol para visualizar archivos .pdb. Esta herramienta surge como una alternativa a herramientas *desktop* como: PyMol, VMD, Jmol y Chimera. Además, 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;

3.2.1. ArgosMol

ArgosMol surge como como una alterntiva Web, facil de usar y tomando en cuenta los principios presentados por Youkharibache (2017). La herramienta es presentada en la Figura 3 y esta disponible en este enlace. Además, ArgosMol está influenciada por otras herramientas Web como:

- **Jolecule.-** Desarrollada por Bosco Ho desde el 2011 hasta la actualidad. Enlace.
- **NGL.-** Presentada por Rose and Hildebrand (2015). Enlace.
- **EzMol.-** Presentada por Reynolds et al. (2018). Enlace.
- **iCn3D.-** Presentada por Wang et al. (2020). Enlace.

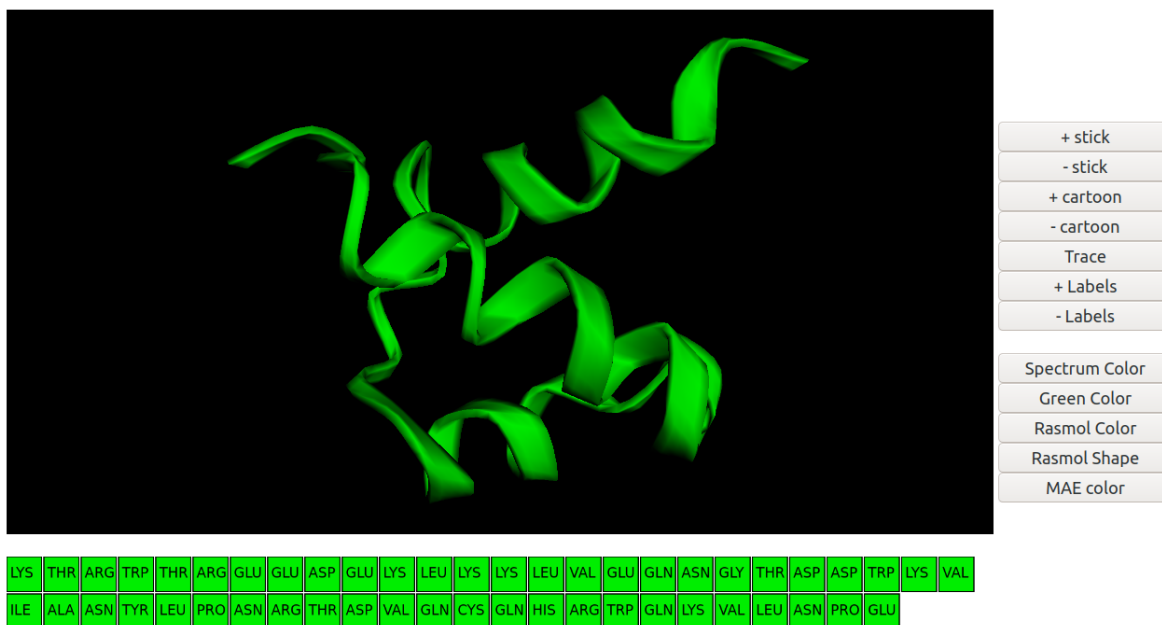


Figura 3: Ejemplo de visualización de proteínas con ArgosMol.

3.2.2. Funcionalidades implementadas

Las funcionalidades implementadas hasta ahora son:

- Permite tres vistas de la proteina y sus posibles combinaciones (ver Figura 4), las tres vistas son:

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.

- *Stick*: Enlaces entre átomos.
- *Cartoon*: Solo estructuras secundarias como hélices y pliegues.
- *Trace*: *Backbone* de la proteína.

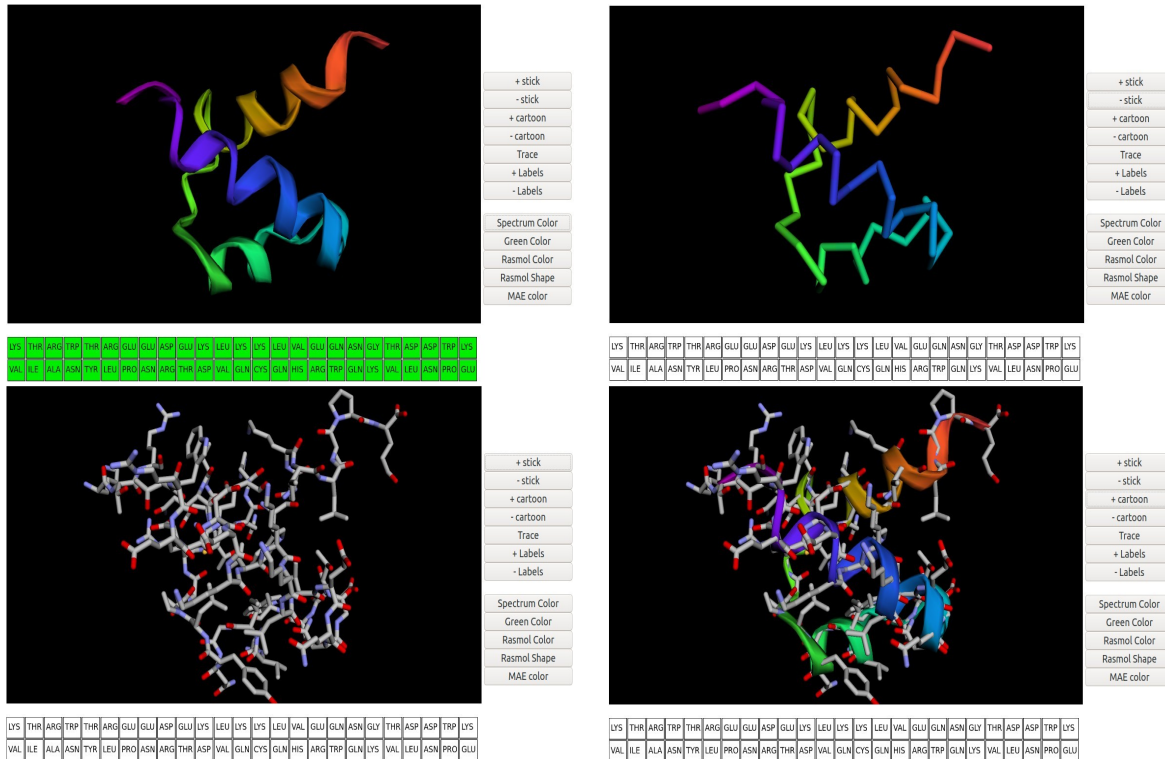


Figura 4: Ejemplo de las tres formas de ver la estructura de una proteína y sus posibles combinaciones entre ellas.

- También, según las recomendaciones del docente y algunos autores, se puede visualizar cada aminoácido por color. Se ha tomado en cuenta 5 esquemas de colores:
 - *Green*: Todos los aminoácidos en color verde.
 - *Spectrum*: Por estructura secundaria.
 - *Rasmol color*: Según la polaridad de cada aminoácido.
 - *Rasmol shape*: Según la forma de cada aminoácido.
 - *MAE*: Propuesta por la herramienta MAEditor.

En la Figura 3 mostramos el esquema *green*, mientras que en la Figura 5, mostramos los cuatro esquemas restantes. Además, debajo del modelo mostramos la secuencia de aminoácidos siguiendo el mismo esquema de colores (esta funcionalidad es propuesta por Youkharibache (2017)).

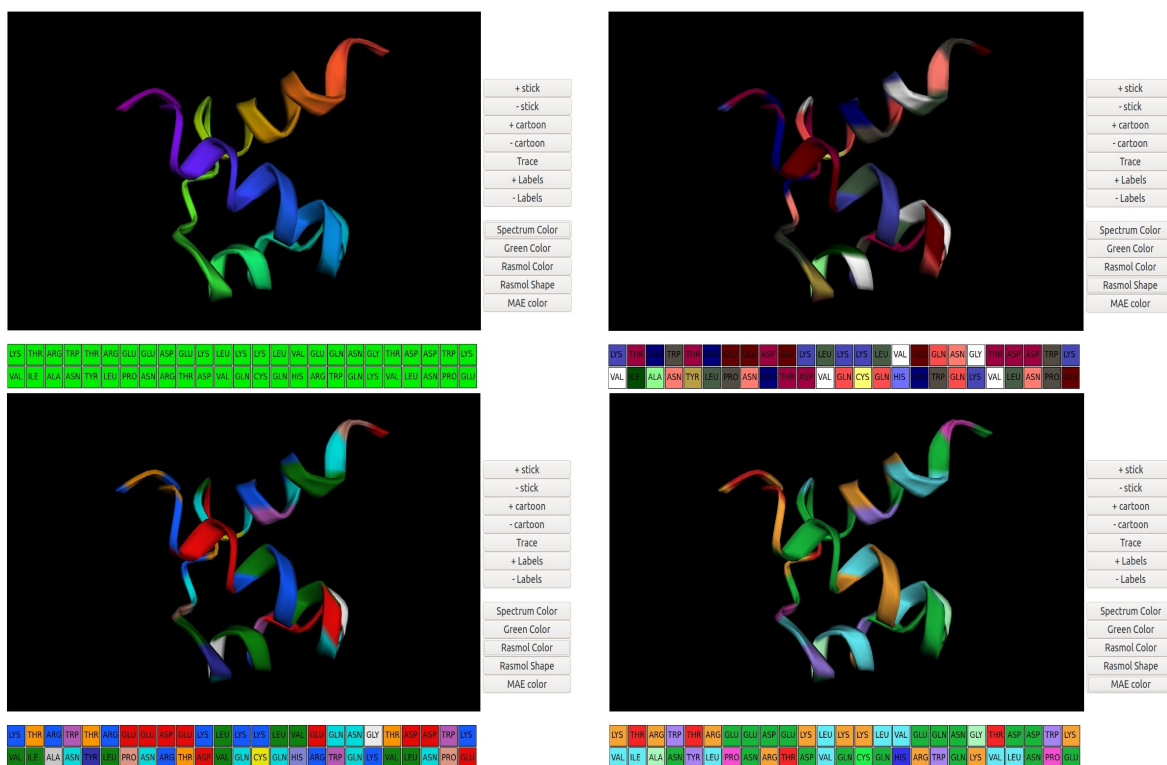


Figura 5: Ejemplo de los cinco esquemas de colores propuestos por ArgosMol. (Arriba Izq.): *spectrum*. (Arriba der.) *rasmol color*. (Abajo izq.) *rasmol shape*. (Abajo der.) *MAE*.

- También podemos ver las etiquetas de cada aminoácido en el modelo 3D. Estas etiquetas pueden activarse con eventos *click* del mouse sobre cada aminoácido o activando la opción de ver todas las etiquetas en conjunto. En la Figura 6, mostramos un ejemplo.

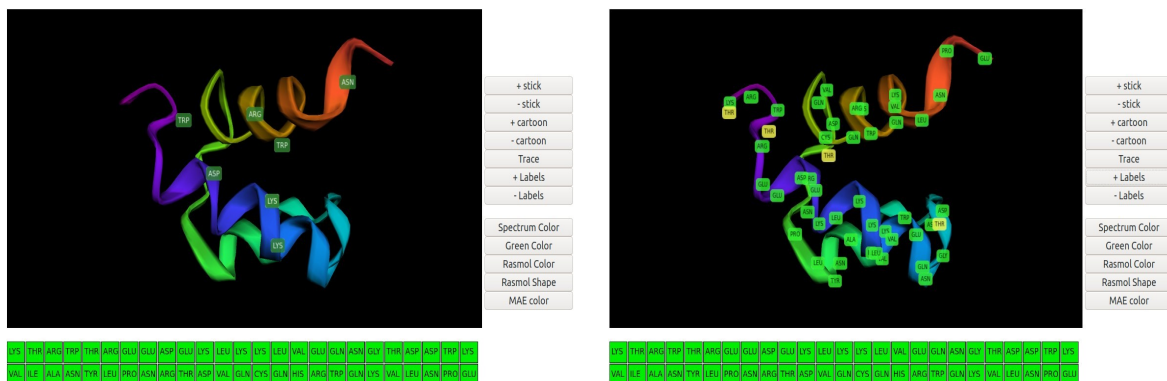


Figura 6: Ejemplo del uso de etiquetas en ArgosMol.

4. Comentarios finales

En este avance, se ha iniciado el desarrollo de una herramienta Web ArgosMol. Esta herramienta permite ver tres formas distintas una proteína; presenta cinco esquemas de colores, manteniendo relación entre el modelo 3D y la secuencia de aminoácidos; y permite ver etiquetas de los aminoácidos. Aún queda por desarrollar los eventos del mouse y teclado, visualización de varias cadenas polipeptídicas y la integración con CONFOLD2.

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.