

# Trabajo Práctico Final

Introducción a la Bioinformática - UNQ

**Cursada: 2° Cuatrimestre 2021**

**Docentes: Ana Julia Velez Rueda**

<b>1. Objetivo</b>	<b>1</b>
1.1 Objetivo Específico	1
<b>2. Contexto</b>	<b>1</b>
<b>3 Requerimientos detallados</b>	<b>3</b>
3.1 Carga de estructura problema	3
3.2 Búsqueda de proteínas homólogas y determinación de regiones conservadas	3
3.3 Cálculo de la energía de unión del ión a la proteína problema	4
3.4 Determinación de regiones conservadas	4
3.5 Visualización	4
3.4 Tecnologías	4
<b>4. Forma de entrega</b>	<b>5</b>
<b>5. Bibliografía</b>	<b>5</b>

## 1. Objetivo

El objetivo del presente trabajo práctico es integrar los conceptos biológicos y bioinformáticos desarrollados durante la materia, junto con los conocimientos, prácticas y habilidades propias de la informática y la programación adquiridas en otras materias, a través de la construcción de un software simple pero innovador que sirva de asistencia al proceso del análisis bioinformático y/o herramienta educativa para la enseñanza de la Biología.

### 1.1 Objetivo Específico

En este contexto nos proponemos desarrollar una aplicación que nos permita la visualización de regiones conservadas de unión a calcio en estructuras homólogas y la energía de unión a la proteína problema.

## 2. Contexto

La historia de las especies y cómo estas han cambiado desde que se desarrollara la vida en la tierra, ha quedado registrada en los genomas de las especies actuales. Los estudios

evolutivos permiten realizar inferencias estructurales y funcionales donde el conocimiento sobre los sistemas aún es insuficiente. Sin embargo, este tipo de inferencia, requiere de la estimación de distancias evolutivas entre especies, basadas en las diferencias entre genes ortólogos. Como bien ya explicamos dos secuencias que comparten un ancestro común se denominan secuencias homólogas (Reeck *et al.*, 1987). La predicción de homología se realiza extrayendo de las secuencias la información conservada durante la evolución, para lo que resulta necesario la comparación de las secuencias para identificar los residuos que tienen en común.

Como aprendimos durante el desarrollo de la materia, los organismos pueden acumular cambios heredables (en su ADN) con el paso del tiempo. Estos cambios a nivel genético, se traducen en cambios a nivel de las proteínas para las cuales estos genes codifican. Y como ya hemos visto la estructura de las proteínas se encuentra estrechamente relacionada a la función (Todd *et al.* 2001). Así mismo, teniendo en cuenta los conceptos relativos a la evolución, no les resultará ilógico pensar que: dado que la forma en que el plegamiento de una proteína se encuentra dirigida por su secuencia, y que está a su vez es codificada por la secuencia de un gen, entonces existirán restricciones evolutivas para los posibles cambios en la secuencia y estructura proteica (Grishin 2001; Zea *et al.* 2018). Es decir, que no todas las posibles mutaciones a nivel genético serán viables o producirán una proteína funcional, que le permita al organismo sobrevivir (Bujnicki 2001; Strokach *et al.* 2019). Es, por tanto, importante comprender y predecir los cambios estructurales que una mutación genética introduce (Bueno *et al.* 2018).

**Para pensar:** ¿qué factores pueden hacer que las mutaciones ocurran? ¿Cómo se relacionan estas con la diversidad biológica?

Estudiar los cambios estructurales que producen las distintas mutaciones es de particular importancia para la bioinformática, ya que nos permite predecir y manipular estos cambios con fines biotecnológicos (Gerlt 1987). En este sentido la construcción de herramientas que nos permitan visualizar y predecir los cambios estructurales que una mutación induce, ayudaría a mejorar dichas aplicaciones.

Nos centraremos en un caso particular de estudio, las proteínas de unión a calcio. El calcio iónico ( $\text{Ca}^{2+}$ ) cumple funciones esenciales tanto en organismos eucariotas como procariotas. Principalmente conocido por su rol en rutas de señalización como segundo mensajero en eucariotas, en procariotas su importancia fisiológica no es menor al estar relacionado con diversos procesos como la división celular, la quimiotaxis, la diferenciación (esporulación) e incluso en varios procesos como molécula de señalización. Tanto los roles fisiológicos como metabólicos del  $\text{Ca}^{2+}$  están modulados por un conjunto de proteínas que toman distintos roles como transportadores, proteínas sensoras de  $\text{Ca}^{2+}$ , bombas de intercambio de iones, proteínas quelantes, etc. La unión del  $\text{Ca}^{2+}$  a dichas proteínas se lleva a cabo por arreglos estructurales denominados “motivos”, que poseen también una impronta secuencial, de ahí que se hable tanto de motivos estructurales como secuenciales. La combinación de un mismo motivo en distintas proteínas efectoras, así como la existencia de distintos motivos, genera un enorme abanico de posibles relaciones entre la estructura-función de las proteínas lo que permite una sutil regulación del  $\text{Ca}^{2+}$  en los

organismos. Existen evidencias que sugieren que la variabilidad en las concentraciones de  $\text{Ca}^{2+}$  en los distintos compartimentos celulares, organismos y circunstancias metabólicas facilitó el surgimiento de distintos motivos, con distintas propiedades (por ej. afinidades) adecuadas para los distintos casos. Aún cuando se conoce ampliamente muchos de los fenómenos fisiológico-moleculares en los que se encuentra involucrado el Calcio, en la actualidad persisten grandes interrogantes concernientes al desarrollo y regulación de los mismos, su alteración y papel en enfermedades.

En este contexto nos proponemos desarrollar una aplicación que nos permita la visualización de regiones conservadas de unión a ligando en estructuras homólogas. Si bien ya existen programas capaces de dar respuesta parciales a algunos de estos cuestionamientos, no existen herramientas interactivas que permitan visualizar de forma dinámica la conservación de motivos de unión a calcio en estructuras homólogas.

**Para pensar:** ¿nos podría servir este software para establecer relaciones secuencia-estructura-función en proteínas homólogas?

## 3 Requerimientos detallados

Como bien explicitamos, el objetivo del presente trabajo es desarrollar un software que permita la visualización de regiones conservadas de unión a ligando en estructuras homólogas y la energía de unión a la proteína problema:

### 3.1 Carga de estructura problema

El sistema debe permitir la carga de una estructura proteica a analizar en el formato PDB<sup>1</sup> o su código PDB, para su posterior procesamiento. Si la estructura ingresada es inválida o el código PDB es inexistente, el software deberá notificar al usuario apropiadamente, mostrando un mensaje claro y útil. El programa debe explicitar en su documentación el formato del input requerido.

Nuestro software solo permitirá el análisis de proteínas de unión a calcio, por lo que se espera la validación del input en tanto a la presencia de calcio en el PDB.

### 3.2 Búsqueda de proteínas homólogas y determinación de regiones conservadas

A partir del código PDB el programa deberá obtener proteínas homólogas con estructura terciaria conocida, para la determinación de regiones conservadas de nuestra proteína problema. Para tal fin se deberán utilizar los software BLAST (Altschul et al. 1990). Luego de obtener las estructuras homólogas a la proteína problema, mediante un alineamiento estructural deberán visualizarse las regiones de unión a calcio.

---

<sup>1</sup>[https://en.wikipedia.org/wiki/Protein\\_Data\\_Bank\\_\(file\\_format\)](https://en.wikipedia.org/wiki/Protein_Data_Bank_(file_format))

### 3.3 Cálculo de la energía de unión del ión a la proteína problema

El cálculo de la energía de unión del ión a la proteína problema se realizará con el software FoldX<sup>2</sup>.

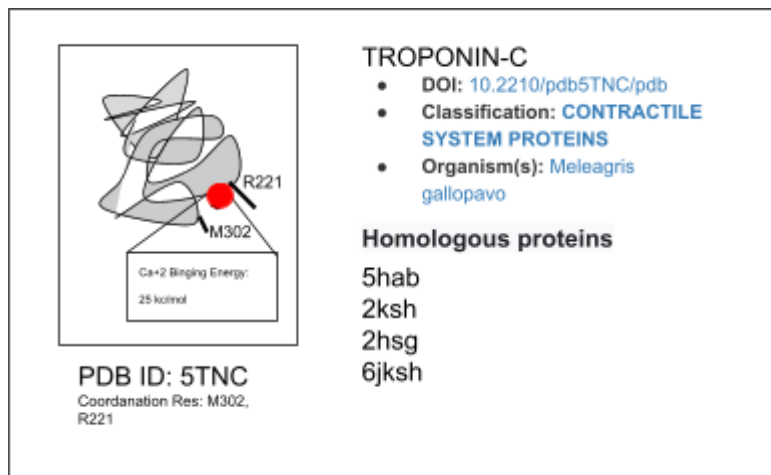
### 3.4 Determinación de regiones conservadas

La determinación de regiones estructuralmente homólogas se deberá realizar mediante alineamientos estructurales (podrá usarse Pymol).

**BONUS:** Como forma adicional, se considerará como un mejor programa aquel que realice los alineamientos estructurales utilizando Mammoth<sup>3</sup> o Dali<sup>4</sup>.

### 3.5 Visualización

Finalmente, el sistema deberá permitir la visualización de la estructura problema con las distintas región de unión a Calcio coloreada y su energía de unión al metal. Además se deberán poder seleccionar las proteínas homólogas para visualizar las regiones conservadas estructuralmente y que potencialmente pueden unir calcio. Para este fin, podrán ser tomado de base visualizadores de estructura pre - existentes como Pymol<sup>5</sup> o VMD<sup>6</sup>.



<sup>2</sup> FoldX Metal binding: <http://foldxsuite.crg.eu/command/MetalBinding>

<sup>3</sup> Mammoth:

<https://ub.cbm.uam.es/software/mammoth.php#:~:text=MAMMOTH,low%2Dresolution%20protein%20tertiary%20model>.

<sup>4</sup> Dali: <http://ekhidna2.biocenter.helsinki.fi/dali/>

<sup>5</sup> <https://pymol.org/2/>

<sup>6</sup> <https://www.ks.uiuc.edu/Research/vmd/>

**BONUS: Como forma adicional, se considerará como un mejor programa aquel que visualice o exponga en la interfaz con el usuario los residuos de coordinación del calcio**

## 3.4 Tecnologías

El software podrá estar implementado utilizando cualquier tecnología. Sin embargo, se recomienda fuertemente utilizar alguna de las siguientes:

- Python. Librerías recomendadas: Biopython, Pandas.
- JavaScript

Herramientas complementarias:

- FoldX: <http://foldxsuite.crg.eu/>
- Blast: [https://biopython.readthedocs.io/en/latest/chapter\\_blast.html](https://biopython.readthedocs.io/en/latest/chapter_blast.html),  
[https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=Download](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download)
- Pymol: <https://pymol.org/2/>
- Dali: <http://ekhidna2.biocenter.helsinki.fi/dali/>
- Mammoth: <https://ub.cbm.uam.es/software/mammoth.php#:~:text=MAMMOTH,low%2Dresolution%20protein%20>

## 4. Forma de entrega

El trabajo práctico se realizará en equipos de hasta 6 integrantes, con **fecha de entrega del proyecto el día miércoles 14 de Julio de 2021**. El mismo deberá estar en un repositorio público en Github y será entregado por medio del GitHub Classroom de la Organización de la materia (<https://github.com/BioinformaticaUNQ>). El proyecto deberá contener un archivo README.md con los datos de l@s integrantes del equipo y la documentación correspondiente.

Los trabajos, además deberán ser presentados por l@s miembros del equipo en una **exposición oral (con DEMO)**, los días **16 y 23 de Julio de 2021**.

## 5. Bibliografía

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–10

Bueno CA, Potoyan DA, Cheng RR, Wolynes PG. 2018. Prediction of changes in protein folding stability upon single residue mutations. *Biophys. J.* 114(3):199a

- Bujnicki JM. 2001. Understanding the evolution of restriction-modification systems: clues from sequence and structure comparisons. *Acta Biochim. Pol.* 48(4):935–67
- Gerlt JA. 1987. Relationships between enzymatic catalysis and active site structure revealed by applications of site-directed mutagenesis. *Chem. Rev.* 87(5):1079–1105
- Grishin NV. 2001. Fold change in evolution of protein structures. *J. Struct. Biol.* 134(2–3):167–85
- Strokach A, Corbi-Verge C, Teyra J, Kim PM. 2019. Predicting the Effect of Mutations on Protein Folding and Protein-Protein Interactions. *Methods Mol. Biol.* 1851:1–17
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673–80
- Todd AE, Orengo CA, Thornton JM. 2001. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* 307(4):1113–43
- Zea DJ, Monzon AM, Parisi G, Marino-Buslje C. 2018. How is structural divergence related to evolutionary information? *Mol. Phylogenet. Evol.* 127:859–66
- Zhang Y, Sagui C. 2015. Secondary structure assignment for conformationally irregular peptides: comparison between DSSP, STRIDE and KAKSI. *J. Mol. Graph. Model.* 55:72–84