



# TESINA DE LICENCIATURA

**Título:** Human Computation en bioinformática. Desarrollo de un videojuego para la clasificación de proteínas.

**Autores:** Ezequiel Colautti – Martín Moro

**Director:** Dr. Julian Echave

**Codirector:** Ing. Armando De Giusti

**Carrera:** Licenciatura en Sistemas

## Resumen

Una de las áreas de investigación más importantes de la biología son las proteínas. Para avanzar en su investigación, la clasificación de las más de 100.000 proteínas depositadas en el Banco de Datos de Proteínas en un conjunto relativamente menor de categorías es una tarea muy importante. Teniendo esto en cuenta, en este trabajo se propone aplicar el paradigma de Human Computation a la resolución de este problema de la Bioinformática. Para eso desarrollamos P3, un GWAP (del inglés, Game With A Purpose) que tiene como objetivo generar los datos necesarios para clasificar las proteínas a partir de la comparación de la representación tridimensional de las estructuras proteicas realizada por los jugadores. Por otro lado estudiamos la posibilidad de generar una clasificación de proteínas basadas en la información de los movimientos de las mismas además de su estructura.

## Palabras Claves

- Human Computation
- GWAP
- Proteínas
- Clasificación estructural de proteínas
- Clasificación de proteínas basada en el movimiento
- Bioinformática

## Trabajos Realizados

En primer lugar se realizó una investigación bibliográfica de los temas a desarrollar. Luego se programó una aplicación basada en el paradigma Human Computation, llamada "P3". P3 es un videojuego web en el que los jugadores deben identificar en un trío de proteínas aquella que consideren diferente de las dos restantes. Los jugadores avanzan por los distintos niveles a partir de la resolución correcta de una cantidad específica de tríos y son desafiados con niveles de dificultad progresiva para motivarlos a continuar jugando y respondiendo correctamente. Se realizaron pruebas con 30 jugadores para generar los datos y poder analizar la performance de la aplicación.

## Conclusiones

Luego de realizada la aplicación y la generación de los datos a través de la misma, en función del análisis de los resultados obtenidos, puede observarse que es posible realizar una clasificación estructural de las proteínas que posea un alto grado de coincidencia con la clasificación SCOP. En cuanto al desarrollo de una clasificación basada en sus movimientos, si bien existen ciertas tendencias que indicarían que existe la posibilidad, se necesitarán más datos para verificar si agregar información sobre movimientos facilita la clasificación y/o conduce a una clasificación diferente que una puramente estructural.

## Trabajos Futuros

En primer lugar sería interesante profundizar en la generación de una clasificación basada en el movimiento de las proteínas analizando los posibles cambios que deberían aplicarse a P3 para cumplir este objetivo y generando un volumen de datos mayor. Por otro lado, sería útil desarrollar una mecánica de juego que permita resolver la clasificación de las más de 100.000 proteínas de una manera más eficiente y con la participación de una menor cantidad de jugadores.