

---

**APLICACIONES DE ESTRUCTURAS DE GRAFOS Y APRENDIZAJE  
PROFUNDO A SISTEMAS DE CLASIFICACIÓN DE INTERACCIÓN  
ANTÍGENO-ANTICUERPO**

**CLAUDIO GUEVARA VÁSQUEZ  
INGENIERO CIVIL EN BIOINFORMÁTICA**

**RESUMEN**

Las interacciones proteína-proteína son de real importancia para la ingeniería de proteínas debido a que forman parte esencial en la mayoría de los procesos moleculares. Un caso particular es la interacción antígeno-anticuerpo, la cual cumple con el rol de inhibir o neutralizar agentes patógenos que afectan negativamente la homeostasis normal del cuerpo. Conocer el funcionamiento específico de un anticuerpo es de gran interés en áreas como la medicina y la farmacología, ya que facilita el diseño de vacunas y medicamentos. Variados métodos experimentales se han desarrollado con el fin de estudiar las interacciones proteína-proteína. Algunos de los ejemplos clásicos corresponden a los microarrays de ADN y proteína, la espectroscopia de masas (MS) y la letalidad sintética. Sin embargo, estos métodos se caracterizan por tener un alto costo de producción y tiempo de desarrollo, ser susceptibles al error humano, y muchas veces, requieren de un elevado conocimiento. Para solventar este problema, cada vez se aplican más técnicas basadas en Machine Learning y Deep Learning, como es el caso de AlphaFold y su capacidad de predecir la estructura secundaria. Sin embargo, la rama de la inteligencia artificial aún debe ser más estudiada y aplicada en la investigación científica. Con base en esto, se realizó una investigación con el fin de predecir la interacción antígeno-anticuerpo por medio de Graph Neural Network. Para lograr esto, las proteínas se representaron por medios de estructuras de grafos, en donde los nodos correspondían a los residuos de las proteínas, mientras que las aristas a las distancias euclidianas entre aminoácidos. Además, se contó con un clasificador de interacción para cada complejo. En general, se obtuvo un rendimiento alrededor del 0,51 y se planteó una serie de puntos a tratar en futuros trabajos para el perfeccionamiento de los modelos, los cuales tienen que ver con arquitecturas de redes neuronales,

---

representación de grafos, métodos de codificación y predicción de complejos proteicos. Demostrando que, a pesar de que los resultados no fueron los esperados, el camino por delante es extenso y queda un largo desarrollo por realizar, con el fin de llegar a elaborar sistemas predictivos en base a esta arquitectura de deep learning.

## ABSTRACT

Protein-protein interactions are significant for protein engineering because they are essential to most molecular processes. A particular case is an antigen-antibody interaction, which fulfills the role of inhibiting or neutralizing pathogenic agents that negatively affect the normal homeostasis of the body. Knowing the specific functioning of an antibody is of great interest in areas such as medicine and pharmacology since it facilitates the design of vaccines and drugs. Various experimental methods have been developed to study protein-protein interactions. Some classic examples are DNA and protein microarrays, mass spectroscopy (MS), and synthetic lethality. However, these methods are characterized by high production cost and development time, being susceptible to human error, and often requiring a high level of knowledge. To solve this problem, more and more techniques based on Machine Learning and Deep Learning are being applied, such as AlphaFold and its ability to predict secondary structure. However, the branch of artificial intelligence still needs to be more studied and used in scientific research. Based on this, an investigation was carried out to predict the antigen-antibody interaction using the Graph Neural Network. Proteins were represented by employing graph structures, where nodes corresponded to protein residues while edges to Euclidean distances between amino acids. In addition, there was an interaction classifier for each complex. In general, a performance of around 0.51 was obtained, and a series of points were raised to be dealt with in future Works to improve the models, which have to do with neural network architectures, graph representation, coding methods, and prediction of protein complexes. Even though the results were not as expected, the road ahead is long, and there is a long development to be done to develop predictive systems based on this deep learning architecture.