

---

**PREDICCIÓN DE SITIOS DE UNIÓN AL ADN EN LA SUPERFAMILIA FUR  
MEDIANTE INTELIGENCIA ARTIFICIAL Y DESCRIPTORES MOLECULARES**

**JÉSSICA FERNANDA LARA MUÑOZ  
INGENIERO CIVIL EN BIOINFORMÁTICA**

**RESUMEN**

La Superfamilia Fur se compone de proteínas reguladoras de absorción de iones, estas poseen una similitud funcional y capacidad de unión al ADN, varían según el cofactor metálico que utilicen dentro de las cuales podemos encontrar del tipo Fur, Zur, Mur, Nur y PerR. El estudio del sitio de unión en proteínas de la Superfamilia Fur y factores de transcripción es un tema de gran interés debido a que este tipo de proteínas pueden ser potencialmente utilizadas en desarrollos y aplicaciones antimicrobianas. En este estudio se caracterizaron 63 sitios de unión al ADN de organismos bacterianos pertenecientes a Fur y diversos factores de transcripción, mediante descriptores moleculares de secuencia y estructura. Utilizando técnicas de inteligencia artificial se realizó la identificación de características relevantes para entrenar los modelos predictivos SVM y RF. Del análisis realizado se determinó que el mejor modelo entrenado con características de secuencia es SVM logrando un rendimiento de exactitud (Accuracy) de 84.6%, mientras que utilizando características estructurales el modelo RF es el que logra un mejor rendimiento con una exactitud (Accuracy) de 77.3%. Los resultados de la predicción indican que es posible desarrollar un modelo predictivo de zonas de unión al ADN en proteínas con una alta precisión utilizando solo la información de secuencia.

## ABSTRACT

The Fur superfamily is composed of ion absorption regulatory proteins. These have functional similarities and DNA binding capacity. The diversity depends on the metal cofactor used, among which they can find Fur, Zur, Mur, Nur, and PerR types. The study of metal binding site proteins in the Fur superfamily and transcription factors is a very interesting topic because these types of proteins can potentially be used in antimicrobial development and its applications. In this investigation, 63 DNA-binding sites of bacterial organisms belonging to the Fur family and various transcription factors were characterized by molecular descriptors of sequence and structure. The use of intelligent artificial intelligence techniques made it possible to identify relevant characteristics for training SVM and RF predicting models. The performed analysis determined that the best predicting model using sequence characteristics is SVM, achieving a performance of 84.6% accuracy while using structural characteristics the RF model achieved the best performance with 77.3% accuracy. The results of the prediction indicate that it is possible to develop a DNA-binding site predictive model in proteins with high accuracy using only sequence information.