

**DISCRIMINACIÓN ENTRE SITIOS DE UNIÓN A METALES MEDIANTE EL USO
DE MÁQUINAS DE VECTORES DE SOPORTE Y LA COMBINACIÓN DE
DIVERSOS TIPOS DE INFORMACIÓN**

**ALFREDO JESÚS PEREIRA TOLOZA
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

Las metaloproteínas son proteínas que requieren la interacción con un metal para desempeñar su función biológica. Aproximadamente un tercio de las proteínas unen al menos un ion metálico en su estructura. Actualmente, en el Protein Data Bank, las metaloproteínas más abundantes son aquellas que unen de Zn, Mg o Ca. Los iones metálicos pueden ayudar por ejemplo, a estabilizar la estructura tridimensional de una proteína, inducir cambios conformacionales para regular la función de una proteína, y participar directamente en la actividad catalítica de enzimas. La región de una proteína que une un metal específico es llamada sitio de unión de metal. Conocer estos sitios de unión es fundamental para entender la función de una proteína, pero su identificación implica procedimientos experimentales largos y costosos. Con este fin, diferentes métodos computacionales han sido propuestos recientemente para la predicción de estos sitios de unión, incluyendo diversos enfoques basados en técnicas de Machine Learning. La mayoría de los métodos previos han empleado un sólo tipo de información a la vez, ya sea evolutiva, geométrica o fisicoquímica. Además, los trabajos más recientes están mayormente enfocados en predecir sitios de unión para un sólo tipo de metal a la vez, principalmente de Zn, sin la capacidad de discriminar entre sitios de unión a diferentes metales. En el presente estudio, se ha desarrollado un método para predecir y discriminar sitios de unión a Zn, Mg y Ca, utilizando Máquinas de Vectores Soporte y la combinación de información evolutiva, geométrica y fisicoquímica. El método ha sido capaz de distinguir de manera eficiente entre estos sitios de unión, alcanzando una exactitud superior al 90% y una tasa de falsos positivos que no excede del 5%. Además, el estudio proporciona evidencia de que la integración de los diversos tipos de información mejora el

desempeño de la discriminación entre diferentes sitios de unión a metales cuando un sólo tipo de información no es suficiente.

Palabras clave: Sitios de Unión a Metal, Máquinas de Vectores de Soporte, Discriminación.

ABSTRACT

Metalloproteins are proteins that require interaction with a metal to perform its biological function. Approximately one third of the proteins bind at least one metal ion in its structure. Currently at the Protein Data Bank, the most abundant metalloproteins are those binding Zn, Mg or Ca. Metal ions may help for instance, to stabilize the three dimensional structure of a protein, to induce conformational changes for protein function regulation, and directly participating in the enzyme catalytic activity. The region of a protein which binds a specific metal is called metal binding site. Knowing these binding sites is critical to understand the function of a protein, but identifying them involves long and costly experimental procedures. In order to do this, different computational methods have been recently proposed for the prediction of these binding sites, including various approaches focused on Machine Learning techniques. Most of the previous methods have employed only one type of information at the time, whether evolutionary, geometrical or physicochemical. In addition, the most recent attempts are mostly concerned to predict binding sites for only one type of metal at the time, mainly Zn, without the ability to discriminate between different metal binding sites. In the present study, it has been developed a method for prediction and discrimination of binding sites to Zn, Mg and Ca, using Support Vector Machines and the combination of evolutionary, geometric and physicochemical information. The method has been able to distinguish efficiently between these binding sites, reaching an accuracy over 90% and a false positive rate that do not exceed 5%. In addition, the study provides evidence that the integration of various types of information improves the performance of the discrimination between different metal binding sites when a single type of information is not enough.

Keywords: Metal Binding Sites, Support Vector Machines, Discrimination