
**DISEÑO DE UNA INTERFAZ GRÁFICA PARA BÚSQUEDA DE PATRONES
ESTRUCTURALES EN EL PROTEIN DATA BANK**

**DIEGO ALEJANDRO ANDRÉS CISTERNA VÁSQUEZ
INGENIERÍA CIVIL EN COMPUTACIÓN**

RESUMEN

La estructura que posee la materia define sus propiedades y funciones. En las últimas décadas, gracias a los avances de la ciencia y la tecnología, investigadores han utilizado variadas técnicas para determinar la estructura de macromoléculas tales como proteínas, ADN y ARN. La información conseguida es útil para describir las interacciones que se dan entre moléculas, y así poder entender las estructuras y funciones que permiten muchos procesos biológicos importantes. El volumen de información de macromoléculas que se maneja hoy en día es muy grande. Por esto se han creado bancos de datos que intentan unificar y mantener accesible la Información a nivel global. Este es el caso del Protein Data Bank (PDB). Sin embargo, poder analizar los datos almacenados es una tarea difícil, que solo se lleva a cabo mediante herramientas de software especializadas. La mayoría de las herramientas disponibles actualmente poseen interfaces estándar y se limitan a realizar búsquedas a través de formularios, con los cuales no se pueden elaborar consultas avanzadas que entreguen datos realmente trabajados. Para efectuar una consulta compleja, las herramientas ofrecen confeccionarlas y ejecutarlas directamente, por ejemplo, en código SQL o a través de lenguajes de programación. Este proyecto se enfoca en entregar una solución práctica para un problema específico, la búsqueda de patrones estructurales a nivel de interacción entre un ligando y los aminoácidos que lo rodean, en su sitio de unión en las macromoléculas. La solución consiste en el diseño de una interfaz gráfica que permita dar forma a un patrón estructural basado en elementos de grafo, y que luego tal patrón sea transformado en una consulta SQL para buscar coincidencias dentro de una base de datos relacional que contiene la información de PDB. Respecto a las contribuciones, en este proyecto se implementó un prototipo funcional que satisface las necesidades originadas por el problema abordado, el prototipo fue validado, y evaluado positivamente por un grupo de usuarios calificados. Además, en el proceso se desarrolló un método que permite transformar asociaciones de pares de

componentes gráficos en consultas SQL, y que al unirlos pueden representar patrones estructurales de alta complejidad. Por último, el aporte a la comunidad científica es facilitar el estudio de estructuras por medio de una interfaz comprensible y fácil de usar, que revoluciona el modo de interacción habitual.

Palabras clave: patrones estructurales - Interacciones proteínas- ligando- Interfaz gráfica de usuario - sitio de unión – PDB

ABSTRACT

The structure of the matter defines its properties and functions. In the last decades, thanks to the advances of the science and the technology, researchers have used various techniques to determine the structure of macromolecules such as proteins, DNA, and RNA. The information obtained is useful to describe the interactions occurring between molecules, and thus be able to understand the structures and functions that allow many important biological processes. The volume of formation of macromolecules that is handled today is very big. Therefore, data banks have been created to try unifying and maintaining accessible the information at a global level. This is the case of the Protein Data Bank (PDB). However, being able to analyze the stored data is a difficult task, that is only carried through specialized software tools. The majority of the tools available currently have standard interfaces and are limited to execute searches through forms, with which is not possible to elaborate advanced queries that return data actually worked. To make a complex query, the tools offer to compose them and execute them directly, for example, in SQL code or through programming languages. This project focuses on delivering a practical solution to a specific problem, the search for structural patterns at an interaction level between a ligand and the amino acids surround it, in their binding site in the macromolecules. The solution consists of the design of a graphic interface that allows giving form to a structural pattern based on graph elements, and that then such pattern be transformed into a SQL query to look for coincidences inside a relational database that contains the information of PDB. Respect to the contributions, in this project was implemented a functional prototype that satisfies the necessities caused by the problem addressed, the prototype was validated, and evaluated positively by a group of qualified users. Moreover, in the process was developed a method that allows transforming associations of pairs of graphic components into SQL queries, and that by uniting them can represent structural patterns of high complexity. Finally, the contribution to the scientific community is to facilitate the study of structures through an understandable and easy to use interface, what revolutionizes the habitual way of interaction.

Keywords: Structural patterns – Protein - ligand interactions - Graphical user interface -Binding site - PDB