

UCLV
Universidad Central
"Marta Abreu" de Las Villas



FQF
Facultad de
Química y Farmacia

Departamento de Química

TRABAJO DE DIPLOMA

“Creación de un flujo de trabajo automatizado, en KNIME, para el estudio de la Relación Estructura – Actividad de una base de datos de compuestos orgánicos con actividad antimalárica”

Autor: Elier E. Abreu Martínez

Tutores: Dr.C. Aliuska Morales Helguera

Dr.C. Christophe Molina

Santa Clara, agosto del 2020
Copyright©UCLV

UCLV
Universidad Central
"Marta Abreu" de Las Villas



FQF
Facultad de
Química y Farmacia

Department of Chemistry

BACHELOR THESIS

“Creation of an automated workflow, in KNIME, for
the Structure-Activity Relationship study of a
database of organic compounds with antimalarial
activity”

Author: Elier E. Abreu Martínez

Supervisors: Ph.C. Aliuska Morales Helguera

Ph.C. Christophe Molina

Santa Clara, August 2020
Copyright©UCLV

Este documento es Propiedad Patrimonial de la Universidad Central “Marta Abreu” de Las Villas, y se encuentra depositado en los fondos de la Biblioteca Universitaria “Chiqui Gómez Lubián” subordinada a la Dirección de Información Científico Técnica de la mencionada casa de altos estudios.

Se autoriza su utilización bajo la licencia siguiente:

Atribución- No Comercial- Compartir Igual



Para cualquier información contacte con:

Dirección de Información Científico Técnica. Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní. Km 5½. Santa Clara. Villa Clara. Cuba. CP. 54 830

Teléfonos.: +53 01 42281503-1419

RESUMEN

El análisis de la Relación Estructura – Actividad (SAR, por sus siglas en inglés) desempeña un papel fundamental en la comprensión de los determinantes estructurales de la actividad biológica y en la optimización de compuestos para convertirse en líderes para la evaluación preclínica. Los programas encaminados a realizar este análisis ofrecen una baja interactividad en la manipulación de los resultados obtenidos. En este estudio, se presenta el primer flujo de trabajo automatizado, en KNIME, que permite combinar dos métodos del análisis de la SAR, creación e inspección de un gráfico de similitud tipo red (NSG, por sus siglas en inglés) y detección y remoción de *Activity Cliff*. Para ello se tomaron los compuestos provenientes del GNF y se curaron obteniéndose una base de datos de 4682 compuestos con su EC_{50} (nM) reportado, que fueron etiquetados como “activos” ($EC_{50} \leq 1000 nM$) e “inactivos” ($EC_{50} > 1000 nM$).

El análisis de la SAR primero crea el NSG y extrae la información de la SAR disponible en cada uno de los clústeres y luego identifica los subconjuntos de compuestos que inducen la discontinuidad SAR local en los conjuntos de datos (*Activity Cliff*). Por tanto, fue identificado en la base de datos subconjuntos con un elevado contenido de información de la SAR local, incluidos los quimiotipos antipalúdicos conocidos. Esta información debería ser útil para priorizar y seleccionar compuestos candidatos antipalúdicos.

ABSTRACT

Structure – Activity Relationship (SAR) analysis plays an important role in understanding structural determinants of biological activity and optimizing compounds to become leaders for preclinical evaluation. The programs designed to carry out this analysis offer low interactivity in the manipulation of the results obtained. In this study, the first automated workflow is presented, in KNIME, which allows combining two methods of SAR analysis, creation and inspection of a network similarity graph (NSG) and detection and removal of *Activity Cliff*. For this, the compounds from GNF were taken and cured, obtaining a database of 4682 compounds with their EC_{50} (nM) reported, which were labeled as "active" ($EC_{50} \leq 1000 nM$) and "inactive" ($EC_{50} > 1000 nM$).

The SAR analysis first creates the NSG and extracts the SAR information available in each of the clusters and then identifies the subsets of compounds that induce the local SAR discontinuity in the database (*Activity Cliff*). Therefore, highly informative subsets of local SAR, including known antimalarial chemotypes, were identified in the database. This information should be useful in prioritizing and selecting antimalarial candidate compounds.