



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA DE SÃO CARLOS
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
Grupo de Química Medicinal & Biológica - NEQUIMED

**INTEGRAÇÃO DE FERRAMENTAS EM BIO E
QUIMIOINFORMÁTICA PARA O ESTUDO DE INIBIDORES
DE CISTEÍNO PROTEASES**

Candidato: Thiago Kelvin Brito Matos

Orientador: Prof. Dr. Andrei Leitão

SÃO CARLOS

2023

THIAGO KELVIN BRITO MATOS

Integração de ferramentas em bio e quimioinformática para o estudo de
inibidores de cisteíno proteases

Tese apresentada ao Instituto de Química de
São Carlos da Universidade de São Paulo,
como parte dos requisitos para obtenção do
Título de Doutor em Ciências.

Área de concentração: Química Orgânica e
Biológica

Orientador: Prof. Dr. Andrei Leitão

SÃO CARLOS

2023

RESUMO

Inibidores de cisteíno proteases têm apresentado uma grande gama de atividades farmacológicas *in vitro* e *in vivo* devido à importância desta classe de alvos para processos fisiopatológicos no organismo humano, em parasitos e vírus. Este trabalho visa estabelecer modelos computacionais para a predição da atividade biológica e modo de interação de novos inibidores de cisteíno proteases, além da aplicação da quimioinformática usando modelos de aprendizado de máquinas para prever a solubilidade em água de novas estruturas químicas. O trabalho foi baseado na mineração de dados referentes a inibidores de cisteíno catepsinas humanas (B, K, L, S) com definição de modelos qualitativos e/ou quantitativos para a predição da potência de novos inibidores de cisteíno proteases a serem sintetizados no grupo de pesquisa usando análise de fragmentos moleculares, docagem covalente e dinâmica molecular. Deste estudo, duas estruturas químicas foram selecionadas como mais promissoras para a inibição da catepsina L, a partir da modificação da região P3 da molécula visando a formação de uma ligação hidrogênio com o resíduo Gly61 no subsítio S3 da enzima. Outro estudo aplicou a docagem covalente em peptídeos contendo nitrila, com a descrição da informação das interações intermoleculares e regiões com impedimento estérico que auxiliaram a compreensão do perfil bioativo desta série de inibidores de cisteíno proteases. A mesma estratégia foi aplicada em inibidor covalente da CALB, em um trabalho em colaboração. Assim, este trabalho envolveu mineração de dados em banco de dados curado, modelagem molecular, análise de ligantes e macromoléculas de forma a integrar um fluxograma de análise de uma nova estrutura química visando uma predição automatizada de alguns parâmetros farmacodinâmicos e farmacocinéticos a partir da representação estrutural inicial.

Palavras-chave: bioinformática, quimioinformática, química computacional, mineração de dados, aprendizado de máquinas, descoberta e planejamento de fármacos.

ABSTRACT

Cysteine protease inhibitors have presented a wide range of *in vitro* and *in vivo* pharmacological activities due to the importance of this class of targets for pathophysiological processes in the human body, in parasites and viruses. This work aims to establish computational models for predicting biological activity and modes of interaction of new cysteine protease inhibitors, in addition to the application of cheminformatics using machine learning models to predict water solubility of new chemical structures. The work was based on the data mining of human cathepsin inhibitors (B, K, L, S) defining qualitative and/or quantitative models for predicting the potency of new cysteine protease inhibitors to be synthesized in the research group using molecular fragment analysis, covalent docking, and molecular dynamics. From this study, two chemical structures were selected as the most promising for cathepsin L inhibition, from the modification of the P3 region of the molecule aiming at the formation of a new hydrogen bond with the residue Gly61 in the S3 subsite of this enzyme. Another study applied covalent docking in nitrile-containing peptoids, describing information on intermolecular interactions and regions with steric hindrance that helped understand the bioactive profile of this series of cysteine protease inhibitors. The same approach was also applied to a covalent inhibitor of CALB in a collaborative work. Thus, this work involved data mining in a curated database, molecular modeling, ligand analysis, and macromolecules to integrate a workflow to analyze a new chemical structure aimed to become an automated prediction of some pharmacodynamic and pharmacokinetic parameters from the initial structural representation.

Keywords: bioinformatics, cheminformatics, computational chemistry, data mining, machine learning, drug discovery and drug design.

