

Título em Português:	Estudos Computacionais para a Identificação de Moléculas Bioativas para Doenças Tropicais Negligenciadas
Título em Inglês:	Computational studies for the identification of bioactive molecules for Neglected Tropical Diseases
Autor:	Pablo Henrique Figueredo
Instituição:	Universidade de São Paulo
Unidade:	Instituto de Física de São Carlos
Orientador:	Leonardo Luiz Gomes Ferreira
Área de Pesquisa / SubÁrea:	Biofísica Molecular
Agência Financiadora:	USP - Programa Unificado de Bolsas

Estudos Computacionais para a Identificação de Moléculas Bioativas para Doenças Tropicais Negligenciadas

Pablo Henrique Figueredo

João Victor Silva-Silva, Prof. Dr. Adriano D. Andricopulo

Prof. Dr. Leonardo Luiz Gomes Ferreira

Instituto de Física de São Carlos/ Universidade de São Paulo

pablofigueredo260@usp.br

Objetivos

(i) Construção de um banco de dados, a partir da literatura científica, com compostos ativos contra a forma amastigota intracelular de *Leishmania infantum* (ii) Análise das propriedades físico-químicas dos compostos do banco de dados (iii) Identificação de relações estrutura-atividade (SAR) para determinar descritores moleculares e subestruturas químicas associados à atividade dos compostos (IC_{50}).

Métodos e Procedimentos

Para a construção do banco de dados, foi realizado um levantamento bibliográfico, entre os anos de 2022–2024, no Google Acadêmico, utilizando “*Leishmania infantum* antileishmanial activity amastigote” como palavras-chave. Para cada composto de interesse, foram registradas informações sobre nome, atividade biológica (IC_{50}), citotoxicidade (CC_{50}) e DOI do artigo de origem. Com base nesses dados, as estruturas moleculares foram construídas no software MarvinSketch para geração dos códigos SMILES. Em seguida, foram calculadas as propriedades físico-químicas no software DataWarrior [1], com foco nos descritores da “Regra dos Cinco” de Lipinski (massa molecular, cLogP, número de doadores e aceptadores de ligações de hidrogênio) [2], determinando-se o número de violações para

cada composto. A partir dessas informações, realizou-se uma análise de scaffolds no DataWarrior para identificar os núcleos estruturais mais recorrentes entre os compostos que atenderam aos critérios de seleção pré-estabelecidos: ausência de violações à Regra de Lipinski e $IC_{50} \leq 10 \mu M$.

Resultados

A partir do levantamento bibliográfico, foi construído um banco de dados contendo 232 compostos com atividade reportada contra *L. infantum*. A análise das propriedades com base na “Regra dos Cinco” de Lipinski revelou um perfil físico-químico favorável para a maioria das moléculas. Conforme ilustrado na Figura 1, 135 compostos (aproximadamente 58,2%) não apresentaram violações, indicando boa compatibilidade com fármacos de administração oral. Ao correlacionar o número de violações com a potência biológica, observa-se que uma parcela significativa dos compostos mais ativos (menor IC_{50}) concentra-se no grupo sem violações. A Figura 2 apresenta a análise de scaffolds realizada com o subconjunto de compostos mais promissores, selecionados por não apresentarem violações à Regra de Lipinski e $IC_{50} \leq 10 \mu M$. Os dois candidatos com maior atividade no banco de dados, composto **46** ($IC_{50} = 0,45 \mu M$) e **209** ($IC_{50} = 0,105 \mu M$),

apresentam scaffolds moleculares únicos, (1-[(benzotriazol-1-ylmethyl)diselanyl)methyl]benzotriazole) e (2-phenylchromen-4-one), respectivamente. Essa singularidade estrutural se mantém mesmo ao se considerar a totalidade do banco de dados.

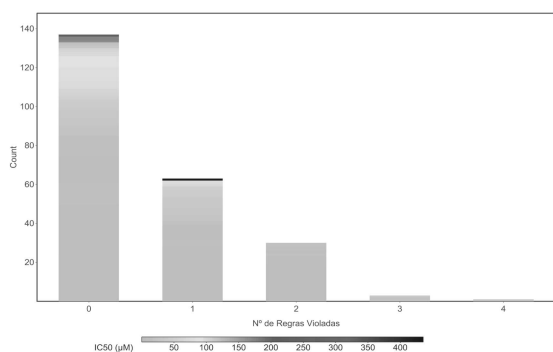


Figura 1: Análise da atividade anti-Leishmania (IC_{50}) em função do perfil de violações à Regra de Lipinski.

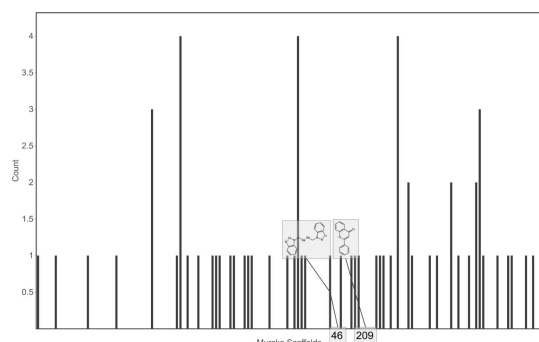


Figura 2: Análise de scaffolds destacando a singularidade estrutural dos compostos 46 e 209, com maior atividade anti-Leishmania (IC_{50})

Conclusões

O estudo indicou que os candidatos 46 e 209 apresentaram maior atividade contra a forma amastigota intracelular de *L. infantum*. Notavelmente, seus núcleos estruturais são pouco comuns nas publicações recentes (2022–2024), levantando questionamentos sobre os motivos de sua exploração limitada nesse período. Na ausência de razões que os

desqualificam, essas estruturas representam uma oportunidade promissora para futuros estudos de desenvolvimento de fármacos contra a leishmaniose.

Os autores declaram não haver conflito de interesses.

Agradecimentos

Ao meu orientador, Prof. Dr. Leonardo Luiz Gomes Ferreira, aos meus colaboradores, Dr. João Victor Silva-Silva e Prof. Dr. Adriano D. Andricopulo. À Universidade de São Paulo, Instituto de Física de São Carlos e Laboratório de Química Medicinal e Computacional. Ao apoio do Programa Unificado De Bolsas De Estudo Para Apoio À Formação De Estudantes De Graduação (PUB-USP)

Referências

- [1] SANDER, T., et al. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *Journal of Chemical Information and Modeling*, v.55, p. 460-473, jan. 2015
- [2] LIPINSKI, C. A, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, v.23, n. 1-3, p. 3-25, jan. 1997.

Computational Studies for the Identification of Bioactive Molecules for Neglected Tropical Diseases

Pablo Henrique Figueredo

João Victor Silva-Silva, Prof. Dr. Adriano D. Andricopulo

Prof. Dr. Leonardo Luiz Gomes Ferreira

São Carlos Institute of Physics / University of São Paulo

pablofigueredo260@usp.br

Objectives

(i) To build a database, from scientific literature, with compounds active against the intracellular amastigote form of *Leishmania infantum* (ii) To analyze the physicochemical properties of the compounds in the database (iii) To identify structure-activity relationships (SAR) to determine molecular descriptors and chemical substructures associated with the compounds activity (IC_{50}).

Materials and Methods

For the construction of the database, a bibliographic survey was conducted, between the years 2022-2024, on Google Scholar, using "Leishmania infantum antileishmanial activity amastigote" as keywords. For each compound of interest, information on its name, biological activity (IC_{50}), cytotoxicity (CC_{50}), and the DOI of the source article were recorded. Based on this data, the molecular structures were built in the MarvinSketch software to generate the SMILES codes. Subsequently, the physicochemical properties were calculated using the DataWarrior software [1], focusing on the descriptors of Lipinski's "Rule of Five" (molecular weight, cLogP, number of hydrogen bond donors and acceptors) [2], determining the number of violations for each compound.

From this information, a scaffold analysis was performed in DataWarrior to identify the most recurrent structural cores among the compounds that met the pre-established selection criteria: no violations of Lipinski's Rule and $IC_{50} \leq 10 \mu M$.

Results

From the bibliographic survey, a database containing 232 compounds with reported activity against *L. infantum* was built. The analysis of properties based on Lipinski's "Rule of Five" revealed a favorable physicochemical profile for the majority of the molecules. As illustrated in Figure 1, 135 compounds (approximately 58.2%) showed no violations, indicating good compatibility with orally administered drugs. When correlating the number of violations with biological potency, it is observed that a significant portion of the most active compounds (lower IC_{50}) is concentrated in the group with no violations. Figure 2 shows the scaffold analysis performed with the subset of the most promising compounds, selected for having no violations of Lipinski's Rule and an $IC_{50} \leq 10 \mu M$. The two candidates with the highest activity in the database, compound **46** ($IC_{50} = 0,45 \mu M$) and **209** ($IC_{50} = 0,105 \mu M$) feature unique molecular scaffolds, (1-[(benzotriazol-1-yl)methyl]diselanyl)methyl]benzotriazole and (2-phenylchromen-4-one),

respectively. This structural uniqueness is maintained even when considering the entire database.

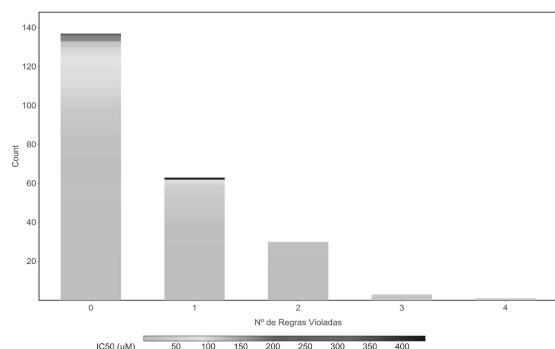


Figure 1: Analysis of anti-Leishmania activity (IC_{50}) as a function of the Lipinski's Rule violation profile.

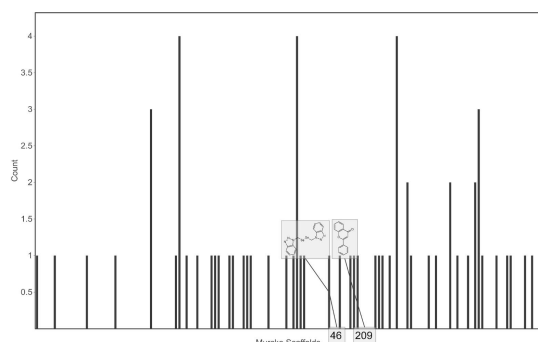


Figure 2: Scaffold analysis highlighting the structural uniqueness of compounds 46 and 209, with the highest anti-Leishmania activity (IC_{50})

Conclusions

The study indicated that candidates 46 and 209 showed the highest activity against the intracellular amastigote form of *L. infantum*. Notably, their structural cores are uncommon in recent publications (2022-2024), raising questions about the reasons for their limited exploration during this period. In the absence of disqualifying reasons, these structures represent a promising opportunity for future

drug development studies against leishmaniasis.

The authors declare no conflict of interest.

Acknowledgements

To my advisor, Prof. Dr. Leonardo Luiz Gomes Ferreira, and my collaborators, Dr. João Victor Silva-Silva, Prof. Dr. Adriano D. Andricopulo. To the University of São Paulo, the São Carlos Institute of Physics, and the Laboratory of Medicinal and Computational Chemistry. To the support from the Unified Scholarship Program for Supporting Undergraduate Student Education (PUB-USP).

References

- [1] SANDER, T., et al. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *Journal of Chemical Information and Modeling*, v.55, p. 460-473, jan. 2015
- [2] LIPINSKI, C. A, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, v.23, n. 1-3, p. 3-25, jan. 1997.