

**Universidade Federal de Pernambuco
Centro de Ciências Biológicas
Programa de Pós-Graduação em Genética**

Crhisllane Rafaele dos Santos Vasconcelos

***Avaliação in silico* do proteoma predito de *Leishmania sp.*
para seleção de alvos terapêuticos**

**Recife
2020**

Crhisllane Rafaele dos Santos Vasconcelos

**Avaliação *in silico* do proteoma predito de *Leishmania sp.*
para seleção de alvos terapêuticos**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco na área de concentração Biologia Molecular como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Orientador: Antonio Mauro Rezende

**Recife
2020**

Catálogo na Fonte:
Elaine C Barroso, CRB-4/1728

Vasconcelos, Chrisllane Rafael dos Santos
Avaliação *in silico* do proteoma de *Leishmania* spp. para seleção de alvos terapêuticos
/ Chrisllane Rafael dos Santos Vasconcelos – 2020.

123 f. : il., fig., tab.

Orientador: Antonio Mauro Rezende

Tese (doutorado) – Universidade Federal de Pernambuco. Centro
Biotecnologias. Programa de Pós-graduação em Genética, Recife, 2020.
Inclui referências e anexos

1. Leishmaniose 2. Expressão gênica 3. Drogas I. Rezende, Antonio Mauro
(orient.) II. Título

616.9364

CDD (22.ed)

UFPE/CB-2021-232

Crhisllane Rafaele dos Santos Vasconcelos

**Avaliação *in silico* do proteoma predito de *Leishmania sp.* para
seleção de alvos terapêuticos**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco na área de concentração Biologia Molecular como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Aprovado em 03/11/2020

Banca Examinadora:

Dr. Antonio Mauro Rezende
Instituto Aggeu Magalhães – Fundação Oswaldo Cruz/PE

Dr. Valdir de Queiroz Balbino
Universidade Federal de Pernambuco

Dr. Rafael de Freitas e Silva
Universidade de Pernambuco

Dr. Nilmar Silvio Moretti
Universidade Federal de São Paulo

Dr. Adriano Cappellazzo Coelho
Universidade de Campinas

AGRADECIMENTOS

Durante o percurso do doutorado percebi que esses 4 anos consistem mais em ter ciência do universo de conteúdo que não dominamos do que realmente dominar um conteúdo. Isso não significa que não nos tornamos especialistas em algo, mas que saímos como doutores cientes que pouco sabemos de outras áreas. No entanto, esse pouco do “universo” que agora conhecemos não é tão simples de ser conquistado, e por isso essa página está presente em tantas dissertações e teses, porque nunca chegamos aqui sozinhos.

Antes de qualquer outro agradecimento, e mesmo com todos os questionamentos que surgiram, agradeço a Deus. Não é fácil caminhar se não houver esperança e fé em um amanhã melhor.

Agradeço a meus pais Sônia e Davi, que me estimularam a criatividade, aguçaram-me a curiosidade, acreditaram em mim e me deram subsídios para continuar o percurso acadêmico. Sem vocês não conseguiria chegar aqui.

Agradeço a minha família, pelo amor e por entender minha ausência quando os dados e scripts me consumiam.

Agradeço a meu orientador e amigo Antonio Rezende, que não me deu as respostas óbvias, mas me instigava a pesquisar; que me ensinou com paciência; que me abriu os caminhos e me deu coragem. Tenho orgulho de ser sua aluna.

Agradeço a Filiphe Vilar, que esteve presente em todos os momentos, me ensinou a paz e a calma para continuar e quando os *bugs* surgiam me enviava as simples soluções.

Agradeço a meus tios José André, Isaias e Cícero, por me ensinarem que as noites também são feitas para estudar. Vocês são meu exemplo.

Agradeço a Rafaela Mikaella por surgir e abrir os braços quando mais precisei, por ser amiga e ter se tornado família; por ter sido força e paciência. Se caminhei a longa quarentena que perdurou o final dessa tese, com certeza foi porque você estava aqui. Estendo esse agradecimento a toda a sua família que tanto me apoiou.

Agradeço aos meus amigos: Michelle Barros, Rodrigo Loyo, Lays Trajano, Leylla Moreira e Kamila Galdêncio. Na verdade, deveria agradecer a Deus por vocês existirem. Obrigada pelo amor, carinho, tempo e conversas. Sem vocês essa caminhada não teria sido recheada de boas risadas.

Agradeço a João Pitta, por ser irmão, parceiro de “bancada”, amigo, vizinho, e

todas as outras boas coisas que você é. Minhas tardes e noites trabalhando não seriam tão lindas e com cheiro de café se você não estivesse aqui.

Agradeço a Beatriz Toscano e a Carol Luz por todas as conversas, pelo apoio e por ter dividido seus experimentos de bancada com a bioinformática ou simplesmente só ter invadido a sala para transbordá-la de carinho. Vocês fazem muita diferença naquela sala.

Agradeço ao Baixo Clero, em especial a Filipe Zimmer, Alexandre Freitas, Lays Ceschini e Elverson Melo, vocês são a melhor equipe de trabalho. Obrigada pelas risadas, por se empenharem em tirar as dúvidas uns dos outros, pelos churrascos e claro, por deixarem a sala com o cheiro de chimarrão.

Agradeço aos meus amigos: Marina, Arthur, Pamella, Letícia e Phellype, por todo apoio nessa caminhada, por dividirem a mesa de estudos quando virei noites com artigos, por terem sido ombros, braços e pernas. Em especial a Pamella, que comemorou mais do que ninguém o meu nome na lista de aprovados para o Doutorado.

Agradeço a Mayara Martins por ser amiga presente mesmo com a distância, por acalmar minha mente e coração quando os resultados não saiam ou as coisas desandaram.

Obrigada a todos os quais eu ajudei nos projetos e que também me ajudaram, os artigos listados no final dessa tese são uma grande prova que contribuição e amizade geram ciência.

Obrigada a banda *Whiskey in the Jar*, Antonio Rezende, Rafael Rosa, Rodrigo, Raissa, Rafael França, Luydson Vasconcelos e todos que sempre estavam nos ensaios, nunca imaginei que tocaria o que eu gosto com os amigos do trabalho no trabalho. Os ensaios eram os momentos que a mente entendia a paz no barulho.

Obrigada a todos que fazem a Instituto Aggeu Magalhães, eu não sei como explicar o quão bem eu me sinto trabalhando aqui.

Obrigada a todos os professores e equipe da secretaria que formam o PPGG e a Fiocruz-PE, em especial a Gabriel Wallau, por todo conhecimento passado e por todas as parcerias científicas.

Agradeço a toda a equipe do Scicast, em especial ao Fernando Malta, Juliana, Marcelo e Thais Boccia, por me levarem a pesquisar e aprender muito além da minha área e por toda a parceria de divulgação científica e amizade. Quero sempre fazer parte dessa família.

Por fim, mas não menos importante, agradeço a plataforma de bioinformática da Fiocruz-PE, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), por todos os subsídios fornecidos que resultaram nessa tese.

Se você não está presente nessa lista, mas deveria ter sido listado, eu te agradeço também por tudo. E com certeza você sabe o quanto sou esquecida, mas isso não significa que eu tenha menos gratidão por você.

Tenho a impressão de ter sido uma criança brincando à beira-mar, divertindo-me em descobrir uma pedrinha mais lisa ou uma concha mais bonita que as outras, enquanto o imenso oceano da verdade continua misterioso diante de meus olhos.

(Isaac Newton)

RESUMO

As leishmanioses são um grupo de doenças infecciosas negligenciadas, com aproximadamente 1,3 milhão de novos casos a cada ano, para os quais as terapias disponíveis apresentam sérias limitações. Portanto, é de extrema importância a aplicação de métodos eficazes e de baixo custo, capazes de selecionar os melhores alvos terapêuticos para acelerar o desenvolvimento de novas terapias contra essas doenças. Assim, propomos a utilização de métodos computacionais integrados capazes de avaliar a drogabilidade dos proteomas preditos de *Leishmania braziliensis* e *Leishmania infantum*, espécies responsáveis pelas diferentes manifestações clínicas das leishmanioses no Brasil. Os proteomas foram avaliados com base em seus contextos estruturais, químicos e funcionais, aplicando métodos que integram dados sobre função molecular, processos biológicos, localização subcelular, sítios de ligação de drogas, drogabilidade e expressão gênica. Esses dados foram comparados com aqueles extraídos de alvos de fármacos já conhecidos, o que possibilitou avaliar os proteomas de *Leishmania* quanto à sua relevância biológica e tratabilidade. Por meio dessa metodologia, identificamos mais de 100 proteínas de cada espécie de *Leishmania* com características de drogabilidade e potencial interação com os fármacos disponíveis. Dentre essas, 31 e 37 proteínas de *L. braziliensis* e *L. infantum*, respectivamente, nunca foram testadas como alvos de drogas e têm demonstrado evidências de expressão gênica em estágio evolutivo de interesse farmacológico. Alguns desses alvos em *Leishmania* apresentaram similaridade de alinhamento inferior a 50% quando comparados ao proteoma humano, tornando essas proteínas farmacologicamente atrativas, por apresentarem risco reduzido de efeitos colaterais. A metodologia utilizada neste estudo também permitiu avaliar as oportunidades de reaproveitamento de compostos como fármacos anti-leishmaniose, inferindo potencial interação entre proteínas de *Leishmania* e aproximadamente 1000 compostos, dos quais apenas 15 já foram testados como tratamento para leishmaniose. Além disso, foi disponibilizada uma lista de potenciais alvos a serem testados com drogas descritas no BindingDB, como a potencial interação das proteínas DEAD box RNA helicase, TRYR e PEPCK com o composto Estaurosporina.

Palavras-chave: *Leishmania*; Leishmaniose; Reposicionamento de drogas; Proteínas drogáveis; Alvos biológicos; Desenvolvimento de drogas.

ABSTRACT

Leishmaniasis is a group of neglected infectious diseases, with approximately 1.3 million new cases each year, for which the available therapies have serious limitations. Therefore, it is extremely important to apply efficient and low-cost methods capable of selecting the best therapeutic targets to speed up the development of new therapies against those diseases. Thus, we propose the use of integrated computational methods capable of evaluating the druggability of the predicted proteomes of *Leishmania braziliensis* and *Leishmania infantum*, species responsible for the different clinical manifestations of leishmaniasis in Brazil. The protein members of those proteomes were assessed based on their structural, chemical, and functional contexts applying methods that integrate data on molecular function, biological processes, subcellular localization, drug binding sites, druggability, and gene expression. These data were compared to those extracted from already known drug targets, which made it possible to evaluate *Leishmania* proteomes for their biological relevance and treatability. Through this methodology, we identified more than 100 proteins of each *Leishmania* species with druggability characteristics, and potential interaction with available drugs. Among those, 31 and 37 proteins of *L. braziliensis* and *L. infantum*, respectively, have never been tested as drug targets, and they have shown evidence of gene expression in the evolutionary stage of pharmacological interest. Also, some of those *Leishmania* targets showed an alignment similarity of less than 50% when compared to the human proteome, making these proteins pharmacologically attractive, as they present a reduced risk of side effects. The methodology used in this study also allowed the evaluation of opportunities for the repurposing of compounds as anti-leishmaniasis drugs, inferring potential interaction between *Leishmania* proteins and approximately 1000 compounds, of which only 15 have already been tested as a treatment for leishmaniasis. Besides, a list of potential targets to be tested using drugs described at BindingDB, such as the potential interaction of the DEAD box RNA helicase, TRYR, and PEPCK proteins with the Staurosporine compound, was made available to the public.

Palavras-chave: *Leishmania*; Leishmaniasis; Drug repurposing; Druggable proteins; Biological targets; Drug discovery.

LISTA DE ILUSTRAÇÕES

Figura 1 – Classificação das espécies de <i>Leishmania</i>	20
Figura 2 – Ciclo Biológico de <i>Leishmania spp</i>	22
Figura 3 – Investimento no desenvolvimento de medicamentos e números de Drogas aprovadas entre 1995 a 2015.....	27
Figura 4 – Esquema de categorias de nós em uma rede.....	35
Figura 5 – Modelo hipotético da integração da rede a dados de ontologia gênica. .	36
Figura 6 – Comparação de medidas de similaridade entre estruturas químicas.	39
Figura 7 – Métodos de detecção de sítio de ligação.	46
Figura 8 – Metodologia de predição de sítios de ligação baseada em grade.....	47
Figura 9 – Representação da metodologia de predição de sítios de ligação baseada em esferas.....	48
Figura 10 – Construção da forma alfa aplicada à busca por sítio de interação.	49
Fig. 1 – Performance evaluation through the AUC values obtained during the 100 training/tests of machine learning models used to predict interaction between proteins.....	59
Fig. 2 – Protein-Protein Interaction Network using Cytoscape 3.5.1.	60
Fig. 3 – Interaction Protein Networks predicted through structural information adding the networks predicted by Rezende et al. [38].....	61
Fig. 4 – Integration of the sub networks formed by the 20 proteins with the highest Degree of connectivity in predicted protein interaction networks of <i>L. braziliensis</i> (a) and <i>L. infantum</i> (b)	62
Fig. 5 – Integration of the sub networks formed by the 20 proteins with the highest value of Bottlenecks in predicted protein interaction networks of <i>L. braziliensis</i> (a) and <i>L. infantum</i> (b)	62
Fig. 6 – Integration of the sub networks formed by the 20 proteins with the highest value of Betweenness Centrality in predicted protein interaction networks of <i>L. braziliensis</i> (a) and <i>L. infantum</i> (b)	63
Figure 1 – Diagram with results of the multidisciplinary approach to select targeting drugs from Leishmania species proteome.....	73
Figure 2 – Workflow presenting the scheme and criteria used for the selection of	

druggable proteins of pharmacological interest in <i>Leishmania sp.</i> proteomes	74
Figure 3 – Subnetwork of protein interaction for <i>Leishmania sp.</i> focused on potential drug targets and hub proteins.....	75
Figure 4 – Degree centrality, similarity, and coverage against human proteome of potential target proteins of <i>Leishmania sp.</i>	76
Figure 5 – Example of predicted structures and druggable binding sites for potential protein targets in proteomes of <i>Leishmania sp.</i>	77

LISTA DE TABELAS

Table 1 –	Total protein structure predicted by each program.	58
Table 2 –	Total protein per tool with lower free energy structure and higher percentage of torsion angles in the most favorable region of the ramachandran plot	58
Table 3 –	Total proteins in each cell compartment predicted by the Wolfpsort tool....	59
Table 4 –	Interactions described as possible by each tool and consensus.....	60
	Table 5: Evaluation of the topological characteristics of protein.....	60
Table 6 –	Evaluation of the topological characteristics of the protein interaction networks predicted through structural information and merged to the networks predicted by Rezende et al. [38].....	61
Table 1 –	Total druggable binding sites.....	72
Table 2 –	Potential drug targets with relative PIN essentiality and <50% similarity to the human proteome.	76

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

Item	Definição
3D	<i>Three-dimensional</i> / Tridimensional
AnB	Anfotericina B
AnB-D	Anfotericina B formulação com Desoxicolato
AnB-L	Anfotericina B formulação lipídica
AUC	<i>Area Under the Curve</i> / Area Sob a Curva
BATS	<i>Blast Automatic Targeting for Structures</i>
BC	<i>Betweenness Centrality</i> / Centralidade de Intermediação
BLAST	<i>Basic Local Alignment Search Tool</i>
BN	<i>Bottleneck Centrality</i> / Centralidade de Gargalos
CC	<i>Clustering Coefficient</i> / Coeficiente de Clusterização
DC	<i>Degree Centrality</i> / Grau de centralidade
DOPE	<i>Discrete Optimized Protein Energy</i>
GBM	<i>Gradient Boosting Method</i>
GO	<i>Gene Ontology</i> / Ontologia Gênica
LM	<i>Linear Regression Model</i> / Modelo de Regressão Linear
MSP	<i>Mean Shortest Path</i> / Média do Caminho Mais Curto
MN	<i>Merged Network</i> / Rede agrupada
NB	<i>Native Bayer</i>
NN	<i>Neural Network</i> / Rede Neural
NPTIM	<i>Predicted Networks Through Interolog Mapping</i> / Redes Preditas por mapeamento de interações ortólogas
NPTSI	<i>Predicted Networks Through Structural Information</i> / Redes Preditas Utilizando Informação Estrutural
PDB	<i>Protein Data Bank</i>
Phyre2	<i>Protein Homology/Analogy Recognition Engine Version 2.0</i>
PIN	<i>Protein Interaction Network</i> / Redes de Interações Proteicas
PPI	<i>Protein-Protein Interaction</i> / Interação entre Proteínas
RF	<i>Random Forest</i> / Floresta Aleatória
RMSD	<i>Root Mean Square Deviation</i> / Raiz Quadrada do Desvio Médio
ROC	<i>Receiver Operating Characteristic</i> / Característica de Operação do Receptor
SRA	<i>Sequence Read Archive</i>

SVM

Support Vector Machine / Máquina de Vetor de Suporte

WHO

World Health Organization / Organização Mundial de Saúde

SUMÁRIO

1	INTRODUÇÃO	16
2	REVISÃO DA LITERATURA	18
2.1	LEISHMANIOSE.....	18
2.1.1	Ciclo Biológico.....	21
2.1.2	Tratamentos disponíveis para Leishmaniose	23
2.2	DESENVOLVIMENTO DE FÁRMACOS.....	26
2.3	SELEÇÃO DE POTENCIAIS ALVOS PARA FÁRMACOS	32
2.3.1	Redes de interação proteica.....	33
2.3.2	Dados de similaridade.....	37
2.3.2.1	Similaridade entre drogas.....	38
2.3.2.2	Similaridade entre alvos de drogas.....	39
2.3.3	Drogabilidade no contexto estrutural	44
2.3.4	Análise de dados de expressão gênica	50
2.4	INTEGRAÇÃO DOS DADOS	51
3	OBJETIVOS	52
3.1	OBJETIVO GERAL.....	52
3.2	OBJETIVOS ESPECÍFICOS	52
4	CAPÍTULO I - BUILDING PROTEIN-PROTEIN INTERACTION NETWORKS FOR LEISHMANIA SPECIES THROUGH PROTEIN STRUCTURAL INFORMATION	53
5	CAPÍTULO II – SYSTEMATIC <i>IN SILICO</i> EVALUATION OF <i>LEISHMANIA</i> <i>SPP.</i> PROTEOMES FOR DRUG DISCOVERY	67
6	DISCUSSÃO GERAL.....	84
7	CONCLUSÃO.....	86
	REFERÊNCIAS.....	87
	ANEXOS	97

1 INTRODUÇÃO

As leishmanioses abrangem um conjunto de doenças causadas por protozoários do gênero *Leishmania* com relatos de ocorrência em 98 países e incidência de aproximadamente 1,3 milhões de novos casos registrados anualmente. Apesar dos esforços para o desenvolvimento de vacinas e novos fármacos contra as leishmanioses, nenhuma vacina eficaz está disponível. As opções farmacológicas possuem sérias limitações à sua utilização, como elevada toxicidade, o desenvolvimento de resistência por parte dos parasitos e custo incompatível com países endêmicos subdesenvolvidos. Assim, selecionar novos alvos promissores é uma estratégia fundamental para o desenvolvimento de drogas alternativas e vacinas. A descoberta destes alvos de maneira eficiente é importante para acelerar esse desenvolvimento e priorizar potenciais candidatos além dos classicamente conhecidos.

Nesse sentido, os avanços tecnológicos e a crescente disponibilidade de dados genômicos têm ajudado na compreensão biológica dos parasitas envolvidos nas leishmanioses. Porém, para que se possa selecionar alvos proteicos para fins terapêuticos, é necessário avaliar diversas características do possível candidato, como sua essencialidade na fisiologia do organismo, seu índice de drogabilidade, que está relacionado com a possibilidade do alvo ser modulado por um composto farmacológico, e estabelecer sua função biológica. Realizar isto de maneira experimental para um proteoma demanda um grande esforço laboral e um alto custo, tornando inviável essa caracterização em larga escala. Desta maneira, abordagens computacionais capazes de assimilar e integrar dados, juntamente com o conhecimento biológico acerca da doença e dos agentes etiológicos, apresentam-se como uma estratégia para reduzir custos no processo de descoberta de novas linhas de tratamento e acelerar as fases iniciais do desenvolvimento de drogas.

Uma abordagem computacional global, sistemática, objetiva e multidisciplinar pode permitir uma avaliação eficaz e orientada de dados, capaz de, utilizando uma grande lista de alvos, selecionar aqueles biologicamente e farmacologicamente favoráveis para desenvolvimento de drogas. A utilização desses critérios para seleção

de alvos drogáveis já foi empregada anteriormente na busca de novas linhas de tratamento oncológico e parasitológico, permitindo a exploração de novos alvos drogáveis e inferindo o redirecionamento de compostos já disponíveis.

Dessa maneira, o presente estudo apresenta uma análise multiômica em larga escala que integra o espaço químico e biológico do proteoma de *Leishmania braziliensis* e *infantum*, através da análise computacional de dados de drogabilidade, função molecular, expressão gênica, localização subcelular e similaridade de estrutura tridimensional com o objetivo de priorizar potenciais alvos terapêuticos para análises experimentais. Além disso, propomos moléculas ativas com possibilidade de redirecionamento e interação com os potenciais alvos, possibilitando uma redução de prazo e custo no desenvolvimento de drogas com potencial leishmanicida.

2 REVISÃO DA LITERATURA

2.1 LEISHMANIOSE

As Leishmanioses são um conjunto de doenças causadas por protozoários flagelados do gênero *Leishmania* com capacidade de infectar mamíferos desde a era Mesozoica (AKHOUNDI et al., 2016). Atualmente, há estimativas de 1,3 milhões de casos anuais e 350 milhões de pessoas com risco de contrair a doença. De acordo com a Organização Mundial de Saúde (OMS), as leishmanioses permanecem como um problema de saúde pública mundial, sendo a causa de aproximadamente 20.000 óbitos por ano (GRADONI, 2018; WORLD HEALTH ORGANIZATION/COUNTRY OFFICE FOR INDIA, 2017).

Apesar da ocorrência em 98 países, apenas sete e 12 países compreendem 90% dos casos de leishmaniose visceral e leishmaniose cutânea, respectivamente (WHO/DEPARTMENT OF CONTROL OF NEGLECTED TROPICAL DISEASES, 2017). O Brasil encontra-se entre os países com a maior quantidade de casos para ambas formas de leishmaniose, apresentando uma média anual de 26.965 casos de Leishmaniose cutânea em todas as unidades federadas e 3.565 casos de Leishmaniose visceral em 21 unidades federais (BRASIL et al., 2019; PAN AMERICAN HEALTH ORGANIZATION, 2018).

Os parasitos causadores das Leishmanioses são transmitidos durante o repasto sanguíneo de espécies de flebotomíneos dos gêneros *Phlebotomus* e *Lutzomyia* (MAROLI et al., 2013). Estes vetores possuem uma ampla gama de hospedeiros, incluindo canídeos, roedores e humanos, com padrão epidemiológico predominantemente zoonótico de transmissão silvestre, acometendo humanos em ambientes florestais. No entanto, é possível observar alterações na sua incidência, indicando a evolução do seu ciclo de transmissão para zonas rurais desmatadas e regiões periurbanas (STEVERDING, 2017).

Existem aproximadamente 54 espécies do gênero *Leishmania*, sendo pelo menos 21 destas patogênicas para humanos (Figura 1) (AKHOUNDI et al., 2016, 2017). No Brasil, 7 espécies de *Leishmania* foram identificadas, sendo as principais

espécies a *Leishmania (Viannia) braziliensis*, *Leishmania (Viannia) guyanensis*, *Leishmania (Leishmania) amazonensis* e *Leishmania (Leishmania) infantum* (BRASIL et al., 2019). Tais espécies se apresentam como patógenos intracelulares obrigatórios com a capacidade de invadir células hospedeiras fagocíticas. Embora sejam morfolologicamente semelhantes, são capazes de causar formas clínicas diferentes da doença, a leishmaniose cutânea e a leishmaniose visceral, dependendo da célula fagocitária infectada, da espécie, da cepa, do grau de virulência e do estado imunológico do hospedeiro (AKHOUNDI et al., 2016; GUPTA; OGHUMU; SATOSKAR, 2013).

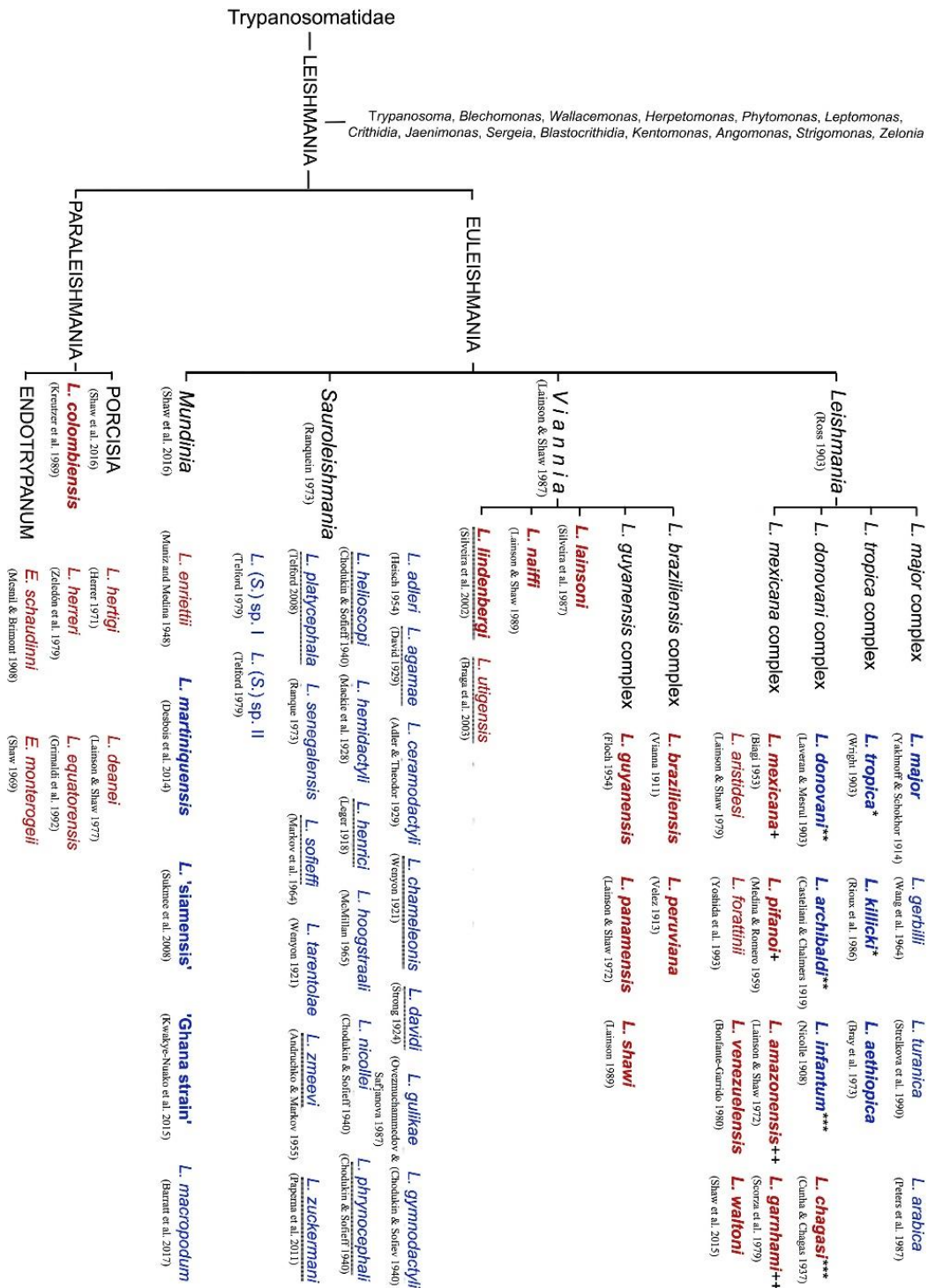


Figura 1: Classificação das espécies de Leishmania. Os sinais (*) e (+) representam sinônimo e sua quantidade representa qual nome de espécie é sinônimo para qual espécie. Sublinhado representa espécies sem classificação final. As aspas (") representam nomes sem descrição formal. As espécies com nome em **negrito** são patogênicas para humanos. A espécies destacadas em azul e vermelho são pertencentes ao Velho e Novo Mundo, respectivamente. As espécies *L. 'siamensis'* e *L. martiniquensis* estão destacadas em azul, mas também foram encontradas no Novo Mundo (AKHOUNDI et al., 2017).

A leishmaniose cutânea é a forma mais comum da doença, com

aproximadamente 1 milhão de casos por ano. Esta, pode se apresentar de três formas distintas: a leishmaniose cutânea localizada, caracterizada por úlceras e lesões na pele; a leishmaniose cutânea difusa, que se apresenta na forma de múltiplos nódulos disseminados pelo corpo sem ulceração; e a leishmaniose muco-cutânea, caracterizada pelo acometimento com ulcerações de membranas mucosas localizadas no nariz, boca e garganta (BRASIL et al., 2019). A leishmaniose muco-cutânea é uma manifestação da leishmaniose cutânea restrita à América Latina, com apresentação clínica marcada por uma extensa desfiguração facial devido a resposta imune do hospedeiro que pode provocar a destruição do septo nasal, lábios e palato. No Brasil, a *L. braziliensis* é o principal agente etiológico capaz de causar as três formas da leishmaniose cutânea (ANVERSA et al., 2018).

A leishmaniose visceral, também conhecida como calazar, é a forma mais grave da leishmaniose, com aproximadamente 300 mil casos relatados por ano e a capacidade de evoluir para óbito em 95% dos casos não tratados (WORLD HEALTH ORGANIZATION/COUNTRY OFFICE FOR INDIA, 2017). É caracterizada clinicamente por febre, hepatoesplenomegalia, anemia, icterícia, entre outras manifestações. No Brasil, essa forma clínica apresenta uma letalidade média de 6,9% e possui como principal agente etiológico a espécie *L. infantum* (syn *chagasi*) (ANVERSA et al., 2018; BRASIL et al., 2019).

2.1.1 Ciclo Biológico

As espécies de *Leishmania* são parasitos digenéticos com um ciclo biológico complexo (Figura 2A), logo seus ciclos de vida envolvem hospedeiros vertebrados e invertebrados, além de dois estágios morfológicos principais: o estágio promastigota (Figura 2B), forma flagelada, extracelular, encontrada no vetor; e o estágio amastigota (Figura 2C), forma aflagelada, intracelular, encontrada nas células de hospedeiros mamíferos (BATES; ROGERS, 2004).

A infecção primária dos flebotomíneos se dá a partir do repasto sanguíneo em hospedeiros infectados, onde o vetor ingere macrófagos parasitados pela forma amastigota de *Leishmania* spp. (Figura 2A,1-3). No intestino do vetor, em um intervalo de 24 a 48 horas, o parasita sofre diferenciação para a forma promastigota procíclica

e passa por um processo de divisão celular (Figura 2A,4-5). Após esta etapa, os parasitos se diferenciam em formas infecciosas (promastigotas metacíclicas) e migram para a válvula estomodeal do vetor (Figura 2A,6-7), completando seu ciclo biológico no próximo repasto sanguíneo, onde as formas promastigotas são inoculadas no hospedeiro através da probóscide do flebótomo, que pode permanecer infectado até o final do seu ciclo de vida (Figura 2A, 8) (BATES; ROGERS, 2004).

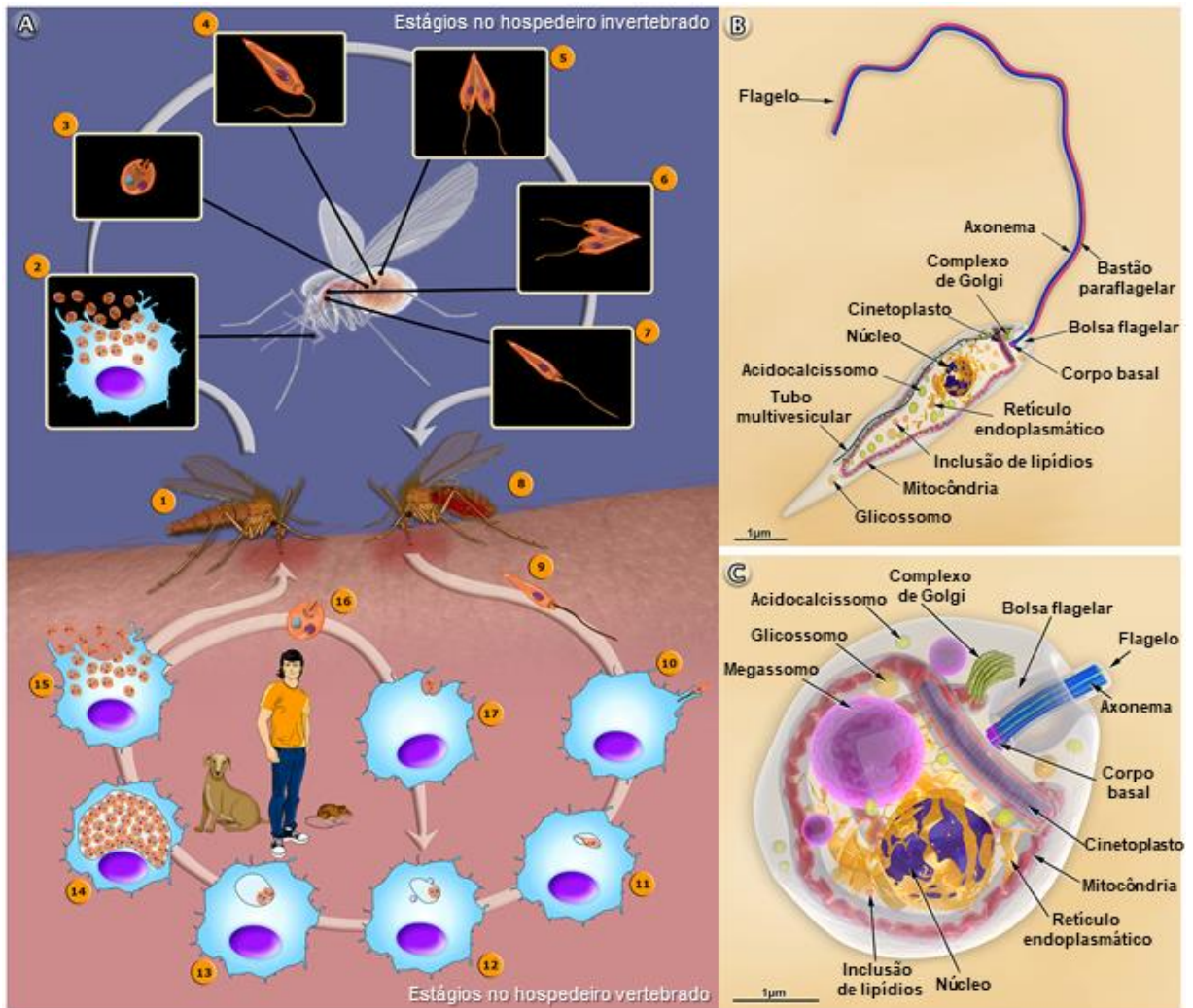


Figura 2: Ciclo Biológico de *Leishmania* spp. (A) Estágio em hospedeiro invertebrado (1) Repasto sanguíneo em hospedeiro infectado. (2) Ingestão de macrófagos infectados com formas amastigotas. (3) Corpo ovoide da forma amastigota. (4) Transformação da forma amastigotas em Promastigotas procíclicas. (5) Promastigotas procíclicas em processo de divisão binária no intestino do vetor. (6) Promastigotas migram em direção à válvula estomodeal no intestino anterior e reiniciam a divisão celular. (7) Processo de transformação para Promastigotas metacíclicas infecciosas. (8) Inoculação de Promastigotas metacíclicas em um novo hospedeiro mamífero via regurgitação durante o repasto sanguíneo. (9) Corpo alongado das Promastigotas metacíclicas que são introduzidas no hospedeiro vertebrado. (10) Fagocitose de Promastigotas metacíclicos por macrófagos. (11) Processo de transformação de Promastigotas metacíclicos em amastigotas. (12) Amastigotas se ligam à membrana do vacúolo parasitóforo. (13) Amastigotas em processo de divisão celular. (14) Intensa multiplicação de

amastigotas. (15) Rompimento celular causado pela alta carga parasitária na célula. (16) Forma amastigota livre. (17) Infecção de um novo macrófago. Representações esquemáticas das organelas encontradas nas fases Promastigota (B) e Amastigota (C) (TEIXEIRA et al., 2013).

As formas promastigotas inoculadas no hospedeiro são fagocitadas por células do sistema mononuclear fagocitário, onde passam novamente por um processo de diferenciação para forma amastigota e multiplicam-se por divisão binária, aumentando a carga parasitária e levando ao rompimento da célula infectada. Os parasitos são então liberados e passam a infectar outras células de forma contínua (BATES; ROGERS, 2004).

Após a infecção, entre os fatores que decidirão o estabelecimento da doença e suas diferentes formas clínicas ou a eliminação do parasito estão: a resposta imune do hospedeiro e a capacidade da cepa em manter o estado infeccioso por evasão imunitária. Podendo, a resposta imune do hospedeiro ser modulada por moléculas produzidas pelo agente infeccioso e por moléculas presentes na saliva do vetor (LESTINOVA et al., 2017). Dessa forma, compreender o balanço entre esses fatores pode auxiliar no aperfeiçoamento de novas estratégias para o desenvolvimento de uma intervenção terapêutica eficaz (GUPTA; OGHUMU; SATOSKAR, 2013).

2.1.2 Tratamentos disponíveis para Leishmaniose

A abordagem a ser empregada no tratamento das leishmanioses, preconizada pela OMS, varia em função da região geográfica, uma vez que as falhas no tratamento estão relacionadas com a genética, estado nutricional do hospedeiro e com a resistência às drogas desenvolvida pelas cepas e espécies específicas da região. Como exemplo temos a espécie *L. braziliensis* de isolados da América Latina que apresentam uma maior tolerância à droga miltefosina quando comparada às espécies *L. lainsoni* e *L. donovani* de isolados do subcontinente Indiano. Da mesma forma, podemos observar uma divergência de resposta ao tratamento com as cepas de *L. guyanensis* e *L. braziliensis* do Peru e do Brasil, respectivamente. Pacientes do Peru infectados com *L. guyanensis* respondem de maneira mais eficiente ao tratamento com antimoniais pentavalentes quando comparados a pacientes infectados com *L. braziliensis*. No entanto, o oposto é observado quando são avaliados pacientes infectados no Brasil (VANAERSCHOT et al., 2014).

Desta forma, as diferentes abordagens terapêuticas preconizadas envolvem as drogas descritas a seguir, usadas de forma isolada ou conjugada.

Antimoniais Pentavalentes: Os antimoniais, na sua forma trivalente, foram relatados pela primeira vez em 1915 como eficazes no tratamento das leishmanioses, sendo substituídos, devido à sua alta toxicidade, pela forma pentavalente a partir de 1937. Tais compostos permanecem até hoje como principal tratamento antileishmanial (BRASIL et al., 2019; NAGLE et al., 2014). Atualmente duas formulações de antimoniais pentavalentes estão disponíveis, o antimoniato de meglumina e estibogluconato de sódio, que podem ser administrados de maneira intramuscular, intravenosa ou intralesional. Apesar de apresentarem uma toxicidade inferior à dos trivalentes, elas possuem diversos efeitos colaterais relacionados a seu conteúdo antimonial, que vão desde vômitos e náuseas até alterações eletrocardiográficas e morte súbita (SOUZA DO LAGO et al., 2018; WORLD HEALTH ORGANIZATION, 2010). Além disso, a utilização destes compostos por mais de 70 anos como droga de primeira linha tem resultado no declínio de sua eficácia clínica, chegando a apresentar resistência ao tratamento em mais de 60% dos casos em regiões da Índia (NAGLE et al., 2014).

Anfotericina B: A anfotericina B (AnB) é um antibiótico poliênico capaz de aumentar a permeabilidade da membrana através de sua ação sobre o ergosterol presente na membrana plasmática de espécies de *Leishmania* (MATSUMORI et al., 2009). Sua atividade antileishmanial foi identificada em 1960 e desde então é utilizado como droga de segunda linha em formulações conjugadas com desoxicolato (AnB-D) ou com lipídios (AnB-L). Sua formulação AnB-D apresenta diversos efeitos adversos, que incluem nefrotoxicidade, hipocalcemia e miocardite, necessitando de hospitalização para monitoramento diário durante o tratamento (BRASIL et al., 2019; SUNDAR; SINGH, 2016; WORLD HEALTH ORGANIZATION, 2010). Suas formulações AnB-L foram desenvolvidas para melhorar a tolerabilidade por parte do paciente. Apesar de menos tóxica apresentando baixa nefrotoxicidade transitória e trombocitopenia, o paciente em tratamento também necessita de monitoramento diário. No entanto, o maior fator limitante da implementação desta terapêutica é o seu alto custo, tornando-se pouco acessível para países em desenvolvimento

(GHORBANI; FARHOUDI, 2018).

Os diversos casos refratários aos antimoniais tem ocasionado a substituição destes por formulações de AnB como droga de primeira linha. Contudo, casos de resistência a AnB já vem sendo relatados em diversos países, trazendo atenção sobre a ameaça da diminuição de eficácia desta linha de tratamento (MORIZOT et al., 2016; PURKAIT et al., 2012; SRIVASTAVA et al., 2011).

Paromomicina: A paromomicina é um antibiótico aminoglicosídeo capaz de inibir a proteossíntese através da interação com o RNA ribossomal (VICENS; WESTHOF, 2001). Sua eficácia no tratamento da leishmaniose cutânea foi identificada em 1966, e contra leishmaniose visceral em 1990 tendo como principal efeito adverso a dor local causada na aplicação. Efeitos colaterais como ototoxicidade e nefrotoxicidade são menos frequentes, mas podem ser observados com ocorrência em 2% e 6% dos pacientes tratados, respectivamente (NAGLE et al., 2014; SUNDAR et al., 2007; WORLD HEALTH ORGANIZATION, 2010). Apesar de apresentar um baixo custo de produção e uma toxicidade inferior aos tratamentos convencionais (anfotericina B e antimoniais pentavalentes), sua taxa de cura é frequentemente inferior, chegando a 14,3% de cura no Sudão (HAILU et al., 2010), fato que resulta na pouca utilização desse tratamento (NAGLE et al., 2014).

Pentamidina: A pentamidina é um antiprotozoário usado principalmente no tratamento da pneumonia por *Pneumocystis jiroveci*, com ação antileishmanial relatada em 1949 (HAZARIKA, 1949). Atualmente ela é usada como droga de segunda linha no tratamento das leishmanioses. Com um menor intervalo de tempo necessário para tratamento completo, a pentamidina se apresentou como solução terapêutica frente aos antimoniais. No entanto, a alta toxicidade é o principal fator limitante na implementação desta linha de tratamento, uma vez que possui a capacidade de causar efeitos adversos irreversíveis, como o diabetes mellitus insulino dependente (BELEHU; NAAFS, 1982; NAGLE et al., 2014; WORLD HEALTH ORGANIZATION, 2010).

Miltefosina: A miltefosina é um alquil fosfolipídio desenvolvido originalmente na década de 1980 como agente anticancerígeno. Sua capacidade antileishmanial foi identificada em 1990, sendo aprovado em 2002 como primeiro tratamento oral para

leishmaniose, usado como droga de terceira linha (CROFT; SNOWDON; YARDLEY, 1996; SUNDAR et al., 2002; WORLD HEALTH ORGANIZATION, 2010). Com efeitos adversos mais raros que os antimoniais, indo de reações gastrointestinais a ocasional hepatotoxicidade e nefrotoxicidade, a miltefosina se apresentou como um avanço na terapêutica disponível para leishmaniose. No entanto, após uma década de seu uso é possível perceber uma redução na sua eficácia, além de relatos de resistência natural a esse fármaco em cepas de *L. infantum* (CARNIELLI et al., 2019; NAGLE et al., 2014; RIJAL et al., 2013; SUNDAR et al., 2012). Outras problemáticas também dificultam a implementação da miltefosina em regiões endêmicas, como seu alto custo e baixa disponibilidade (SUNYOTO; POTET; BOELAERT, 2018).

O uso combinado das drogas disponíveis também é utilizado como estratégia para aumentar a eficácia do tratamento das leishmanioses. Contudo, os efeitos colaterais inevitáveis têm levado ao abandono do tratamento por parte dos pacientes e ao uso de maneira indevida das drogas, ocasionando assim o surgimento de mecanismos de resistência nos parasitos (GHORBANI; FARHOUDI, 2018). Desta forma, as linhas de tratamento disponíveis permanecem com graves desvantagens, sendo necessário estudos capazes de acelerar o desenvolvimento de formas terapêuticas mais eficazes e seguras.

2.2 DESENVOLVIMENTO DE FÁRMACOS

No início do século XX as drogas eram principalmente derivadas de plantas e não passavam por um processo de desenvolvimento com ensaios capazes de gerar informações acerca de sua toxicidade e eficácia (ANDRADE et al., 2016). Só em 1906, com o surgimento da *Food and Drug Administration* (FDA) foi estabelecido regulamentos para o desenvolvimento de drogas. No entanto, as normas estabelecidas envolviam apenas as características físico-químicas e pureza dos medicamentos, o que resultou em diversas drogas contaminadas com metais pesados e na intoxicação de diversas pessoas (ANDRADE et al., 2016; GREAVES; WILLIAMS; EVE, 2004).

Somente após 1937, com a morte de mais de 70 pessoas contaminadas com

dietileno glicol usado na produção do Elixir de Sulfanilamide-Massengill, foram estabelecidos princípios para realização de estudos clínicos que foram posteriormente incorporados ao Código de Nuremberg (ANDRADE et al., 2016; GEILING; CANNON, 1938). Com o marco do Código de Nuremberg emergiram os primeiros grupos de pesquisas multidisciplinares que serviram como base para o surgimento das grandes indústrias farmacêuticas (ANDRADE et al., 2016).

Dessa forma, em 1960 o processo de desenvolvimento de medicamentos foi estabelecido e permanece praticamente inalterado. Mesmo com as modificações que os demais setores de indústrias baseadas em pesquisa estabeleceram em seus processos de pesquisa e desenvolvimento, o setor farmacêutico continua a aplicar o mesmo processo de desenvolvimento de fármacos, o que tem resultado em um baixo número de drogas aprovadas e em um alto custo aplicado, como veremos a seguir (Figura 3) (KAITIN, 2010).

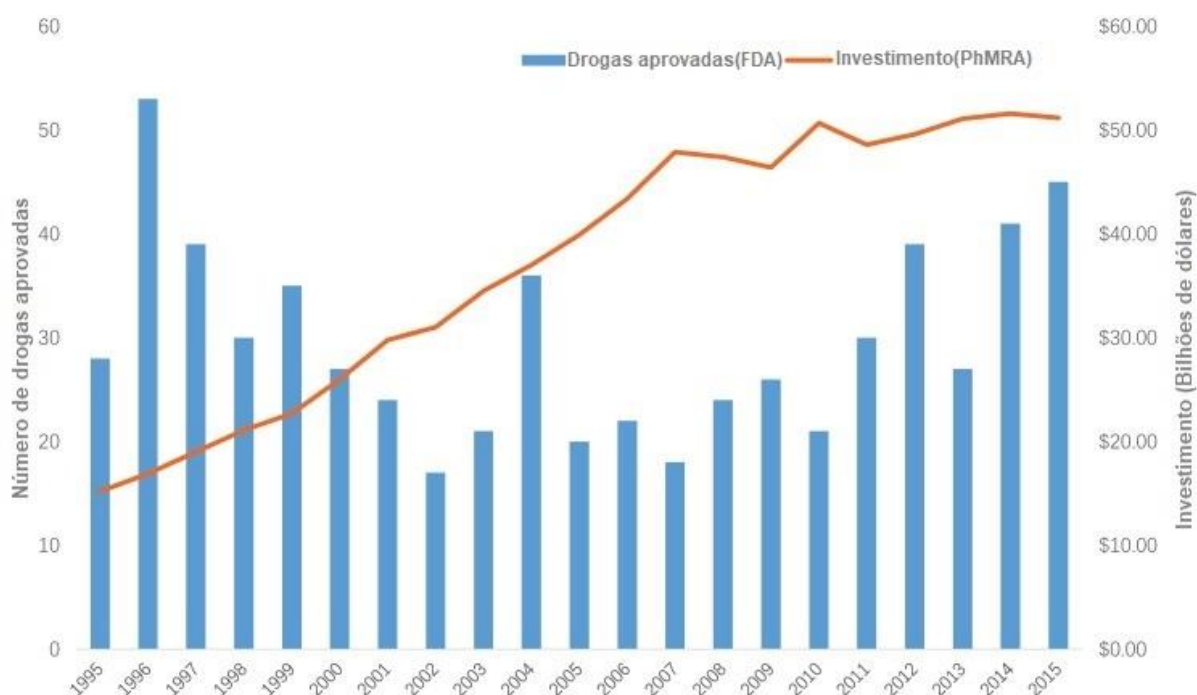


Figura 3: Investimento no desenvolvimento de medicamentos e números de Drogas aprovadas entre 1995 a 2015. O investimento no desenvolvimento de medicamentos por empresas membros da Pharmaceutical Research and Manufacturers of America (PhRMA) está representado com a linha laranja e no eixo vertical direito, enquanto o número de medicamentos aprovados pelo FDA está representado pelas barras em azul e eixo vertical esquerdo.

O processo de descoberta de um novo medicamento pode ser dividido em

farmacologia clássica (direta) ou farmacologia reversa dependendo do modelo utilizado para identificar potenciais substâncias (SAMS-DODD, 2005; SWINNEY, 2013). A farmacologia clássica é baseada na triagem fenotípica onde as substâncias são classificadas de acordo com os efeitos fisiológicos produzidos em uma determinada doença. Essa abordagem permite identificar moléculas capazes de alterar o fenótipo de uma doença por meio da interação com um ou vários alvos indefinidos (SWINNEY, 2013).

Na farmacologia reversa, O processo de desenvolvimento de um novo fármaco tem como ponto chave a descoberta de um alvo muitas vezes baseado em evidências biológicas, seja por sua essencialidade na estabilidade do patógeno ou na manutenção da doença. Isto é, o início do processo visa encontrar um alvo, sendo este muitas vezes uma proteína, com uma ação biológica específica que possa apresentar utilidade terapêutica (KNOWLES; GROMO, 2003; SCHENONE et al., 2013). Esta etapa permite posteriormente a seleção de compostos capazes de modular a função da molécula biológica alvo de maneira desejada. Entre a seleção e sua utilização na linha de tratamento de uma doença, os compostos candidatos serão submetidos a testes químicos, biológicos, toxicológicos e clínicos. Estes testes são fundamentais para garantir a segurança de um fármaco e são responsáveis por grande fração do custo de desenvolvimento de uma droga (BASAK, 2012).

No entanto, mesmo com a aplicação de ambas metodologias, todo este processo possui um baixo desempenho, uma vez que a taxa de sucesso para aprovação clínica de um novo composto é inferior a 19% (DIMASI et al., 2010; HUSSEIN et al., 2016; NICOLAOU, 2014). A maioria das falhas estão associadas às fases iniciais do estudo e estão relacionadas a problemas na otimização da substância, toxicidade ou problemas na seleção do alvo (BROWN; SUPERTI-FURGA, 2003). Estima-se que 60% dessas falhas sejam resultantes da não drogabilidade do alvo selecionado (SCHMIDTKE; BARRIL, 2010). Isto é, o alvo de escolha possui uma baixa probabilidade de ligar-se a pequenas moléculas semelhantes a drogas com potencial suficiente de interferir na sua função biológica (SCHMIDTKE; BARRIL, 2010).

Desta forma, o processo de descoberta de novos fármacos torna-se extenso,

laborioso e dispendioso, com custo estimado entre 500 milhões e 2 bilhões de dólares e com tempo necessário para aprovação que se estende de 12 a 15 anos (BASAK, 2012; DIMASI; HANSEN; GRABOWSKI, 2003). Essa realidade torna-se ainda mais preocupante no que tange às doenças negligenciadas, uma vez que medicamentos para este grupo de doenças geram pouco retorno financeiro para a indústria farmacêutica, reduzindo assim os esforços deste setor em financiar pesquisas para o desenvolvimento de novos fármacos (CHARLTON et al., 2018; TROUILLER et al., 2002).

Em consequência do exposto, as falhas no processo de desenvolvimento de fármacos permanecem e o número de drogas aprovadas segue em declínio, com aproximadamente 10.000 compostos medicinais sintetizados para cada droga aprovada (WALTERS et al., 2011). Esse insucesso tem levado a utilização de novas abordagens, como a “*piggy-back*”, que apresenta uma metodologia contrastante com o modelo tradicional de descoberta de novos medicamentos (ANDREWS; FISHER; SKINNER-ADAMS, 2014; KLUG; GELB; POLLASTRI, 2016; NWAKA; HUDSON, 2006a). Esse tipo de abordagem consiste em comparar alvos validados com os produtos gênicos de uma determinada doença, e utilizando como base a similaridade entre eles é possível selecionar novos potenciais alvos de maneira eficiente e inferir o redirecionamento de drogas já aprovadas (ANDREWS; FISHER; SKINNER-ADAMS, 2014; NWAKA; HUDSON, 2006a).

Um dos sucessos da abordagem “*piggy-back*” que pode ser citado é o redirecionamento de inibidores da proteína farnesyltransferase (PFT), desenvolvidos inicialmente para tratamento tumoral. Esses inibidores foram testados como medicamento antimalárico baseado na semelhança estrutural do sítio de ligação entre PFT de mamíferos e PFT de *Plasmodium falciparum*. Revelando assim a citotoxicidade desses inibidores no parasita causador da malária (GELB et al., 2003; NALLAN et al., 2005; NWAKA; HUDSON, 2006b).

Com o advento dessas abordagens, encontrar novas aplicações para drogas já desenvolvidas tornou-se uma potencial estratégia alternativa na descoberta de novas linhas terapêuticas; Tornando-se ainda mais atrativa devido a característica de polifarmacologia clinicamente relevante presente em muitos medicamentos

(OVERINGTON; AL-LAZIKANI; HOPKINS, 2006). Essa característica reflete a possibilidade de interações do fármaco com outro alvo além do alvo de escolha, e geralmente está associada aos efeitos colaterais e adversos produzidos. No entanto, a polifarmacologia de um composto por vezes é essencial para sua eficácia, além de permitir o seu reaproveitamento em outras linhas de tratamento, resultando em economia significativa de tempo e custo (HOPKINS; MASON; OVERINGTON, 2006; ZHANG et al., 2016).

Mesmo considerando as estratégias para reaproveitamento de fármacos, como citado anteriormente, uma grande porção das falhas no processo de desenvolvimento de drogas é ocasionada pela não drogabilidade dos alvos selecionados, no entanto, deve-se enfatizar que nem todos os alvos drogáveis possuem associação com a manutenção da doença, não sendo então alvos de interesse farmacológico. Desta forma, se faz necessária a utilização de diversas fontes de informações de maneira eficiente para selecionar potenciais alvos terapêuticos no contexto biológico e farmacológico (CHENG et al., 2017; YAMANISHI et al., 2008).

Métodos experimentais de alto rendimento, o avanço nas técnicas de biologia molecular e das ferramentas de bioinformática têm gerado uma explosão nos tipos e na quantidade de informações biológicas disponíveis, não somente no que diz respeito a sequências biológicas (DNA, RNA e proteínas), mas também sobre informação estrutural, de dados de expressão gênica, de regulação e de interação proteica. Tais avanços tornaram possível o surgimento de bancos de dados específicos como o TritypDB (ASLETT et al., 2010), que integra e disponibiliza dados genômicos e proteômicos para 56 patógenos da família *Trypanosomatidae*, entre esses 15 espécies do gênero *Leishmania* (ASLETT et al., 2010; AURRECOECHEA et al., 2017).

A disponibilidade dessas informações em bancos de dados possibilita o surgimento de várias abordagens *in silico* capazes de gerenciar e integrar dados díspares gerando um alto impacto na seleção de potenciais alvos para o desenvolvimento drogas (APIC et al., 2005; CLAUS; UNDERWOOD, 2002; PATEL et al., 2013).

Como exemplo da aplicabilidade dessas informações nas fases iniciais do processo de descoberta de um novo fármaco, pode-se citar o estudo desenvolvido por

Patel e colaboradores, que partindo de dados disponíveis sobre genes associados ao câncer humano, realizou uma análise sistêmica do espaço biológico e químico de 479 genes. Os testes computacionais com dados físico-químicos, de expressão gênica, estrutura tridimensional e similaridade em nível de sequência com alvos de drogas já aprovadas, permitiram a identificação de 46 genes de interesse farmacológico no contexto químico e biológico nunca antes explorados como alvos de drogas para tratamento oncológico (PATEL et al., 2013).

No âmbito das doenças que possuem protozoários como agente etiológico, a integração de dados biológicos e farmacológicos para seleção de alvos terapêuticos já foi aplicada utilizando como objeto de estudo o proteoma de *Plasmodium vivax*, parasita causador da malária. A abordagem permitiu inferir compostos anticancerígenos como potenciais fármacos antimalária, identificando uma ação inibidora dos compostos sobre a enzima dihidrofolato redutase (WILLIAMS et al., 2015).

A integração de informações e a inclusão de dados acerca do espaço químico dos fármacos como mais uma camada de informação, permite não somente a seleção de alvos promissores, como também o redirecionamento de drogas já desenvolvidas, reduzindo ainda mais o tempo e o custo do processo de descoberta de fármacos (HODOS et al., 2016; NAPOLITANO et al., 2013). Utilizar a semelhança entre compostos químicos como mais uma camada de informação no redirecionamento de drogas é possível devido ao princípio de culpa por associação, onde drogas similares tendem a ligar-se a alvos similares. Este princípio é apoiado por Gao e Skolnick, que ao avaliarem a estrutura de aproximadamente 20.000 proteínas do proteoma humano, identificaram a existência de cerca de 1000 conformações estruturais de regiões de ligação com fármacos, ocorrendo assim uma alta semelhança entre os sítios drogáveis dessas proteínas (GAO; SKOLNICK, 2013).

Seguindo esta linha de pensamento e realizando uma avaliação de maneira mais completa e integrada, Yamanishi et al. (2008) e Nascimento et al. (2016) incorporaram dados biológicos com informações sobre similaridade estrutural entre drogas e a similaridade entre seus efeitos colaterais. A combinação destes dados tornou possível inferir interação entre produtos gênicos e drogas já aprovadas

(NASCIMENTO; PRUDÊNCIO; COSTA, 2016; YAMANISHI et al., 2008), permitindo priorizar potenciais alvos e drogas já disponíveis, reduzindo assim os riscos de efeitos adversos e o custo da pesquisa (VANHAELEN et al., 2017).

2.3 SELEÇÃO DE POTENCIAIS ALVOS PARA FÁRMACOS

A seleção do alvo nas etapas de desenvolvimento de fármacos é o primeiro passo crítico, e se apresenta como determinante no estabelecimento de sucesso ou falha do projeto (KHANNA, 2012). Essa seleção torna-se complexa devido à compreensão incompleta da biologia dos patógenos e do hospedeiro, tornando assim necessário o uso de múltiplas metodologias para validar um alvo (GILBERT, 2013).

Tradicionalmente, o processo de identificação de alvos farmacológicos é realizado através do isolamento bioquímico usando a afinidade do alvo com a molécula para realizar essa identificação. Nesse tipo de abordagem, conhecida como métodos bioquímicos diretos, as moléculas são conjugadas a bases sólidas ou a marcadores radioativos e os alvos são identificados por detecção direta da interação entre molécula e composto. No entanto, esse tipo de abordagem possui alguns obstáculos, como por exemplo a utilização de grupos reativos do alvo para fixação na base sólida, quando esses mesmos grupos são necessários para a interação alvo-molécula, bloqueando assim a região de interação ou resultando em uma baixa afinidade entre o alvo e a molécula (SCHENONE et al., 2013; ZHENG; CHAN; ZHOU, 2004).

Outras abordagens para identificação de alvos também são utilizadas como os métodos de interações genéticas, que através de técnicas como nocaute gênico comparam o fenótipo gerado no nocaute com os efeitos de um composto, evidenciando assim que a proteína do gene com expressão bloqueada pela técnica é alvo do fármaco em questão; E métodos de inferência computacional que visam a identificação de um alvo partindo do reconhecimento da interação física entre molécula e composto através da predição da posição, orientação e afinidade da droga no sítio de ligação da molécula alvo (MENG et al., 2012; SCHENONE et al., 2013).

Ambos os métodos destacados aqui podem ser aplicados de maneira

independente, embora a validação de uma molécula biológica como alvo seja dada de maneira mais acurada quando utilizada a integração dos resultados dessas metodologias. (SCHENONE et al., 2013). No entanto, a maioria dessas técnicas resultam na descoberta de drogas baseada apenas na análise de um pequeno grupo-alvo selecionado de forma *ad hoc*, explorando apenas uma pequena porção do proteoma total e negligenciando novos alvos potencialmente drogáveis com importância terapêutica. Assim, quando o objetivo é selecionar alvos para o desenvolvimento de novos fármacos, onde análises em larga escala de moléculas e compostos são realizadas, os métodos computacionais se apresentam como uma escolha inicial, fornecendo poderosas metodologias para integrar dados, acelerando as primeiras fases do estudo e reduzindo o custo de execução (PATEL et al., 2013; SANSEAU et al., 2012; SCHENONE et al., 2013).

A detecção de potenciais alvos farmacológicos por métodos computacionais pode ser direcionada ao alvo ou ao ligante através de informações extraídas de: topologia de rede de interação proteica, similaridade em nível de sequência ou estrutural de proteínas com alvos de drogas aprovadas, similaridade da atividade molecular, similaridade da estrutura química dos ligantes e seus efeitos colaterais, drogabilidade estrutural proteica e expressão gênica (DUDLEY; DESHPANDE; BUTTE, 2011; SLIWOSKI et al., 2014). Nos próximos tópicos será discutido como esses dados podem ser usados na seleção de potenciais alvos de drogas.

2.3.1 Redes de interação proteica

As redes de interação proteína-proteína fornecem uma visão global dos processos biológicos mediante sua representação por meio de grafos, compostos por nós simbolizando proteínas e arestas simbolizando interações (BARABÁSI; OLTVAI, 2004; HALAKOU et al., 2017). Em sua completude, uma rede de interação proteica representa o interactoma de uma célula ou do organismo em um contexto biológico (RUAL et al., 2005).

Tais grafos são descritos como $G=(V,E)$, sendo $V(G)$ a representação do conjunto de nós, e $E(G)$ a representação do conjunto de conectores, que por sua vez podem apresentar-se como interações direcionais, utilizados por exemplo em redes

com rotas metabólicas, ou não direcionais, que são utilizados para representar a relação de interação entre proteínas. Em redes direcionais a aresta é orientada em um determinado sentido ($v1 \rightarrow v2$) com nós sucessores e antecessores e uma função (f) de associação entre os elementos, assim o dígrafo é descrito por $G=(V,E,f)$. Em redes não direcionais as arestas não possuem orientação direcional ($v1-v2$), sem o estabelecimento de sucessores e antecessores ou do direcionamento do fluxo de informação, no entanto indicam uma associação entre os nós (RAVASZ et al., 2002; VERLI, 2014).

Sendo as interações entre proteínas essenciais para quase todos os processos biológicos, as informações contidas nas redes podem ser utilizadas para compreender a fisiologia celular, seja essa em estado de homeostase ou em estado de doença (KAR; GURSOY; KESKIN, 2009; RUAL et al., 2005). Essas informações podem ser acessadas através da análise matemática derivada do ramo do conhecimento conhecido como teoria dos grafos, que por sua vez considera características topológicas que podem refletir propriedades biológicas (CSERMELY et al., 2013).

A primeira dessas características está relacionada com a distribuição das interações no grafo, onde as redes de interação proteica reais são ditas como sendo livres de escala, isto é, não apresentando uma uniformidade no número de interações de cada nó. Assim, é possível observar uma pequena quantidade de nós que atuam como nós centrais (“*hub*”) (Figura 4), altamente conectados aos outros, enquanto muitos outros nós possuem poucas interações (BARABÁSI; OLTVAI, 2004; VERLI, 2014).

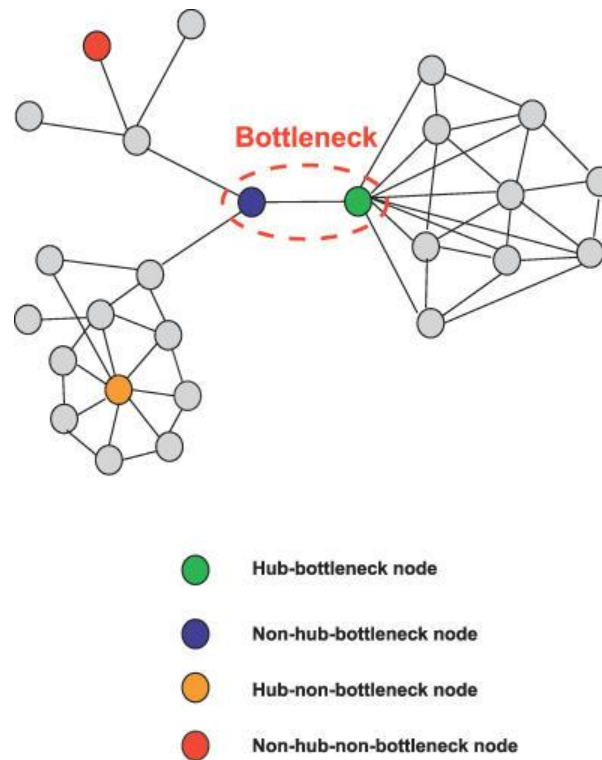


Figura 4: Esquema de categorias de nós em uma rede. Cada nó na rede apresenta uma categoria e está simbolizado por uma cor diferente. O nó verde representa um nó Hub-bottleneck; O nó azul representa um nó bottleneck não hub. O nó laranja representa um nó hub, no entanto não bottleneck. E o nó vermelho representa um nó não hub e não bottleneck (YU et al., 2007).

Esta característica permite rotular as proteínas quanto ao seu grau de interação (*Degree*), tornando possível a identificação das proteínas *hubs* (Figura 4), que devido ao seu alto grau de interação, geralmente se apresentam como essenciais para a estabilidade do fluxo de informação no grafo. A utilização de compostos capazes de interferir nas conexões deste tipo de proteína apresenta uma maior probabilidade de desestabilizar a estrutura da rede. Biologicamente, a interferência em tais proteínas é capaz de gerar alterações funcionais na rede por quebra do fluxo de informação (JEONG et al., 2001; YU et al., 2004).

Assim, a seleção de proteínas *hubs* se apresenta como uma metodologia promissora na identificação de potenciais alvos para o desenvolvimento de fármacos, uma vez que um ataque direcionado a esta proteína pode gerar um efeito cascata comprometendo a viabilidade celular (CSERMELY et al., 2013).

A outra característica topológica está relacionada com a modularidade da rede. Os módulos são grupos de nós densamente conectados entre si, quando comparados

com o restante da rede, e estão associados ao conceito de agrupamento. No contexto geral de uma rede, esta característica pode ser avaliada pelo coeficiente de agrupamento (*Clustering coeficiente*) que é representado pela fórmula $C_i = 2n/k_i(k_i - 1)$, onde K é o tamanho da vizinhança de nós do nó i , e n é o número de conectores nesta vizinhança. Deste modo, quanto maior o coeficiente de agrupamento, mais conectado será o módulo (VERLI, 2014).

Biologicamente, a identificação de módulos numa rede de interação proteica é de grande relevância, pois tais agrupamentos podem refletir módulos funcionais, ou seja, grupos de proteínas relacionados a uma mesma atividade celular. A obtenção de tal informação pode ser integrada a dados de ontologia gênica referentes a uma determinada função desempenhada na maquinaria celular, sua localização subcelular, ou seu envolvimento em um processo biológico (Figura 5). Por fim, a integração desses dados permite associar proteínas não caracterizadas, isto é, sem função definida, a um determinado processo biológico (CSERMELY et al., 2013; VERLI, 2014).

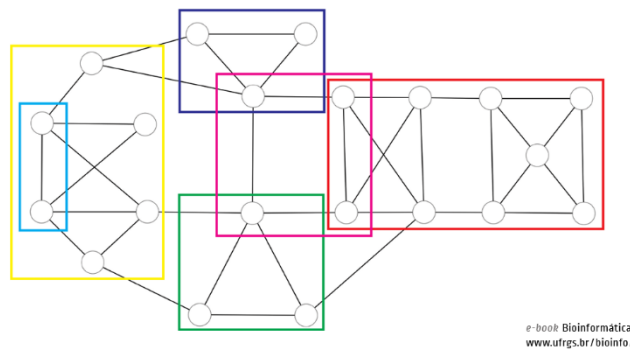


Figura 5: Modelo hipotético da integração da rede a dados de ontologia gênica. Na figura cada cor representa a identificação de um processo. Devido à relação de proteínas com mais de uma ontologia, na rede é possível perceber interseção entre módulos ou subconjuntos. Como exemplo, se o módulo em amarelo representa a ontologia referente à transcrição, o subconjunto azul poderia conter proteínas do complexo de iniciação da RNA polimerase II (VERLI, 2014).

A análise da distribuição de nós e da modularidade permitem ainda ponderar um parâmetro conhecido como *Betweenness*, que reflete a quantidade de caminhos mais curtos entre todos os pares de nós que passam por um nó. Nós com um alto valor deste parâmetro são importantes para integridade do fluxo de informação, uma vez que compõem o caminho da comunicação entre diferentes partes da rede (LUO

et al., 2007; YU et al., 2007). O cálculo do valor de *Betweenness* para um nó n , é dado pelo número de caminhos mais curtos entre os nós i e j , e a fração destes caminhos que passam por n . Sendo matematicamente representado pela fórmula: $Bet(n) = \sum_{i \neq n \neq j \in V} \frac{\sigma_{ij}(n)}{\sigma_{ij}}$, onde σ_{ij} representa os caminhos entre os nós i e j , e $\sigma_{ij}(n)$ representa o total de caminhos mais curtos que passam pelo nó n (VERLI, 2014).

Uma proteína com alto valor de *Betweenness* e posicionamento intermodular é denominada de *Bottleneck*. Esta proteína possui uma elevada capacidade de interagir com outras proteínas sem a necessidade de interação direta (Figura 5). Desta forma, a identificação dessas proteínas também se apresenta como uma poderosa ferramenta na seleção de potenciais alvos, uma vez que proteínas com essa característica estão no caminho de quase toda a informação compartilhada entre nós, e a inibição de suas interações possui alta capacidade de comprometer a integridade da rede (CSERMELY et al., 2013).

Desta forma, é notável o potencial do estudo das redes de interações apoiado nas técnicas da teoria de grafos à seleção de potenciais alvos destinados ao desenvolvimento de fármacos. Indo além, a análise detalhada das redes ainda permite compreender o sistema de ação de uma droga, seus efeitos colaterais ou a complexidade de uma doença, tornando o processo de construção dessas redes e suas análises uma poderosa ferramenta no planejamento racional de fármacos (CSERMELY et al., 2013).

2.3.2 Dados de similaridade

A crescente disponibilidade de dados estimulada pelos avanços das técnicas moleculares, têm impulsionado o desenvolvimento de ferramentas computacionais capazes de analisar e combinar informação de modo a gerar novos conhecimentos. Estas informações combinadas com dados acerca da interação entre alvos e drogas tem traçado um caminho promissor para inferência de relação entre novos alvos e compostos já disponíveis (KOUTSOUKAS et al., 2011).

Muitos destes dados, obtidos através de ensaios experimentais, são de livre acesso em bancos de dados públicos tais como: Drugbank (LAW et al., 2014),

PubChem (KIM et al., 2016), ChEBI (HASTINGS et al., 2016) e Stitch (SZKLARCZYK et al., 2016). Esses repositórios possibilitam a obtenção de informações sobre alvos, drogas e suas interações. Outros bancos de dados fornecem a integração das informações disponíveis nas bases públicas, como o BindingDB (LIU et al., 2007), que além de agregar informações de outros bancos, fornece dados sobre a afinidade entre alvos e compostos. Atualmente, o BindingDB contém 1.842.640 dados de interação, 7.484 proteínas alvo e 815.527 compostos.

Compreender as relações entre dados de sequência, estrutura e função em moléculas proteicas ou drogas pode fornecer informações que auxiliam no entendimento de outras proteínas da família. Reconhecer semelhanças entre essas informações e utilizar esses dados para derivar relacionamentos é, portanto, um dos meios fundamentais da bioinformática para seleção de potenciais alvos (YETURU; CHANDRA, 2008). Dessa forma, a maneira mais comum de inferir atividade de um composto sobre um novo alvo utilizando este tipo de dado é a busca por métricas de similaridade, seja de função, de sequência proteica, de estrutura tridimensional proteica ou de similaridade de estrutura química dos compostos (DUDLEY; DESHPANDE; BUTTE, 2011; KUHN et al., 2008).

2.3.2.1 Similaridade entre drogas

No que tange o âmbito das drogas, a busca pela similaridade entre estruturas químicas pode ser realizada partindo da conversão de sua estrutura bidimensional em uma impressão digital, gerando assim um conjunto de vetores que refletem a estrutura da droga para os quais um coeficiente de similaridade ou distância é calculado (KUHN et al., 2008; WILLETT; BARNARD; DOWNS, 1998). Outra forma de obter um valor de similaridade é utilizando sua estrutura tridimensional, sendo obtido um valor de similaridade através do alinhamento entre as estruturas tridimensionais dos compostos por sobreposição espacial. Apesar de apresentarem resultados semelhantes quando comparadas, as medidas de similaridade bidimensional são mais ágeis, no entanto, acabam negligenciando características estruturais que interferem na função dos compostos (Figura 6) (KUHN et al., 2008; THIMM et al., 2004).

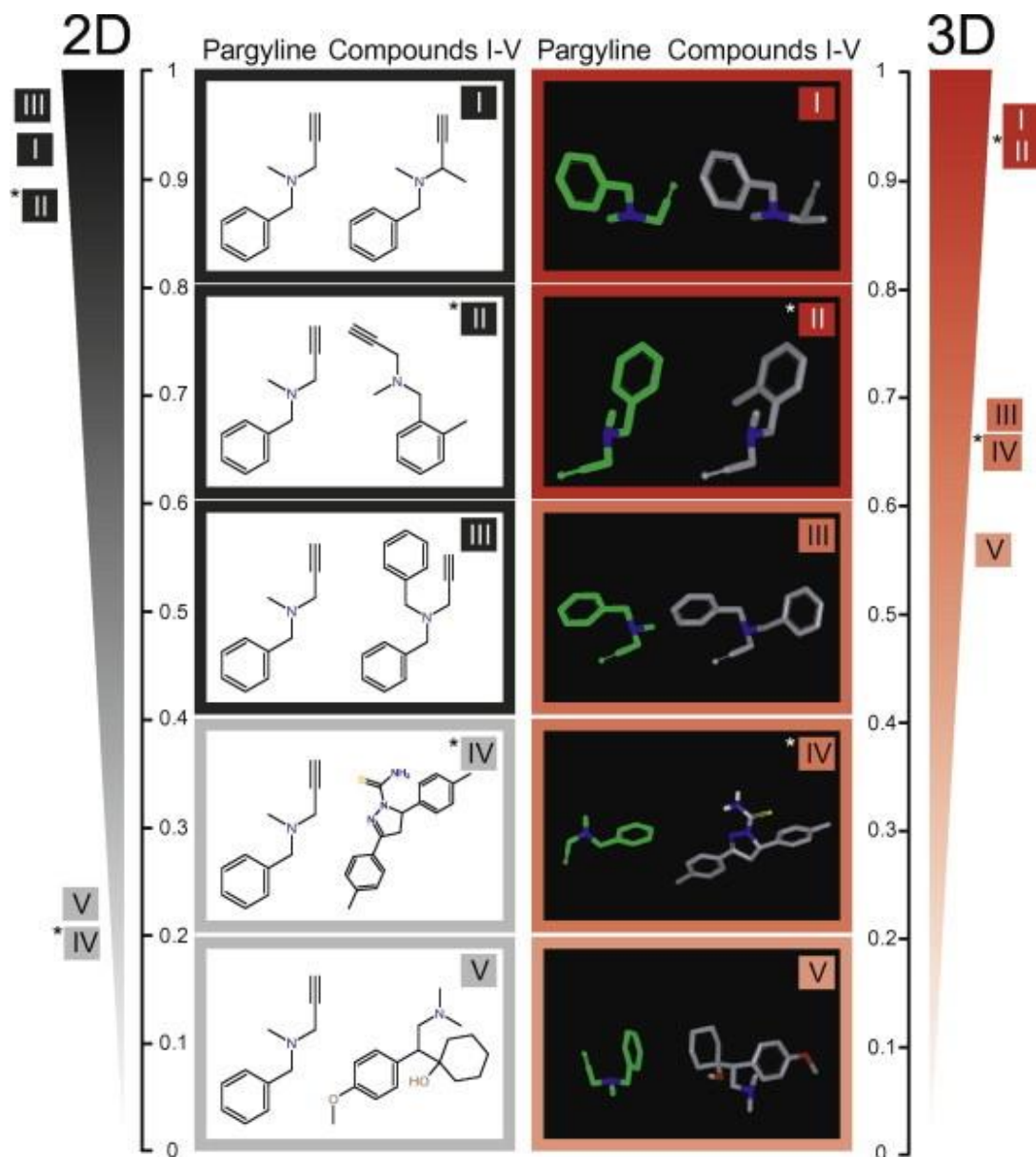


Figura 6: Comparação de medidas de similaridade entre estruturas químicas. Estrutura da droga Pargilina comparada a compostos derivados da Pargilina (compostos I-III), 1-tiocarbamoil-3,5-di- (4-metilfenil) -4,5-di-hidropirazol (composto IV) e a droga Venlafaxina (composto V). Compostos com atividade semelhante à droga Pargilina estão marcados com asterisco. A utilização da estrutura tridimensional revelou uma maior similaridade estrutural entre os fármacos com atividade semelhante à Pargilina, enquanto que a utilização da estrutura bidimensional não permitiu identificar essa similaridade, inferindo uma diferença maior entre a Pargilina, o composto II e o composto IV (KUHN et al., 2008).

2.3.2.2 Similaridade entre alvos de drogas

No que tange o âmbito dos alvos, a semelhança pode ser avaliada no que se refere à função, à sequência e à estrutura das proteínas. A semelhança funcional entre

proteínas pode ser obtida através da similaridade semântica de uma representação formal de sua função, como a ontologia gênica (*Gene Ontology* - GO) (SCHLICKER et al., 2006; ZHAO; WANG, 2018). A GO consiste em um vocabulário controlado estruturado em um grafo direcional acíclico, onde os termos estão relacionados entre si e são capazes de descrever e sintetizar características utilizando linguagem formal. Cada termo está associado a uma ontologia que pode descrever um componente celular (*Cellular Component* - CC), uma função molecular (*Molecular Function* - MF) ou um processo biológico (*Biological Process* - BP) (ASHBURNER et al., 2000). Os termos relacionados a componentes celulares trazem informações acerca da localização subcelular onde a proteína está ativa ou de qual componente celular a mesma faz parte, os termos relacionados aos processos biológicos trazem informações sobre os caminhos e processos nos quais uma proteína está envolvida, enquanto que os termos relacionados a funções moleculares trazem informações acerca das atividades moleculares exercidas por uma proteína (ASHBURNER et al., 2000; CARBON et al., 2019).

O valor de similaridade semântica entre GO pode ser obtido através de métricas baseadas no conteúdo informativo (*Information Content* - IC) do termo ou métricas baseadas em grafo (YU et al., 2010a). As métricas baseadas no IC dependem da frequência dos termos comparados e seu ancestral comum mais próximo, onde o valor de IC é associado a importância do termo, sendo matematicamente representado pela fórmula: $IC = -\log(f_{anot}(c))$, onde $f_{anot}(c)$ representa a fração de produtos gênicos anotados com o termo c ou seus descendentes e reflete a probabilidade do termo ocorrer, sendo menos informativo os termos com maiores ocorrências (RESNIK, 1999; YU et al., 2010a).

Na literatura podem ser encontradas diversas métricas baseadas em IC, dentre as quais as mais simples e comumente usadas são a métrica de similaridade de Resnik (OVASKA; LAAKSO; HAUTANIEMI, 2008; RESNIK, 1999; YU et al., 2010a) e a métrica de similaridade de Lin (LIN, 1998; OVASKA; LAAKSO; HAUTANIEMI, 2008). A similaridade de Resnik consiste unicamente no valor de IC do ancestral comum mais informativo entre dois GO, sendo representada pela fórmula: $SIM_{renik}(t_1, t_2) = IC(A)$, onde A representa o termo ancestral entre os termos t_1 e t_2 com maior valor de IC

(OVASKA; LAAKSO; HAUTANIEMI, 2008; RESNIK, 1999). Assim, essa medida considera o quão informativo é o ancestral comum entre os termos, mas desconsidera a distância entre o ancestral comum e os termos comparados (SCHLICKER et al., 2006).

Em contrapartida, a similaridade de Lin pondera também os valores de IC dos termos t_1 e t_2 , sendo representada pela fórmula: $sim_{lin}(t_1, t_2) = \frac{2IC(A)}{IC(t_1)+IC(t_2)}$ (LIN, 1998; OVASKA; LAAKSO; HAUTANIEMI, 2008). Dessa forma, essa métrica avalia o quão próximos estão os termos do seu ancestral comum mais informativo, no entanto acaba não refletindo o quão informativo é o ancestral comum. Com o intuito de considerar ambas informações, Schlicker et. al. (2006) propôs a métrica *Relevance* que combina as similaridades de Resnik e Lin e é representada pela fórmula: $sim_{rel}(t_1, t_2) = \frac{2IC(A)(1-p(A))}{IC(t_1)+IC(t_2)}$, refletindo assim, tanto a distância entre os termos comparados e o ancestral comum mais informativo, quanto a relevância deste ancestral para anotação (SCHLICKER et al., 2006).

O estudo desenvolvido por Tan et al. (2014) demonstrou o potencial da utilização da similaridade funcional e estrutura química desenvolvendo uma metodologia para redirecionamento de fármacos. Em seus resultados, através dos cálculos de semelhança entre GO juntamente com a similaridade entre estruturas químicas dos compostos foi possível inferir novos alvos para 143 drogas, sendo 71,8% dessas novas indicações apoiadas por ensaios químicos já confirmados na literatura (TAN et al., 2014).

A Similaridade de estrutura primária ou similaridade de sequência pode ser obtida através de técnicas de comparação entre sequências biológicas que buscam por características individuais que estejam na mesma ordem nas sequências analisadas (VERLI, 2014). Dessa forma, os algoritmos de alinhamento realizam uma busca pela melhor correspondência entre as sequências permitindo a inserção de espaços de forma a minimizar as diferenças e encontrar o melhor alinhamento (VERLI, 2014).

As técnicas de alinhamento podem ser divididas entre alinhamentos globais, que consideram toda a extensão das sequências avaliadas; e alinhamentos locais,

que realizam a busca pelo melhor alinhamento entre regiões das sequências. Ao comparar sequências com extensões dissimilares e ao realizar a busca por similaridade entre proteínas com interesse em domínios, motivos ou regiões de interação com drogas, comumente se é utilizado algoritmos que aplicam alinhamento local (LÉON; MARKEL, 2006).

Um dos principais algoritmo utilizado por programas para alinhamento local entre sequências proteicas foi desenvolvido por Smith e Waterman em 1981 (SMITH; WATERMAN, 1981). Este método é baseado em uma abordagem de programação dinâmica que busca um alinhamento local ótimo baseado em matriz de substituição com pontuação definida para *mismatch* (SMITH; WATERMAN, 1981).

Com o avanço dos métodos computacionais, outros algoritmos também foram desenvolvidos para identificar similaridade entre proteínas, como os baseados em função de kernel, *Kernels Mismatch* e *Spectrum*, que calcula a semelhança entre sequências baseado na ocorrência de padrões de tamanho fixo compartilhados entre os dados (LESLIE et al., 2004; PALME; HOCHREITER; BODENHOFER, 2015; SAIGO et al., 2004).

O algoritmo *Spectrum* considera o número de ocorrências de uma subsequência de tamanho k (k -mers) entre as sequências de aminoácidos de duas proteínas p e p' . Assim, para cada k -mer m , é calculado o número de ocorrências $N(m, p)$ e $N(m, p')$, de modo que, como resultado seja obtido um vetor para cada sequência, representado matematicamente por $\phi_{spec}(p) = \{N(m, p)\}_{m \in \mathcal{A}^k}$ e $\phi_{spec}(p') = \{N(m, p')\}_{m \in \mathcal{A}^k}$ onde $|\mathcal{A}|$ consiste no tamanho do alfabeto de aminoácidos. Após a construção dos vetores a similaridade pode ser obtida através do cálculo do produto interno, representado por $K_{spec}(p, p') = \langle \phi_{spec}(p), \phi_{spec}(p') \rangle$ (LESLIE; ESKIN; NOBLE, 2002).

O algoritmo *Mismatch* também utiliza o produto interno entre vetores para calcular a similaridade entre duas sequências proteicas, no entanto permite um número definido de dissimilaridade m em cada k -mer, sendo sempre $m < k$ (LESLIE et al., 2004).

O potencial da busca por alvos de drogas através de métricas de similaridade entre sequências foi avaliado por Mitchell que, utilizando complexos droga-proteína

depositadas no *Protein Data Bank* (PDB), identificou que a similaridade da estrutura química entre ligantes é maior para proteínas com sequências similares (MITCHELL, 2001), concluindo assim que proteínas semelhantes em nível de sequência tendem a ligar-se a compostos semelhantes. Tal conclusão continua a ser corroborada e identificada em outras pesquisas (HEMMINGS; EGAN, 2018; LIU et al., 2007).

No entanto, a falta de similaridade entre sequências não impede a interação de proteínas com compostos semelhantes, uma vez que é de conhecimento que famílias de sequências diferentes podem adotar a mesma conformação. Isso pode ocorrer devido à convergência evolutiva, devido ao número limitado de arranjos espaciais aceitáveis de elementos estruturais secundários que acabam exercendo uma restrição conformacional relacionada a uma determinada função molecular, ou resultante de uma evolução divergente a ponto de tornar a similaridade das sequências não mais reconhecível (KOCH; WALDMANN, 2005). Em ambos os pontos, esse fator ressalta a similaridade estrutural entre proteínas como um ponto a ser explorado na busca por novos alvos para drogas já aprovadas.

Neste contexto, é possível identificar na literatura diversos exemplos que ilustram a relação entre a estrutura de uma proteína e sua função, assim como a relação entre a estrutura de uma proteína e sua interação com um fármaco (YETURU; CHANDRA, 2008). No entanto, também é possível identificar na literatura exemplos de proteínas com sequência e estrutura dissimilares, mas capazes de alcançar funções similares. A exemplo pode-se citar a identificação de pelo menos três proteínas com conformações estruturais e sequências de aminoácidos diferentes, mas capazes de reconhecer e ter função mediada por manose (RAMACHANDRAIAH; CHANDRA, 2000). Além disso, também é de conhecimento que mesmo proteínas com similaridade superior a 90% podem alcançar estrutura e funções diferentes, como as proteínas priônicas (REYNAUD, 2010).

Baseado nessas informações é possível perceber que nenhuma medida isolada é considerada ideal para todos os tipos de busca por características semelhantes. Dessa forma, para realizar uma análise mais completa de potenciais alvos é necessário a utilização de diferentes métricas de similaridade em múltiplas fontes de dados, refletindo assim a natureza multifacetada das características

moleculares (WILLETT; BARNARD; DOWNS, 1998).

2.3.3 Drogabilidade no contexto estrutural

O termo drogabilidade (*druggability*) surgiu no início do século XXI vinculado ao genoma humano, quando se teve como objetivo identificar genes associados a doenças para os quais fosse possível o desenvolvimento de compostos semelhantes a drogas (HOPKINS; GROOM, 2002). Posteriormente, outros estudos atribuíram novas definições ao termo, vinculando-o ao produto do gene e mais especificamente aos sítios de ligação de um fármaco. Assim, atualmente o termo ‘drogável’ refere-se a proteínas com importância para a manutenção de um estado de doença cuja atividade molecular pode ser modulada através da interação com compostos semelhantes à droga (DESAPHY et al., 2012; HUSSEIN et al., 2016; SCHMIDTKE; BARRIL, 2010).

Após a primeira utilização do termo drogabilidade, outras terminologias passaram a ser utilizadas para descrever a capacidade de interação de uma determinada droga a um alvo. No entanto, apesar de possuírem significados semelhantes, os termos drogabilidade, capacidade de ligação (*ligandability* ou *bindability*) e tratabilidade (*tractability*) podem referir-se a definições diferentes. O termo tratabilidade é utilizado para descrever proteínas moduladas *in vitro* por compostos químicos, enquanto que a capacidade de ligação reflete a possibilidade de interagir com pequenas moléculas independente de sua atividade farmacológica (HUSSEIN et al., 2016).

Considerando a definição de drogável, apesar de muito vinculada à estrutura tridimensional do sítio de ligação, é possível atribuí-la também aos alvos identificados pelos métodos de similaridade citados no tópico 2.3.2, sendo estes métodos relatados como preditores de drogabilidade baseados em evolução (VOLKAMER et al., 2010). Desta forma, para tornar mais compreensível, utilizaremos o termo aqui vinculado somente a capacidade de interação com uma droga no contexto estrutural. A principal diferença da aplicabilidade dessas duas formas de identificar alvos drogáveis, é que utilizando unicamente dados de similaridade só é possível inferir drogabilidade a proteínas pertencentes a famílias proteicas já caracterizadas como alvos de drogas.

Enquanto que a utilização da estrutura tridimensional proteica permite aplicar o termo drogável para novos alvos, trazendo à luz informações acerca da estrutura do sítio de ligação e atribuindo a este um índice de drogabilidade (KOZAKOV et al., 2015; VOLKAMER et al., 2012).

Os métodos computacionais que utilizam a informação tridimensional do alvo realizam a previsão de drogabilidade partindo da identificação do local em que o ligante interage com a proteína, sendo estes sítios geralmente localizados em cavidades na superfície da estrutura proteica (TARI, 2012). Desta forma, as ferramentas computacionais podem identificar essas regiões através de metodologias baseadas em energia de ligação alvo-droga, sendo o *docking* molecular o mais comumente utilizado, ou baseados em geometria pela busca das cavidades na superfície da estrutura proteica (NICOLAOU, 2014; VOLKAMER et al., 2010).

Os métodos de energia por *docking*, também conhecidos como métodos de triagem virtual de alto rendimento (*Virtual High Throughput Screening*) possuem como base os princípios de complementariedade físico-química entre sítio ativo e ligante. Utilizando duas bibliotecas compostas por estruturas tridimensionais dos alvos e pequenas moléculas, realizam a busca da orientação estrutural preferencial entre cada proteína e ligante. Esses métodos possuem como vantagem a disponibilização de informações estruturais acerca da natureza da interação alvo-droga e a afinidade da ligação. No entanto, possuem um alto custo computacional sendo pouco aplicáveis frente a um proteoma completo e todos os compostos disponíveis (MINIE et al., 2014; XU; HUANG; ZOU, 2018).

Os métodos baseados em geometria podem realizar a identificação das cavidades a partir de três princípios diferentes: grade, esfera e teoria da forma-alfa (Figura 7) (KAWABATA, 2010; ZHANG et al., 2011; ZHOU; YAN, 2014). Após a identificação, diversas propriedades da cavidade são calculadas, tais como hidrofobicidade, tamanho, compacidade, áreas de superfície doador e receptor de hidrogênio, e composição de aminoácidos na cavidade. Desta forma, é possível inferir um índice de drogabilidade para cada cavidade identificada (KOZAKOV et al., 2015).

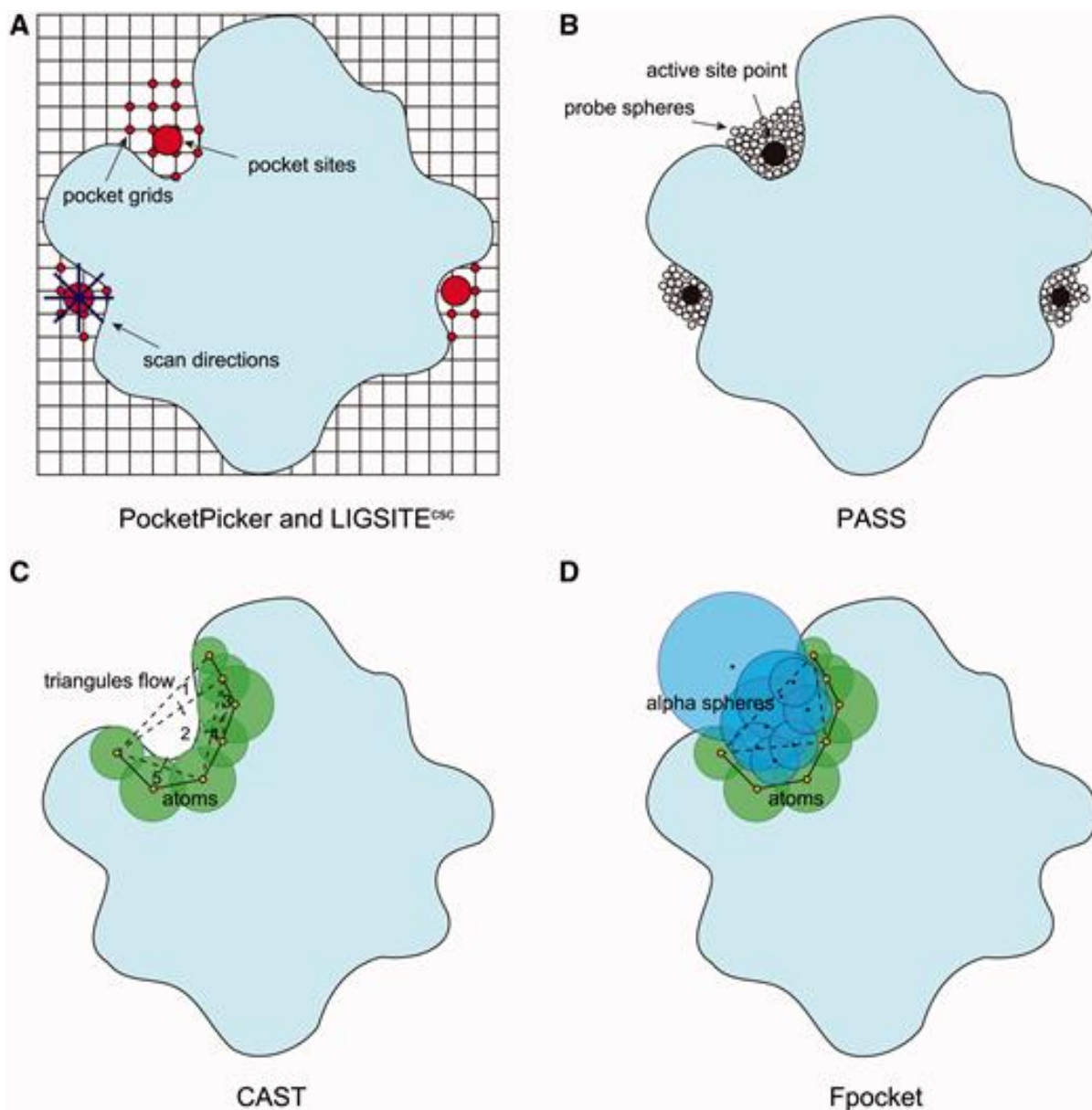


Figura 7: Métodos de detecção de sítio de ligação. A: Método baseado em grade aplicado pelas ferramentas PocketPicker e Ligsite. B: Método baseado em esfera aplicado pela ferramenta Pass. C: Método baseado na teoria da forma alfa utilizando fluxo de triângulos para detecção do bolso aplicado pela ferramenta CAST; D: Método baseado na teoria da forma alfa utilizando o conceito de esferas alfa para detecção do bolso aplicado pela ferramenta Fpocket (ZHOU; YAN, 2014).

Nos métodos baseados em grade (Figura 8), a estrutura da proteína é projetada em uma grade tridimensional, sendo os pontos da grade em relação a proteína categorizados como acessíveis ou inacessíveis ao solvente. As ferramentas então realizam uma varredura na grade ao longo dos eixos X,Y,Z em busca de grupos de pontos acessíveis ao solvente cercados por pontos não acessíveis ao solvente. Essas regiões representam bolsos e são nomeadas de eventos proteína-solvente-proteína

(WEISEL; PROSCHAK; SCHNEIDER, 2007; ZHANG et al., 2011). Este tipo de metodologia é aplicado por ferramentas como o Ligsite (HENDLICH; RIPPMANN; BARNICKEL, 1997) e PocketPicker (WEISEL; PROSCHAK; SCHNEIDER, 2007).

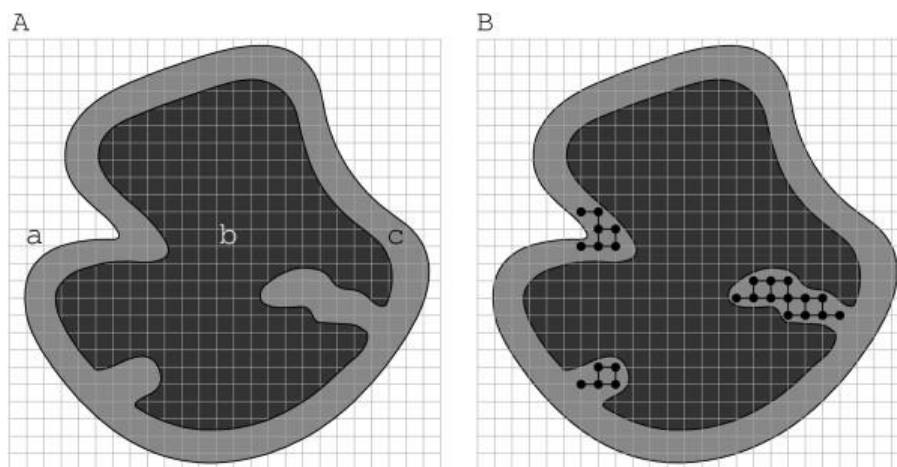


Figura 8: Metodologia de predição de sítios de ligação baseada em grade. A: Pontos longe da proteína (a) e longe da superfície (b) não são considerados nos cálculos, sendo somente considerados os pontos próximos a superfície da estrutura (c). B: pontos da grade indicando regiões com cavidades (WEISEL; PROSCHAK; SCHNEIDER, 2007).

Nos métodos baseados em esferas (Figura 9), a estratégia consiste em preencher a superfície da proteína com esferas formando uma primeira camada, para cada esfera é então calculado o número de átomos da proteína dentro de um raio de oito *ångströms*; esse valor é nomeado de “contagem de enterro”. Camadas adicionais de esferas são aplicadas para preencher as cavidades e posteriormente é calculado o número de esferas vizinhas de cada esfera, sendo este valor integrado à contagem de preenchimento definindo assim o peso da esfera. Este valor é então utilizado para identificar os locais ricos em esferas que por sua vez representam bolsos de ligação (WEISEL; PROSCHAK; SCHNEIDER, 2007; ZHANG et al., 2011). Essa metodologia é aplicada por ferramentas como SURFNET (LASKOWSKI, 1995), PASS (BRADY; STOUTEN, 2000) e POCASA (YU et al., 2010b).

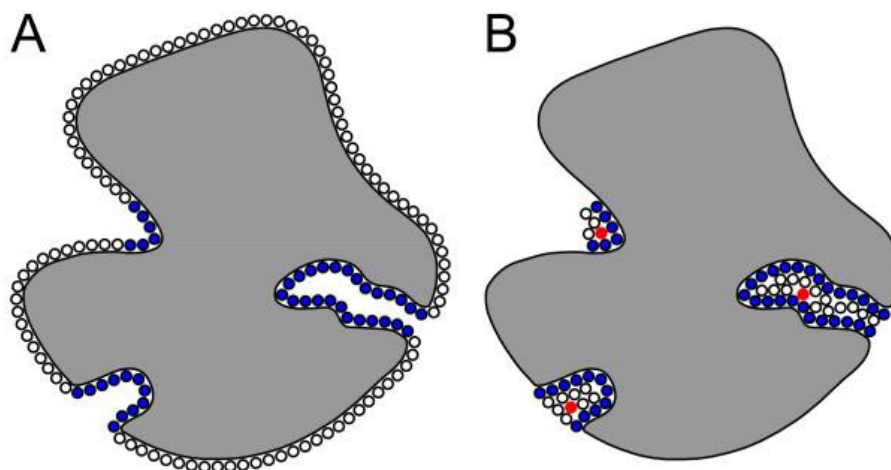


Figura 9: Representação da metodologia de predição de sítios de ligação baseada em esferas. Esferas em azul nas cavidades representam a primeira camada, esferas vermelhas representam o ponto central de uma região rica em esferas (WEISEL; PROSCHAK; SCHNEIDER, 2007).

Os métodos baseados na teoria da forma alfa (Figura 10) (EDELSBRUNNER; MÜCKE, 1994) são aplicados atualmente pelas ferramentas Cast (LIANG; EDELSBRUNNER; WOODWARD, 1998) e Fpocket (LE GUILLOUX; SCHMIDTKE; TUFFERY, 2009), no entanto ambas ferramentas aplicam cálculos distintos para identificar as cavidades. A teoria da forma alfa foi proposta por Edelsbrunner e Mücke em 1994 para cálculo de estrutura de pontos baseados na triangulação de Delaunay construída a partir do diagrama de Voronoi (Figura 10-A) (ZHOU; YAN, 2014). Ao ser aplicada a estruturas proteicas, a triangulação de Delaunay representa a relação entre átomos vizinhos e sua borda reflete a forma alfa (Figura 10-B). Ambas ferramentas aplicam a forma alfa para representar a superfície da proteína (Figura 10-C), no entanto a ferramenta Cast aplica fluxo de triângulos vazios para detectar os bolsos (Figura 10-C), enquanto a ferramenta Fpocket aplica o conceito de esferas alfa de tamanhos distintos, que são esferas com quatro átomos em sua superfície e nenhum átomo no centro (Figura 10-D). Assim, o diâmetro da esfera reflete a curvatura dos 4 átomos e o agrupamento de esferas de raios distintos reflete a cavidades e rugosidade da estrutura proteica (ZHOU; YAN, 2014).

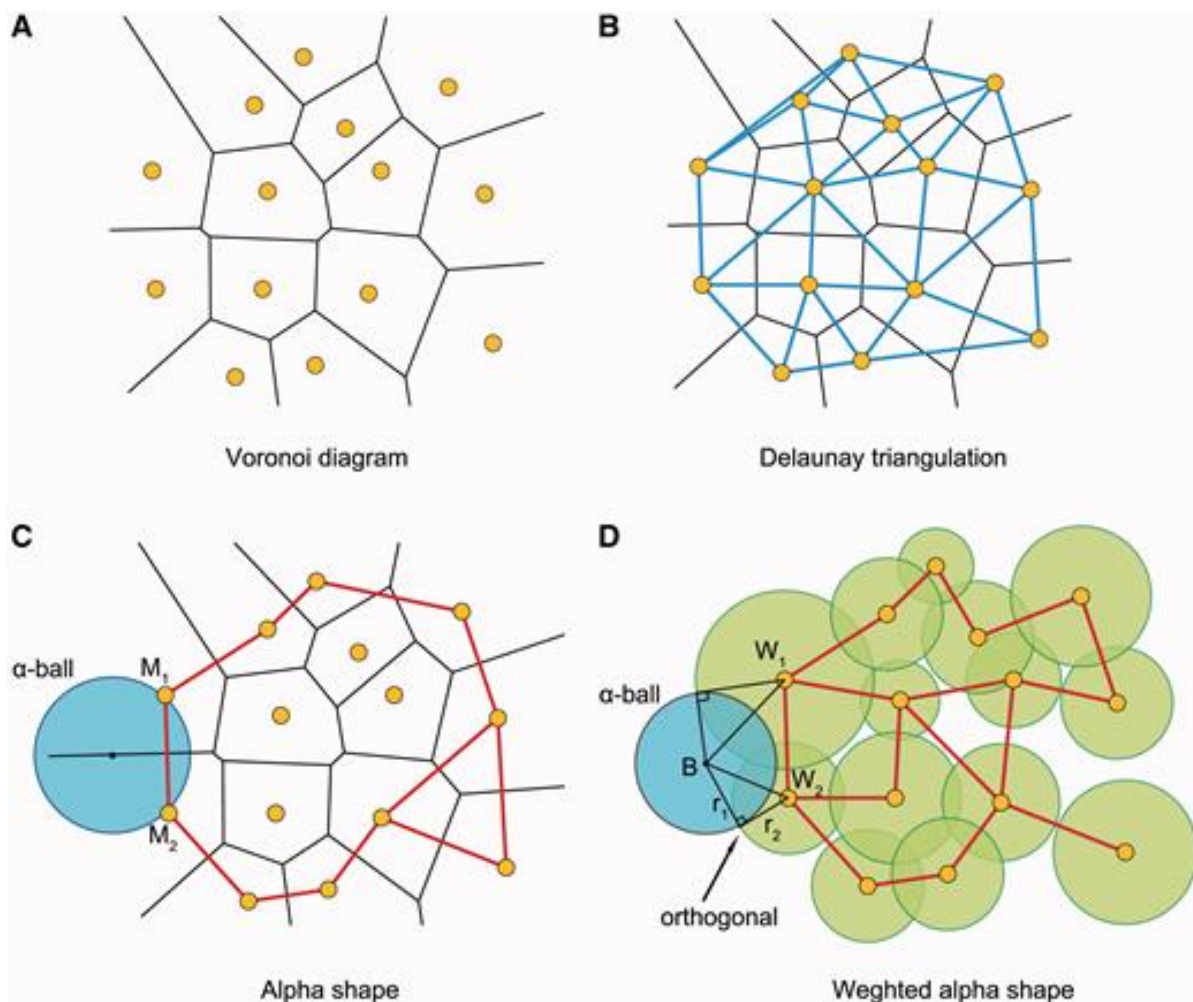


Figura 10: Construção da forma alfa aplicada à busca por sítio de interação. A: diagrama de Voronoi para um conjunto de pontos. B: Triangulação de Delaunay obtida conectando os pontos que compartilham faces do diagrama de Voronoi. C: Bordas da triangulação representando a forma alfa básica que é aplicada à estrutura da proteína. Em azul a representação da esfera alfa vazia com os átomos M_1 e M_2 em sua superfície. D: Forma alfa ponderada (ZHOU; YAN, 2014).

Em estudos realizados por Volkamer *et al.*(2010), a ferramenta FPocket apresentou um alto desempenho, quando comparada as demais ferramentas citadas baseadas em geometria, sendo então considerado um dos métodos mais precisos na detecção de regiões de interação alvo-droga (VOLKAMER *et al.*, 2010).

A identificação de sítios de ligação de drogas permite ainda comparar estruturalmente essas regiões com sítios de ligação de alvos já identificados. Avaliar a similaridade entre os locais de ligação difere da similaridade das estruturas completas, pois os locais de ligação com fármacos são pequenos contendo apenas alguns resíduos; e ainda diferem também da similaridade a nível de sequência, uma

vez que esses resíduos frequentemente não são contíguos na estrutura primária da proteína (YETURU; CHANDRA, 2008). Além disso, estudos sugerem uma natureza finita das possibilidades de conformação e propriedades de sítios de ligação, mostrando assim a importância de utilizar essa métrica na busca por alvos similares para redirecionamento de drogas e esclarecendo a questão acerca da promiscuidade dos fármacos (SKOLNICK et al., 2015).

2.3.4 Análise de dados de expressão gênica

Identificar transcritos e quantificá-los têm sido uma das atividades centrais da biologia molecular para diversas aplicações desde a descoberta da função do ácido ribonucleico (RNA) como intermediário entre o genoma e o proteoma. O ensaio de sequenciamento de RNA de alto rendimento (RNA-seq) tem sido uma das principais formas de realizar a identificação e quantificação dos transcritos em um único experimento, gerando assim o que é nomeado de dados de expressão gênica (CONESA et al., 2016).

Os dados de expressão gênica refletem a atividade do RNA mensageiro (mRNA) de genes sob uma condição específica, sendo essa condição determinada de acordo com o objetivo do estudo. Dessa forma, conhecer a abundância de um transcrito permite identificar a expressão diferencial de genes entre tecidos, estágios de desenvolvimento ou condições de estresse (CONESA et al., 2016; ISIK et al., 2015).

Na análise de alvos de drogas, os dados de expressão gênica podem ser utilizados para identificar os efeitos causados por um fármaco ou até identificar seus potenciais alvos. Esses resultados são alcançados através de experimentos com um organismo controle versus este mesmo organismo submetido a uma determinada intervenção medicamentosa. Dessa forma, a análise da expressão diferencial com estes dados permite identificar quais genes apresentaram alterações em seus níveis de expressão refletindo assim os efeitos causados e as vias afetadas por um determinado fármaco (ISIK et al., 2015).

No que tange a identificação de novos alvos para o desenvolvimento de drogas, usando como exemplo a identificação de alvos para oncofármacos, os dados de

expressão gênica podem ser utilizados para priorizar produtos gênicos mais expressos em células cancerígenas (PATEL et al., 2013). Assim como, para a identificação de potenciais alvos em espécies que possuem estágios morfológicos distintos, os dados de expressão gênica podem ser utilizados para priorizar os genes mais expressos no estágio morfológico de maior interesse para terapia.

2.4 INTEGRAÇÃO DOS DADOS

Mesmo com os resultados das ferramentas apresentadas e com o número de abordagens disponíveis, nenhuma metodologia sozinha é capaz de compreender o espaço farmacológico e biológico em estudos de exploração de alvos para drogas. Deste modo, faz-se necessária a aplicação de métodos integrativos, capazes de explorar o espaço químico e genômico de modo a selecionar potenciais alvos considerando sua essencialidade e sua drogabilidade. A integração de dados se apresenta como uma metodologia promissora para ranquear proteínas e tomá-las como ponto de partida no desenvolvimento de drogas de maneira racional e direcionada (YAMANISHI et al., 2008).

Assim, o presente estudo apresenta uma análise integrada dos dados multiômicos de *Leishmania braziliensis* e *infantum*, os principais agentes causadores das formas de leishmaniose no Brasil, através da análise computacional de dados de drogabilidade, função molecular, processos biológicos, expressão gênica, localização subcelular e similaridade de estrutura tridimensional, integrando esses dados com análises de similaridade entre drogas já desenvolvidas, com o objetivo de priorizar potenciais alvos terapêuticos para análises experimentais. Além disso, propomos moléculas ativas com possibilidade de redirecionamento e interação com os potenciais alvos, possibilitando uma redução de prazo e custo no desenvolvimento de drogas com potencial leishmanicida e tornamos disponível um banco de dados que pode ser utilizado para análises futuras, experimentais ou de aprendizagem de máquina.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Selecionar potenciais alvos para reposicionamento de fármacos contra agentes etiológicos da leishmaniose tegumentar (*L. braziliensis*) e leishmaniose visceral (*L. infantum*) a partir da integração de dados estruturais, químicos, funcionais, celulares e de interação proteica.

3.2 OBJETIVOS ESPECÍFICOS

1. Avaliar o contexto topológico do proteoma predito das espécies de *L. braziliensis* e *L. infantum* presentes nas redes de interação proteica já existentes.
2. Avaliar a similaridade dos proteomas citados no objetivo anterior contra proteínas já descritas como alvos de drogas aprovadas.
3. Realizar a classificação funcional e de domínios desses proteomas a partir de ontologias gênicas (*Gene Ontology*, GO).
4. Predizer a localização subcelular proteica dos proteomas citados.
5. Avaliar a drogabilidade das proteínas deste organismo através de modelos e estruturas já existentes.
6. Integrar e avaliar os dados obtidos nos objetivos 2 a 5 construindo matrizes de similaridade.
7. Verificar a expressão das proteínas destes organismos através de dados públicos de experimentos de RNA-Seq disponíveis na base de dados *Sequence Read Archive* (SRA).
8. Ranquear e disponibilizar uma lista de potenciais proteínas alvo e suas respectivas drogas.

4 **CAPÍTULO I - BUILDING PROTEIN-PROTEIN INTERACTION NETWORKS FOR LEISHMANIA SPECIES THROUGH PROTEIN STRUCTURAL INFORMATION**

Crhisllane Rafaele dos Santos Vasconcelos,^{1,3} Túlio de Lima Campos,^{1,2} e Antonio Mauro Rezende^{1,2,3}

1Microbiology Department of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE Brazil

2Bioinformatics Plataform of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE Brazil

3Genetics Department of Universidade Federal de Pernambuco, Recife, PE Brazil

Building protein-protein interaction networks for *Leishmania* species through protein structural information

Artigo publicado pela revista BMC Bioinformatics no dia 6 de março de 2018. A versão online deste artigo contém material suplementar no formato de website.

DOI: 10.1186/s12859-018-2105-6

PMCID: PMC5840830

PMID: 29510668

Recife

2020

RESEARCH ARTICLE

Open Access



Building protein-protein interaction networks for *Leishmania* species through protein structural information

Chrisllane Rafael dos Santos Vasconcelos^{1,3*}, Túlio de Lima Campos^{1,2} and Antonio Mauro Rezende^{1,2,3*}

Abstract

Background: Systematic analysis of a parasite interactome is a key approach to understand different biological processes. It makes possible to elucidate disease mechanisms, to predict protein functions and to select promising targets for drug development. Currently, several approaches for protein interaction prediction for non-model species incorporate only small fractions of the entire proteomes and their interactions. Based on this perspective, this study presents an integration of computational methodologies, protein network predictions and comparative analysis of the protozoan species *Leishmania braziliensis* and *Leishmania infantum*. These parasites cause Leishmaniasis, a worldwide distributed and neglected disease, with limited treatment options using currently available drugs.

Results: The predicted interactions were obtained from a meta-approach, applying rigid body docking tests and template-based docking on protein structures predicted by different comparative modeling techniques. In addition, we trained a machine-learning algorithm (Gradient Boosting) using docking information performed on a curated set of positive and negative protein interaction data. Our final model obtained an AUC = 0.88, with recall = 0.69, specificity = 0.88 and precision = 0.83. Using this approach, it was possible to confidently predict 681 protein structures and 6198 protein interactions for *L. braziliensis*, and 708 protein structures and 7391 protein interactions for *L. infantum*. The predicted networks were integrated to protein interaction data already available, analyzed using several topological features and used to classify proteins as essential for network stability.

Conclusions: The present study allowed to demonstrate the importance of integrating different methodologies of interaction prediction to increase the coverage of the protein interaction of the studied protocols, besides it made available protein structures and interactions not previously reported.

Background

Leishmaniasis represents a series of infections that have as etiological agents species of parasites of the genus *Leishmania*. Belonging to the group of neglected tropical diseases, with more than 90 endemic countries and approximately 1 million new cases per year, leishmaniasis has become a worldwide public health problem [1]. Despite efforts to develop vaccines and new drugs against these diseases, no effective vaccine has been made available, and existent drugs have serious limitations on their use, such as high toxicity, resistant parasites selected by

drug pressure and incompatible costs in countries underdeveloped [2–4].

Observing the number of reported cases of leishmaniasis and the difficulties in the treatment and prevention, it is clear the need for approaches that allow a wider understanding of the mechanisms of the diseases, and then we will be able to accelerate the steps toward the development of new drugs. It is already known that comprehension about interactions between proteins and the behavior of this biological system are key information to achieve that goal [5–7], and once this data is obtained in 'omics' scale, it allows the prediction of biological function [8–11], identification of changes at gene expression regulation associated with a disease [6, 12], identification of major modules and essential proteins associated [6, 13]. In the end, the analysis

* Correspondence: crisillane@gmail.com; antonio.rezende@cpqam.fiocruz.br

¹Microbiology Department of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

of this data generates critical information for the development of new specific drugs, also making possible to predict side effects of new drugs and to understand the side effects of drugs already used [14–16].

Several methodologies, capable of handling and generating large-scale protein interaction data, have been employed, such as the experimental techniques of yeast two-hybrid and affinity purification coupled with mass spectrometry [17]. However, because the problems involving experimental methods, such as cost, laboriousness and susceptibility to systemic errors, over the years, several computational methods have been developed and used to predict protein interaction networks (PIN) [18, 19].

The computational methods can be categorized in different approaches: compiling existing data available in the literature, named text mining [20], data prediction methods based on primary-structure, evolution and tertiary-structure, such as the methods by sequence homology [21–23], co-location [24], similarity of phylogenetic distribution [25] and rigid-body docking [26–28]. Thus, applying bioinformatics tools, extracting and manipulate biological information have been possible to predict protein interaction networks quickly, efficiently and generally with satisfactory numbers of nodes and interactions [6].

Protein interaction networks have been used in some studies with the objective of selecting promising therapeutic targets [29–31], and the protein interaction data contained in this type of network has already been used in the pharmaceutical industry to development of new drugs [32]. Despite the most of the studies involving protein interaction data embraced by the pharmaceutical industry are concentrated in the area of oncology, this breakthrough highlights the value of information contained in a PIN, and it encourages researchers to obtain such data in other areas, like infectious diseases, where analyses using PINs have already been carried out for *Mycobacterium tuberculosis* [33], *Plasmodium falciparum* [34], and *Brugia malayi* [35], which are agents that cause tuberculosis, malaria and filariasis, respectively.

PIN analysis is one of the most promising methodologies for identifying therapeutic targets, understanding drug action and predicting side effects [36]. The use of this approach to the development of new drugs for leishmaniasis is possible, but few data of protein interaction for *Leishmania* species are available. Large-scale experimental methodologies have been used, but they have been directed to host-leishmania interaction [37], so most of the available networks were obtained by computational methods such as PIN predicted through sequence similarity [38, 39] and those predicted through text mining, co-occurrence and co-expression deposited in the String database [40]. However, despite the multiple methodologies used, less than 50% of

the proteome of the *Leishmania* species are present in these PINs.

Due to the limited data available on protein interaction for species of *Leishmania*, and considering the importance of this information to accelerate the steps for development of new drugs, we predict here a PIN for *Leishmania braziliensis* and *Leishmania infantum* using physical interaction data between protein structures. It is worth to mention those two species were selected as they belong to two distinct subgenera, *Viannia* and *Leishmania*, respectively, and they are the main leishmania pathogens in Brazil [41, 42], causing mainly cutaneous and muco-cutaneous disease (*Viannia*) and visceral disease (*Leishmania*). Therefore, a meta approach [43, 44] that combines two different methods of predicting PIN was applied: the rigid-body method, that predicts interaction through an exhaustive search of orientations of a protein in relation to the other one based on its atomic coordinates; and the template-based method, that use structural similarity between proteins and known protein complexes [28]. This methodology has not yet been used for *Leishmania* proteomes, hence it allows a complementation for existent available networks, providing new information on interactions and inserting new proteins into these networks. At the end, it is possible to improve and increase the possibilities of data extraction for selection of potential new drug targets.

Methods

Prediction of protein structures

The sequences of the predict proteomes of *L. braziliensis* and *L. infantum* version 8.0 were obtained from the Tri-TrypDB database [45]. The use of computational methods to predict three-dimensional conformation of the proteins was necessary because just few structures for those proteomes were deposited in the Protein Data Bank (PDB) [46]. To perform this task, we applied template-based protein structural modeling methodologies through the Modeller [47] version 9.14 and Modpipe version 2.2.0 [48] algorithm packages, and the Mholline [49] and Protein Homology/analogy Recognition Engine version 2.0 (Phyre2) [50] web-servers.

The modeling algorithm of the Modeller package (model-single) predicts three-dimensional models from the comparative modeling using the alignment of the target sequence against the template sequence, and extracting the spatial constraints from the atomic coordinate file of the template, obeying the terms of a probability density function based on empirical data [47]. The templates were selected using the specific protein alignment algorithm (blastp) of the Basic Local Alignment Search Tool (BLAST) package [51], which made possible to analyze the sequence identity and coverage alignment of the leishmania proteomes against the data

deposited in the PDB. Only templates with a minimum of 50% identity and 80% coverage were used. Afterward, two tools were used to perform the Modeller input alignment between the target and template sequences. First, the algorithm for alignment of the modeller package (align2d) [47], and second, the Mafft tool version 7.0 [52]. Align2d is based on dynamic programming algorithm [53], and it takes into account the atomic coordinates of the template [47]. In contrast, Mafft is based on Fast Fourier Transform, and it uses iterative refinement that takes into account evolutionary information to generate alignment [52]. Both alignments were used to predict three-dimensional structures. Modpipe is an automated version of the Modeller package, and it was used to enable a different template search applying profile-profile and sequence-profile alignment [48].

The Mholine server also uses the modeling algorithm of the Modeller package, but it uses the Blast Automatic Targeting for Structures (BATS) and Filter tools to evaluate the quality of the templates, and then to select the best template for comparative modeling [49].

Unlike the tools already mentioned, the Phyre2 server has its own structural modeling algorithm, which implements ab-initio modeling for the portion of the protein which no template has been found. In addition, Phyre2 selects templates based on alignment of Hidden Markov Models via HHsearch [50, 54].

In general, the available template-based protein modeling tools can efficiently predict protein structures when they are executed with high quality templates and identity values between query and template proteins are greater than 25% [55]. In addition, for using structures, which have been predicted by these methods, to computational assays of protein interaction, it is often necessary to perform a full-atomic refinement simulation to increase the quality of the models [56, 57]. Therefore, all predicted structures were submitted to the Modrefiner [57] refinement algorithm.

The quality of the models was evaluated against stereochemical and energetic parameters using Procheck [58] tool and against the standard Discrete Optimized Protein Energy (DOPE) function of the Modeller package [59]. The evaluation of these parameters allows checking conformational stability and approximation of the model to the correct folding [60]. Thus, only models that obtained values for these parameters according to the recommendation of the used tools (torsion angles in a more favorable region in ramachandran plot calculated by Procheck $> 90\%$ and normalized DOPE ≤ -1) were submitted to computational tests of protein interaction.

Prediction of protein interactions using docking methods

The protein models were grouped according to the subcellular localization predicted by the Wolfpsort tool [61],

thus reducing the possibility of false positive interaction prediction, besides decreasing the computational time spent on interaction predictions through docking. The three-dimensional protein models of each group were applied to two docking methodologies: first template-based docking through the Prism Protocol [62] tool, and second, the rigid-body docking through the Megadock [63] tool version 4.0.2.

The Prism Protocol requires as input atomic coordinates of two proteins, and a template set formed by pairs of proteins that are known to interact. This tool applies the Naccess [64] and Multiprot [65] softwares to compare the residues responsible for the interaction in the template set with the surface residues of a pair of target proteins, and then Prism Protocol uses this information to infer interaction between a pair of target proteins. In the end, a predicted protein complex is subjected to flexible refinement and energy minimization using the Fiberdock [66] tool [62]. The generated complexes are ranked according to the global energy binding score, and they are selected if they have a score equal to or less than 0. This threshold is the same one used by the developers of the tool to predict interactions in the Prism Web Server [67].

In parallel, the Megadock tool uses only the atomic coordinates of two proteins, and considering shape complementarity, electrostatic and hydrophobic interactions, it computes a set of interaction solutions for a candidate pair of proteins [63]. The prediction of protein interaction through the de novo docking methodology, like Megadock applies, can be described as a binary classification problem, where the resulting set represents a possible or non-possible interaction. To perform this classification, we first used two algorithms based on clustering for evaluating the docking solutions, the Megadock package clustering algorithm [63, 68] and the Calibur tool [69]. The first one generates an affinity value for a predicted Protein-Protein Interaction (ppi-score), this value takes into account the similarity between the solutions and the z-score of the docking score [63, 68], while the second tool groups the solutions by Root Mean Square Deviation (RMSD), and it finds a suitable distance for that grouping, which we call here Calibur-score. This distance is then used to infer whether this interaction represents a true interaction or does not [69]. These scores were submitted to machine learning algorithms in order to classify the complexes generated by rigid-body method.

Prediction of protein interactions using machine-learning techniques

Initially, we obtained a benchmark data set for the construction of machine-learning predictors of protein

interactions. To do so, all the steps performed by the de novo docking were also applied to a set of positive interaction data, composed of 119 protein pairs that are known to interact, obtained from the Benchmark 4.0 database [70], and to a set of negative interaction data, composed of 147 non-interacting protein pairs obtained from the Negatome database [71]. Hence, our final training/test dataset was composed by 266 total entries, where the Calibur and PPI-scores were used as feature inputs, and the outputs were set as “1” for interacting protein pairs, and “0” for the non-interacting pairs. The construction of the learning models was performed using R (<https://cran.r-project.org>) along with the following libraries: *stats* (Linear Regression Model), *e1071* (Support-Vector Machine and Naive Bayes), *randomForest* (Random Forest), *neuralnet* (Neural Network/Perceptron) and *gbm* (Gradient Boosting Method). For performance assessment and visualization, we used *ROCR*, *PRROC*, *ggplot2* and *plotly* packages.

Six popular machine-learning algorithms for binary classification were trained with default parameters. We performed 100 training/test iterations where we randomly selected 70% of the positive and 70% of the negative interaction data, using them as training sets for each model, then we used the remaining 30% as test sets, calculating the accuracy and area under the curve (AUC) of the Receiver Operating Characteristic (ROC) graph. In addition, Precision and Recall values were calculated for each iteration. We highlight that it was not part of the present work to exhaustively find optimum parameters for each machine-learning method used. After all iterations, we generated boxplots showing the AUCs for each model, and performed statistical tests (pairwise t-tests and TukeyHSD) comparing the performance across the different algorithms. Finally, the model that presented the best performance was selected to classify *Leishmania* interaction data.

The best models built based on the training sets generated response values, ranging from 0 to 1, for the interaction prediction of each pair of proteins. Precision, recall and specificity values were analyzed to define a response value threshold to classify the positive or negative interaction controls. Following the *Leishmania* predictions, protein pairs with response values above this threshold were selected and used as input for Cytoscape for network visualization and topological analysis.

Topological analysis and selection of essential proteins for the network

Most of the biological networks present free-scale topology, that is, the distribution of the number of connections for each node (degree) follows a power law, where there are few network components (nodes) with a high

degree and many network components with a low degree [72]. This feature is strongly related to the stability of the networks, as it makes them resistant to random attacks [72–74].

Other properties of biological networks are their clustering tendency, which can be reflected by the Clustering Coefficient (CC), and their small world effect, caused by having a small number of steps separating any two components of the network, which can be evaluated through the Mean Shortest Path (MSP) [74]. The evaluation of these properties in the network allows validating the data, considering that these characteristics are different from random networks [74, 75].

Thus, we used the Cytoscape software along with the Network Analyzer plugin to evaluate the networks produced based on the free-scale model proposed by Barabási and Oltvai [10]. The CC and MSP were also compared to 1000 random networks produced by the Random Network plugin (<http://apps.cytoscape.org/apps/randomnetworks>), and the differences were analyzed through empirical *p*-value.

In addition, to assess the behavior of an interaction network, some topological features can be used to select essential proteins for PIN stability. This is possible due to the relationship between the protein centrality and its role in cell survival [76–78]. In this way, the CytoHubba [79] plugin was used to calculate the Degree Centrality (DC), Betweenness Centrality (BC) and Bottleneck (BN) for each protein.

Results

Prediction of protein structures

Protein structures were predicted for 31.13 and 31.39% of the *L. braziliensis* and *L. infantum* proteomes, respectively, by at least one of the modeling tools (Table 1). About those sets of predictions, approximately 4% of both proteomes obtained structures with values referring to free energy and stereochemical properties in according to the thresholds recommended by the evaluation tools. With the use of the structural refinement tool, the percentages of accepted models raised to 8.11 and 8.56% for *L. braziliensis* and *L. infantum* proteomes, respectively (Table 1).

The use of multiple structure prediction tools allowed predicting structures for a reasonable quantity of proteins. Thus, based on the accepted models in according to the thresholds used, and in order to select the most accurate predicted three-dimensional structure, we selected for each protein with more than one accepted model, the predicted structure with the lowest free energy and highest percentage of torsion angles in the most favorable region of the ramachandran plot (Table 2).

Table 1 Total protein structure predicted by each program

Species	Total proteome	Align2d/Modeller		Mafft/Modeller		Modpipe		Mholline		Phyre2		Total	
		p.s	a.s	p.s	a.s	p.s	a.s	p.s	a.s	p.s	a.s	p.s	a.s
<i>L. braziliensis</i>	8357	518	163	518	165	1529	479	1858	63	1374	259	2604	681
<i>L. infantum</i>	8239	512	181	512	176	1579	504	1856	74	1415	255	2587	708

p.s total of Predicted Structures

a.s total of Accepted Structures (Structures with values referring to free energy and stereochemical properties according to the thresholds determined by the standardized Dope algorithm and by the Procheck tool)

Performance evaluation of machine learning models

As presented in the methodology section, machine learning algorithms were evaluated against positive and negative interaction datasets used as controls. Based on this analysis, the *gbm* (model available at <https://crhislane-wixsite.com/ppinleishmania>) technique showed a better performance when compared to other machine learning algorithms, obtaining an AUC = 0.88 (Fig. 1). We were able to improve the *gbm* model by setting “shrinkage=0.1, n.trees=100, interaction.depth=3, bag.fraction=0.5, train.fraction=0.8, n.minobsinnode=10, cv.folds = 5, class.stratify.cv = TRUE” parameters. The *gbm* algorithm calculates a response value ranging from 0 to 1, for which a minimum threshold of 0.46 has been determined based on controls to indicate interaction between the proteins. This threshold has recall equal to 0.69, specificity equal to 0.88 and precision equal to 0.83.

The use of the response value generated by *gbm* model to evaluate the outcome of the interactions also showed a higher performance when compared to the analysis using only the ppi-score, which obtained an AUC of 0.72, recall equal to 0.65 and precision equal to 0.68. Even when we compared to other studies that used the same interaction prediction tools [28, 63], our recall and precision of response value were higher.

Prediction of protein interaction

The interaction prediction was performed between proteins that shared the same cell compartment (Table 3). Two proteins of *L. infantum* (LinJ.30.2360, LinJ.31.2540) were the only ones classified in the Peroxisome and Golgi locations, respectively. In this way, they did not share location with any other protein incorporated in the study, making the interaction test impossible, and it was necessary to exclude them from the study. Proteins that had more than one cellular location were maintained in more than one group. In this way, groups of proteins were submitted to the two techniques of

interaction prediction using docking, resulting in 82,494 and 88,055 tested interactions by rigid-body method for *L. braziliensis* and *L. infantum*, respectively. Of these, 19,808 and 21,029 interactions were also tested through the template-based method.

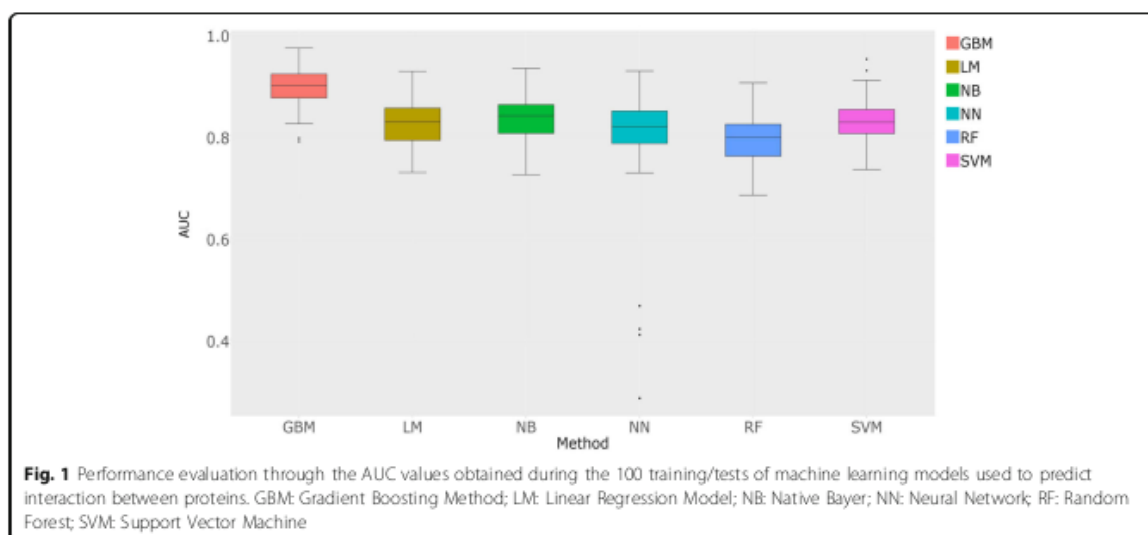
As previously stated, interactions predicted by the template-based method were classified as potential interactions when the global energy binding score was less than or equal to 0. For the interactions predicted by the body-rigid method, due to the amount of solutions generated for each pair of proteins (10,800 solutions), we used clustering tools from which ppi-score and calibur-score were obtained. These values were then submitted to the machine learning algorithm model *gbm*, and the interactions with a response value equal to or greater than 0.46 were described as potential interaction. It worth to remind that *gbm* training model and the threshold of response value were defined based on two sets of experimentally solved protein structures; one set of proteins known to be interacting and one set of non-interacting proteins. Therefore, the leishmania protein interaction predictions are based on validated data (See Methods section).

To predict a highly precise interaction network, we apply a meta-approach, using the consensus between both docking methodologies, as proposed by Ohue et al. [28]. Following this methodology, only interactions described as possible by both methodologies were used to build the protein networks (Table 4) (Fig. 2) (Additional files 1 and 2). It is understandable that true positives can be lost applying this meta-approach, but our main goal here was the reduction of potential false positives, thereby increasing the quality of the protein interaction networks generated.

A protein network is characterized by a graph composed of nodes representing the proteins and the edges representing the physical interactions between proteins. The networks predicted here had their quality assessed

Table 2 Total protein per tool with lower free energy structure and higher percentage of torsion angles in the most favorable region of the ramachandran plot

Species	Align2d/Modeller	Mafft/Modeller	Modpipe	Mholline	Phyre2	Total
<i>L. braziliensis</i>	88	44	336	28	185	681
<i>L. infantum</i>	96	47	344	31	190	708



through comparison against 1000 random networks, where the values of Clustering Coefficient and Mean Shortest Path were obtained (Table 5). The Clustering Coefficient, which measures the density of interactions close to a protein in the network [80], was significantly higher in the networks of *Leishmania* species than in random networks. The same behavior was observed when the Mean Shortest Path was evaluated. Both measures are related to the robustness of the network, and the comparisons with random networks suggest the predicted networks are compatible with biological networks, and they are not a product of random insertion of interactions.

In order to quantify the new information generated by this methodology and to improve the protein interaction networks of *Leishmania* species, we incorporated the network predicted here with the networks predicted by Rezende and collaborators through the Interolog Mapping method [38] (Fig. 3). The protein interaction networks, resulting from the merging of the networks predicted by both methodologies, continued to present a behavior consistent with biological networks, and different from random networks (Table 6).

With the merged networks, it was possible to verify the use of structural information added 201 and 181 proteins to the network of *L. braziliensis* and *L. infantum*, respectively. In addition, it was possible to predict 6002 interactions for *L. braziliensis* and 7119 interactions for *L. infantum*, which were not obtained by the

Interolog Mapping method, increasing the knowledge about the interactomes of these species.

Topological analysis for selection of essential proteins

The analysis of the topological context of each protein was performed in the predicted networks through structural information (NPTSI) and in the networks predicted through Interolog Mapping (NPTIM) [38], separately, as well in the merged network (MN), resulting from the interaction data obtained in both methodologies. The topological index local-based method Degree was calculated for all the proteins present in the networks (Additional file 3), being possible to select the 20 proteins with the highest number of direct interactions with neighbor proteins (Fig. 4).

Through the degree of connectivity, it was possible to observe that the insertion of new proteins and interactions forming MN did not significantly alter the list of hub proteins presented in the NPTIM (Fig. 4), since the 3 proteins that were substituted among the 20 most connected proteins, remained between the 25 most connected proteins in MN (Additional file 3).

Global-based methods were also used to evaluate the topological context of each protein considering the shortest path. For this, the metrics BottleNeck and Betweenness Centrality were calculated for all proteins in the networks (Additional files 4 and 5). Obtaining such values allowed us to observe that, in contrast to the

Table 3 Total proteins in each cell compartment predicted by the Wolfpsort tool

Species	cytoskeleton	cytosol	endoplasmic reticulum	extracellular	mitochondria	nuclear	plasma membrane
<i>L. braziliensis</i>	23	351	5	66	150	138	17
<i>L. infantum</i>	19	389	2	62	141	134	20

Table 4 Interactions described as possible by each tool and consensus

Species	Megadock	Prism	Consensus
<i>L. braziliensis</i>	56,520	9216	6198
<i>L. infantum</i>	64,163	10,032	7391

Degree, the insertion of new information into PINs changed drastically the list of bottlenecks proteins (Figs. 5 and 6).

From the evaluation of both metrics of global centrality in MN, it was possible to consider the consensus proteins between the metrics, that is, the bottlenecks proteins selected from the calculations of BC and BN, being these a total of 8 bottlenecks proteins (LbrM.04.0500; LbrM.07.1200; LbrM.10.0880; LbrM.13.0260; LbrM.20.0710; LbrM.20.1010; LbrM.22.0110; LbrM.25.2330) among the 20 evaluated by each metric in the *L. braziliensis* network and 5 bottlenecks proteins (LinJ.10.0830; LinJ.13.0280; LinJ.22.0013; LinJ.27.0620; LinJ.27.2260) in the *L. infantum* network.

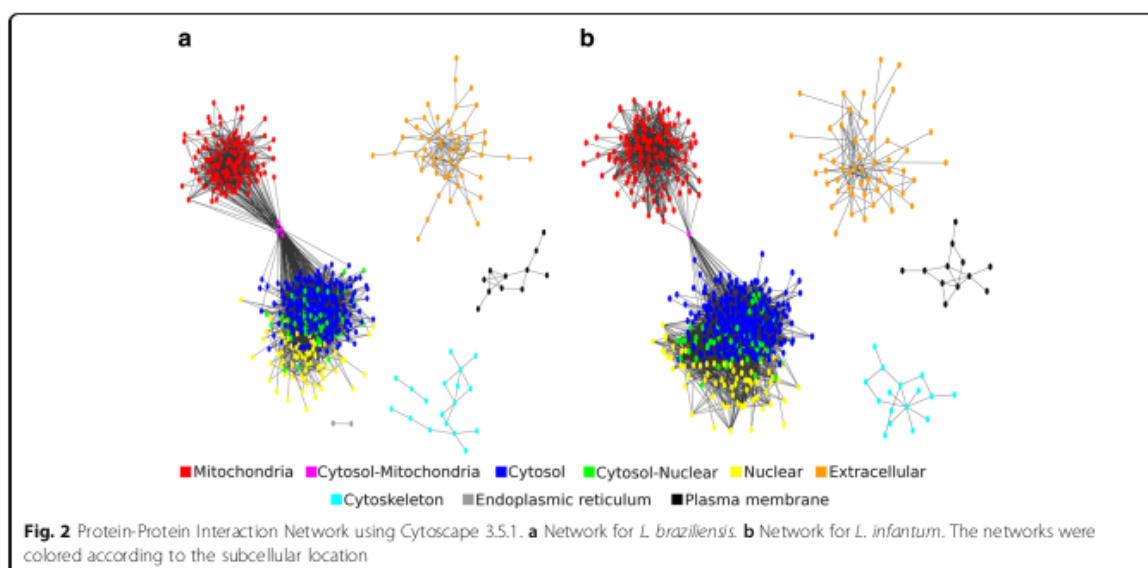
The search for the intersection between the nodes selected by all the evaluated metrics (BN, BC and DC) allowed to identify the proteins that present local and global centrality characteristics, being these (LbrM.20.1010 and LbrM.22.0110) in the *L. braziliensis* network and (LinJ.22.0013 and LinJ.27.2260) in the *L. infantum* network. Proteins with this level of centrality were described by Han Jing-Dong and collaborators as “date hubs”. These ones are responsible for the dynamics of the networks, since they are related to the ability of a protein to interact with different proteins at different times [81].

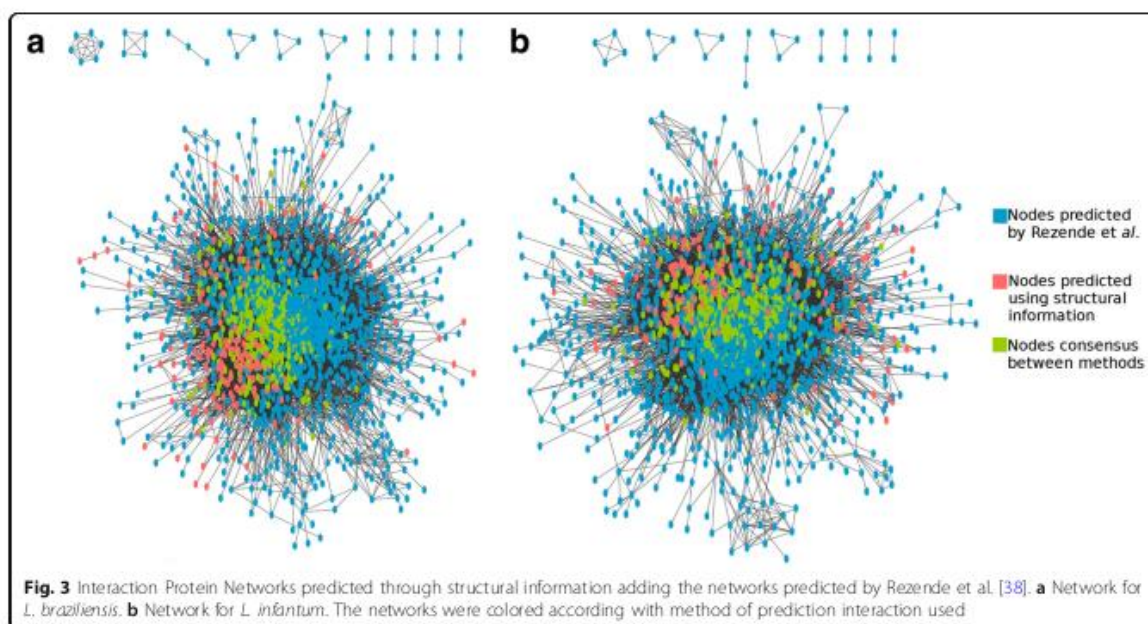
Table 5 Evaluation of the topological characteristics of protein interaction networks predicted through structural information

<i>L. braziliensis</i>			
Scale free model	Correlation	R^2	
	0.671	0.795	
Comparison with random networks			
Measure	Predicted network	Random network	P-value
Clustering Coefficient	0.212	0.161 ± 0.005	$p < 0.05$
Mean Shortest Path	2.680	2.510 ± 0.007	$p < 0.05$
<i>L. infantum</i>			
Scale free model	Correlation	R^2	
	0.751	0.811	
Comparison with random networks			
Measure	Predicted network	Random network	P-value
Clustering Coefficient	0.233	0.169 ± 0.004	$p < 0.05$
Mean Shortest Path	3.000	2.488 ± 0.006	$p < 0.05$

Discussion

L. braziliensis and *L. infantum* are the main species causing leishmaniasis in Brazil. They belong to different subgenera (*Viania* and *Leishmania*, respectively) defined by Lainson and Shaw [82]. Therefore they present some evolutionary differences which can be observed on the clinical disease they can cause. Those differences can be described as the presence of retrovirus in *Viannia* subgenus, which can reflect in the metastatic ability of *L. braziliensis* [83], the different profiles of aneuploidy for both subgenera, which provide a different number of chromosome copy, and can be related with genes expression regulation [84] and drug resistance [85].





Therefore, not just because both species are important pathogens in Brazil, they were also selected here because they can illustrate the difference between the subgenera in the context of protein interactions.

The prediction of protein interaction network based on structural information is a very challenging process, especially when it is applied to species with little structural information obtained experimentally. This is the current reality of *L. braziliensis* and *L. infantum* that at the Uniprot [86] database have only 7 and 10 proteins

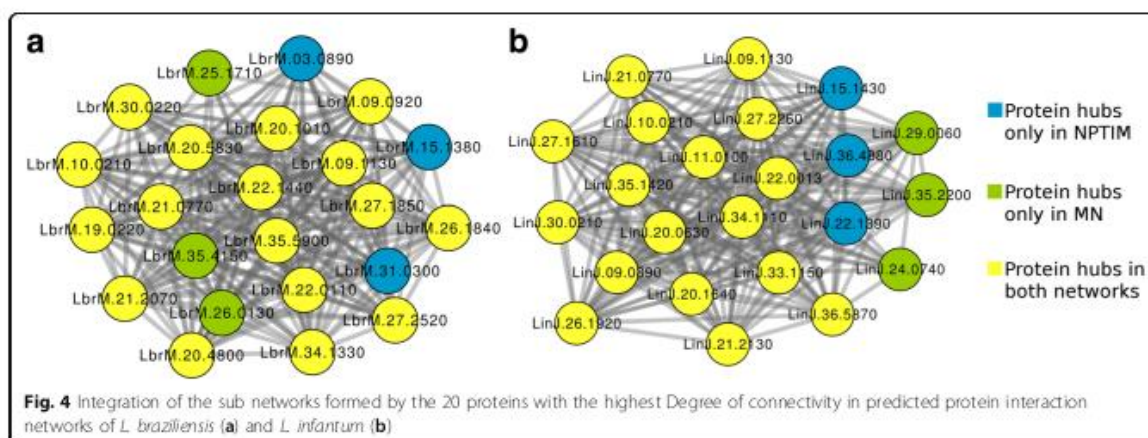
Table 6 Evaluation of the topological characteristics of the protein interaction networks predicted through structural information and merged to the networks predicted by Rezende et al. [38]

<i>L. braziliensis</i>			
Scale free model	Correlation	R^2	
	0.905	0.832	
Comparison with random networks			
Measure	Predicted network	Random network	<i>P</i> -value
Clustering Coefficient	0.381	0.144 ± 0.002	<i>p</i> < 0.05
Mean Shortest Path	2.832	2.555 ± 0.003	<i>p</i> < 0.05
<i>L. infantum</i>			
Scale free model	Correlation	R^2	
	0.917	0.837	
Comparison with random networks			
Measure	Predicted network	Random network	<i>P</i> -value
Clustering Coefficient	0.381	0.149 ± 0.002	<i>p</i> < 0.05
Mean Shortest Path	2.817	2.537 ± 0.003	<i>p</i> < 0.05

with available structures, respectively. However, the increasing availability of different computational tools has enabled the protein structural prediction in large-scale, which allowed this study to provide a promising number of predicted protein structures for *Leishmania* species.

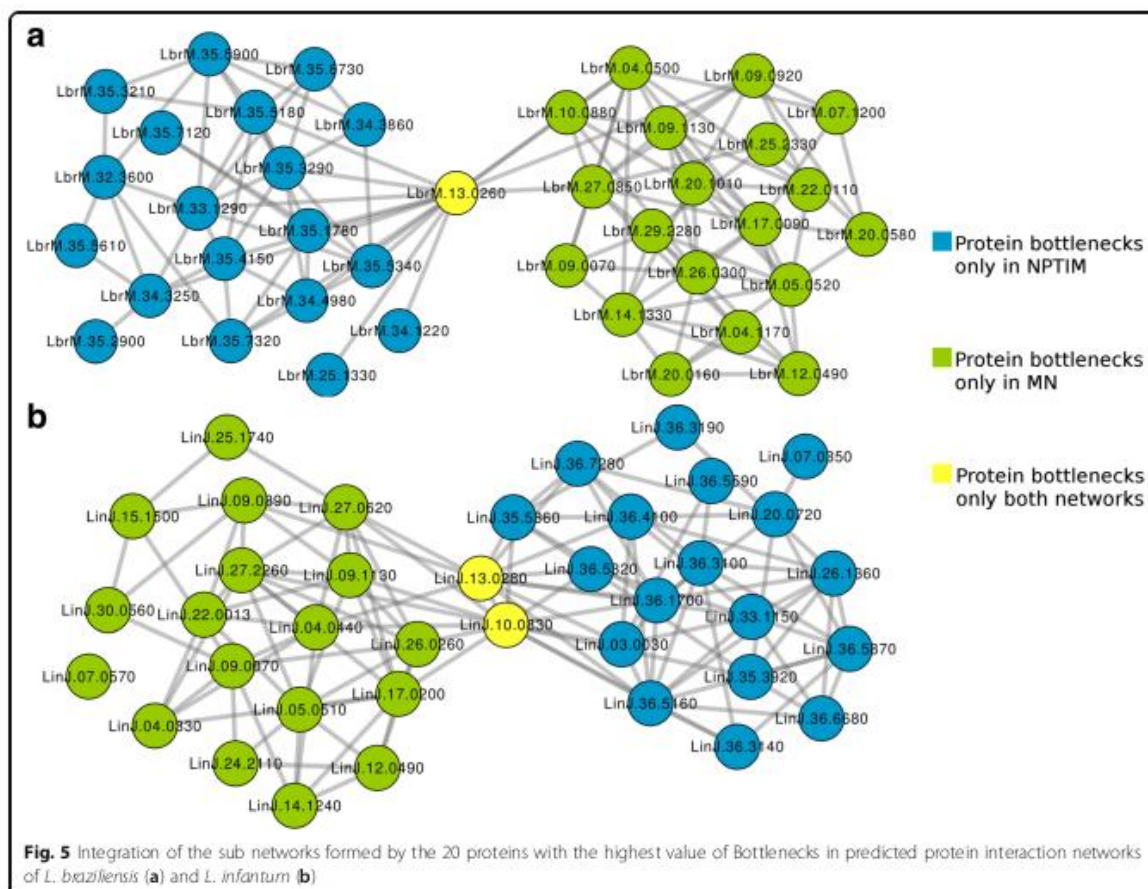
Even using a set of parameter values to guarantee models with high quality, we know models might be different from native structure of their proteins. However, as it has been demonstrated by several studies [87–92], the comparative modeling used in this study and the use of sequence similarity are methodologies that provide relevant information for prediction of protein interaction, and they are a helpful alternative approaches to structural biology, being able to provide structural representatives for a large amount of unresolved structure proteins, as it was seen for the data obtained for leishmania.

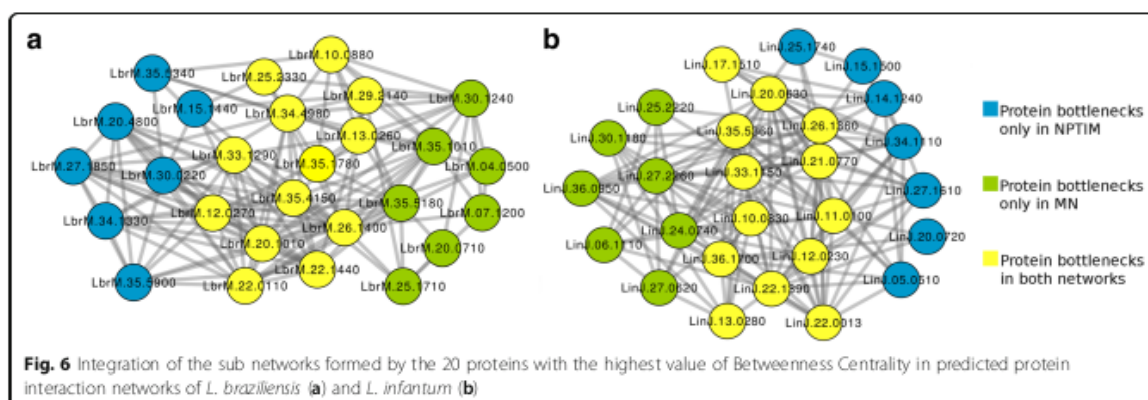
Obtaining three-dimensional (3D) structures for *Leishmania* proteins opens a parallel path for functional prediction and discovery of new potential targets for drugs based on structural features [93, 94]. This is possible because the function conservation is strictly associated with conservation of the 3D structure [95]. In addition, the availability of these structures allows a search for druggable regions that can be used to design new drugs. Furthermore, with the protein interaction information, it is possible to identify if the druggable regions are part of protein interaction interfaces, and therefore, they can be used to interrupt a protein interaction, and causing damage in the parasite.



Among all the possibilities that can be reached from obtaining protein structures, the prediction of interaction networks provides invaluable structural details for understanding several biological processes [96]. This applicability was first used in 2006 by Kim, P. M. et al.

where it was possible to identify structural characteristics of interaction in hubs proteins that could not be identified by methodologies based on sequence [97]. However, even experimental techniques for determination of protein interaction on a large scale are subject





to systematic errors and may produce false positives. Similarly, computational docking methods are sometimes unable to distinguish incompatible complexes [27]. To reduce the possibility of errors, we employed machine learning models, trained using positive and negative controls of high confidence. As a result, we obtained a significant difference of AUC, recall and precision when compared to the use of only the affinity values produced by the docking tool. The performance of the *gbm* model was also superior even when compared to the meta-approach applied by Ohue et al. [28].

The classification of interactions as true performed by the *gbm* model produced a network for each species of *Leishmania* addressed in this study. Both networks were evaluated against their topological characteristics, where it was possible to verify their robustness, features compatible with biological networks, and important differences between predicted networks and random networks. One of the characteristics found was the free-scale nature (Table 5), caused by the presence of many proteins that perform few interactions and few proteins, denominated hubs, which perform many interactions. However, this characteristic was improved when the networks predicted through the docking method were incorporated into the networks predicted through the Interolog Mapping method (Table 6). This behavior could be observed because the network constructed here represents only a subset of the true interactome, and as demonstrated by Stumpf and collaborators, the prediction of a network from a small subset can cause a significant deviation of the power law [98].

This free-scale nature of a PIN, as addressed in the Methods section, is strongly related to resilience of network, allowing it to withstand random attacks. This resilience is owed to the fact that the majority of proteins present into an interaction network perform few interactions, thus if they were knockdown, the impact could be not be so strong. However, this same feature makes the

network vulnerable to targeted attacks to hub proteins, which are essentials for network stability [99], because they perform a big number of interaction, and we know if we knockdown them, the organism will suffer a great impact. These proteins with higher degree of connectivity are important for cell survival because their essentiality for the transmission of intra-protein information [100, 101]. Therefore, attacking those proteins can destabilize the network causing a break in the transmission of information [36, 101], and hence, the description of such type of protein is an advantage to select targets to drug development.

The selection of hub proteins within the interaction networks was performed through Degree centrality, and as expected, the NPTSI analysis presented different set of protein hubs from those obtained from NPTIM and MN for both species (Additional file 3). This divergence is caused by the difference of protein universe contained in the compared networks. However, this behavior was not observed when compared to NPTIM and MN hub proteins, indicating that the insertion of new proteins in the NPTIM did not significantly alter their set of protein hubs (Fig. 4). This result is consistent with the preferred attachment model observed in biological networks [102]. This principle reports that proteins inserted into a real network tend to interact with proteins that already have a higher connectivity degree [103].

The preferential attachment phenomenon is extremely important for the evolutionary process of biological networks, since this process is resulting from the presence of highly conserved domains in hub proteins, and it is related to the free-scale behavior of the networks [102, 104]. The high degree of connectivity makes hub proteins essential for network maintenance, presenting a lethal phenotype upon removal of the protein [105]. This result would be attractive for the development of drugs, however, the degree of conservation of these proteins with host proteins increases the risk

of side effects, and may be this is one of the reasons of why they are not the majority of targets already used by the pharmaceutical industry [36]. This can be observed with the analysis of selected date hub proteins, which in UNIPROT showed more than 90% coverage with *Homo sapiens* proteins making difficult or even impossible to use them as targets (data not showed).

On the other hand, the use of random proteins as targets for drugs is not a plausible reality, since the deletion of non-hub proteins, for the most part, has no great impact on the phenotype [105]. As an alternative to this, some studies report that proteins that bind to protein hubs, as well as those with behavior of bottlenecks, which are proteins that alone are responsible for inter-modular interaction, are also promising targets for drugs [106–108]. Proteins with such behavior were recovered from the network using the topological indices BN and BC, generating subnetworks composed of bottlenecks (Figs. 5 and 6). From this analysis, it is possible to observe that the insertion of new proteins and interactions in the networks was able to change the list of proteins that behave like bottlenecks, contrasting with the analysis of the Degree of centrality. It is interesting to notice all analyses presented on Figs. 4, 5 and 6 showed a significant overlap between orthologs from *L. braziliensis* and *L. infantum*. This suggests the conservation of important common process in both species. In addition, this fact can be helpful in order to have the same drug molecule to treat more than one leishmania species.

Our findings are also consistent in favoring central proteins in a global context, since the orthologs proteins LbrM.13.0260 and LinJ.13.0280 reported as nonhubs-bottlenecks have low alignment coverage against STK36 protein ortholog in *H. sapiens*. These results are consistent with the use of this topological feature to select proteins for drug design.

Conclusion

From the data generated in this study, it is possible to perceive the importance of the use of multiple methodologies to be able to get the interactome of these species. This need is observed even to protein interaction data of most commonly studied species, such as *Homo sapiens*, that still presents a small coverage in front of the complete proteome [109]. Here, we use a methodology not yet applied to the proteome of *Leishmania* species and we describe a binary classification system that was capable of recognizing non-possible complexes. Our findings increase the knowledge in the structural context and the interactome of the species addressed, opening possibilities for future studies in the development or redirection of drugs.

Additional files

Additional file 1 Protein interaction network of *Leishmania braziliensis* predicted using structural information through the Docking methodology. (XLSX 145 kb)

Additional file 2 Protein interaction network of *Leishmania infantum* predicted using structural information through the Docking methodology. (XLSX 187 kb)

Additional file 3 Degree of connectivity of all proteins present in the networks used in this study. (XLSX 212 kb)

Additional file 4 Value of the Bottleneck topological index of all proteins present in the networks used in this study. (XLSX 207 kb)

Additional file 5 Value of the Betweenness Centrality topological index of all proteins present in the networks used in this study. (XLSX 290 kb)

Acknowledgements

This work was funded with grants provided by the Brazilian funding agency CNPq (478138/2013-0). CRSV received studentships from CAPES agency. The authors thank the Núcleo de Plataformas Tecnológicas of the Instituto Aggeu Magalhães for the use of Bioinformatics Platform and the Postgraduate Program in Genetics from Universidade Federal de Pernambuco.

Funding

CNPq and CAPES agency.

Availability of data and materials

The data sets generated during the study as well as the interaction prediction model are available in: <https://crhsilane.wixsite.com/ppinleishmania> and in the Additional files.

Authors' contributions

CRSV wrote and run all scripts necessary to parse the data and to automatize the execution of all tools cited in the manuscript, she also wrote the manuscript. TLC provided support to all machine learning analyses. AMR designed the project, wrote the manuscript and supervised the project. All authors read and approved the final version of this manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Microbiology Department of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE, Brazil. ²Bioinformatics Platform of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE, Brazil. ³Genetics Department of Universidade Federal de Pernambuco, Recife, PE, Brazil.

Received: 10 November 2017 Accepted: 1 March 2018

Published online: 06 March 2018

References

- Alvar J, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*. 2012;7:e35671.
- Tiuman TS, Santos AO, Ueda-Nakamura T, Filho BPD, Nakamura CV. Recent advances in leishmaniasis treatment. *Int J Infect Dis*. 2011;15:e525–32.
- García-Hernández R, Gómez-Pérez V, Castany S, Gamarró F. Fitness of *Leishmania donovani* parasites resistant to drug combinations. *PLoS Negl Trop Dis*. 2015;9:e0003704.

4. Rodrigues V, Cordeiro-da-Silva A, Laforce M, Silvestre R, Estaquier J. Regulation of immunity during visceral Leishmania infection. *Parasit Vectors*. 2016;9:118.
5. Henney A, Superti-Furga G. A network solution. *Nature*. 2008;455:730–1.
6. Vidal M, Cusick ME, Barabási A-L. Interactome networks and human disease. *Cell*. 2011;144:986–98.
7. Xia J, Benner MJ, Hancock REW. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res*. 2014;42:W167–74.
8. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. *Nature*. 2010;466:68–76.
9. Ahn Y-Y, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. *Nature*. 2010;466:761–4.
10. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2001;5:101–13.
11. Vinayagam A, et al. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal*. 2011;4:rs8.
12. Rebsamen M, Kandasamy RK, Superti-Furga G. Protein interaction networks in innate immunity. *Trends Immunol*. 2013;34:610–9.
13. Lecca P, Re A. Detecting modules in biological networks by edge weight clustering and entropy significance. *Front Genet*. 2015;6:265.
14. Vallabhajosyula RR, Chakravarti D, Lutfelli S, Ray A, Raval A. Identifying hubs in protein interaction networks. *PLoS One*. 2009;4:e5344.
15. Bonetta L. Protein-protein interactions: Interactome under construction. *Nature*. 2010;468:851–4.
16. Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y. Relating drug-protein interaction network with drug side effects. *Bioinformatics (Oxford, England)*. 2012;28:522–8.
17. Harrington ED, Jensen LJ, Bork P. Predicting biological networks from genomic data. *FEBS Lett*. 2008;582:1251–8.
18. Skrabanek L, Saini HK, Bader GD, Enright AJ. Computational prediction of protein-protein interactions. *Mol Biotechnol*. 2008;38:1–7.
19. Xia J-F, Wang S-L, Lei Y-K. Computational methods for the prediction of protein-protein interactions. *Protein Pept Lett*. 2010;17:1069–78.
20. Roberts PM. Mining literature for systems biology. *Brief Bioinform*. 2006;7:399–406.
21. Bock JR, Gough DA. Predicting protein-protein interactions from primary structure. *Bioinformatics (Oxford, England)*. 2001;17:455–60.
22. Matthews LR, et al. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or 'interologs'. *Genome Res*. 2001;11:1210–6.
23. Xia J-F, Han K, Huang D-S. Sequence-based prediction of protein-protein interactions by means of rotation forest and autocorrelation descriptor. *Protein Pept Lett*. 2010;17:137–45.
24. Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*. 1998;23:324–8.
25. Juan D, Pazos F, Valencia A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci*. 2008;105:934–9.
26. Smith GR, Sternberg MJE. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol*. 2002;12:28–35.
27. Kiel C, Beltrao P, Serrano L. Analyzing protein interaction networks using structural information. *Annu Rev Biochem*. 2005;77:415–41.
28. Ohue M, Matsuzaki Y, Shimoda T, Ishida T, Akiyama Y. Highly precise protein-protein interaction prediction based on consensus between template-based and de novo docking methods. *BMC Proc*. 2013;7:56.
29. Kar G, Gursay A, Keskin O. Human cancer protein-protein interaction network: a structural perspective. *PLoS Comput Biol*. 2009;5(12):e1000601.
30. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput Biol*. 2009;5(10):e1000550.
31. Hwang WC, Zhang A, Ramanathan M. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin Pharmacol Ther*. 2006;84:563–72.
32. Mullard A. Protein-protein interaction inhibitors get into the groove. *Nat Rev Drug Discov*. 2012;11:7–5.
33. Raman K, Chandra N. Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol*. 2008;8:234.
34. LaCount DJ, et al. A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*. 2003;438:103–7.
35. Taylor CM, et al. Targeting protein-protein interactions for parasite control. *PLoS One*. 2011;6:e18381.
36. Csérmely P, Kovács T, Kiss HM, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther*. 2013;138(3):333–408.
37. Fatoux-Ardore M, Peysselon F, Weiss A, Bastien P, Pratlong F, Ricard-Blum S. Large-scale investigation of Leishmania interaction networks with host extracellular matrix by surface plasmon resonance imaging. *Infect Immun*. 2014;82:594–606.
38. Rezende AM, Folador EL, Resende D d M, Ruiz JC. Computational prediction of protein-protein interactions in Leishmania predicted proteomes. *PLoS One*. 2012;7(12):e51304.
39. Flórez AF, et al. Protein network prediction and topological analysis in Leishmania major as a tool for drug target selection. *BMC Bioinformatics*. 2010;11:484.
40. Szklarczyk D, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;D1:43.
41. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância das Doenças Transmissíveis, and Brasil, Manual de vigilância e controle da leishmaniose visceral.
42. Brasil, Departamento de Vigilância das Doenças Transmissíveis, Secretaria de Vigilância em Saúde, and Ministério da Saúde, Manual de vigilância da leishmaniose tegumentar. 2015.
43. T. Ishida and K. Kinoshita, "Prediction of disordered regions in proteins based on the meta approach," *Bioinformatics*. 2008;24(11):1344–8.
44. H. K. Saini and D. Fischer. Meta-DP: domain prediction meta-server. *Bioinformatics*. 2005;21(12):2917–20.
45. Aslett M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*. 2007;35:D457–62.
46. Berman HM, et al. The protein data bank. *Nucleic Acids Res*. 1998;28(1):235–42.
47. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2014;47:5.6.1–32.
48. Pieper U, et al. "ModBase, a database of annotated comparative protein structure models, and associated resources". *Nucleic Acids Res*. 2011;39:D465–D474.
49. Capriles PVSZ, Guimarães ACR, Otto TD, Miranda AB, Dardenne LE, Degreve WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment. *BMC Genomics*. 2010;11:610.
50. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845–58.
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
52. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–66.
53. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48(3):443–53.
54. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics (Oxford, England)*. 2005;21(7):951–60.
55. Zhang J, Barz B, Zhang J, Xu D, Kosztin I. Selective refinement and selection of near-native models in protein structure prediction. *Proteins*. 2015;83(10):1823–35.
56. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science (New York, NY)*. 2005;309(5742):1868–71.
57. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J*. 2011;101(10):2525–34.
58. R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. "PROCHECK: a program to check the stereochemical quality of protein structures." *J Appl Crystallogr*. 1993;26 (2).
59. Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 2006;15(11):2507–24.
60. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Current Protoc Protein Sci*. 2016;86:2.9.1–2.9.37.

61. Horton P, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007;35(Web Server issue):W585–7.
62. Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protoc.* 2011;6(9):1341–54.
63. Ohue M, Matsuzaki Y, Uchikoga N, Ishida T, Akiyama Y. MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data. *Protein Pept Lett.* 2012;21(8):766–78.
64. Hubbard S, Thornton J. "NACCESS, Computer Program," Department of Biochemistry and Molecular Biology. University College London. 1993.
65. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins.* 2004;56(1):143–56.
66. Mashach E, Nussinov R, Wolfson HJ. FiberDock: flexible induced-fit backbone refinement in molecular docking. *Proteins.* 2010;78(6):1503–19. <http://cosbi.ku.edu.tr/prism/targets.php>. Accessed 4 May 2017.
67. Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y. In silico screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis. *J Bioinforma Comput Biol.* 2009;7(6):991–1012.
68. Li SC, Ng YK. Calibur: a tool for clustering large numbers of protein decoys. *BMC bioinformatics.* 2010;11:25.
69. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. *Proteins.* 2010;78(15):3111–4.
70. Blohm P, et al. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.* 2014;42(D1):D396–D400.
71. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys.* 2002;74(1):47–97.
72. Albert R. Scale-free networks in cell biology. *J Cell Sci.* 2005;118(Pt 21):4947–57.
73. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2002;5:101–13.
74. Tosadori G, Bestvina I, Spoto F, Laudanna C, Scardoni G. Creating, generating and comparing random network models with network randomizer. *F1000Research.* 2016;5:1–12.
75. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol.* 2005;22(4):803–6.
76. Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. *J Biomed Biotechnol.* 2004;2005(2):96–103.
77. Wuchty S, Stadler PF. Centers of complex networks. *J Theor Biol.* 2003; 223(1):45–53.
78. Lin C-Y, Chin C-H, Wu H-H, Chen S-H, Ho C-W, Ko M-T. Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology. *Nucleic Acids Res.* 2008;36(Web Server issue):W438–43.
79. Friedel CC, Zimmer R. Inferring topology from clustering coefficients in protein-protein interaction networks. *BMC bioinformatics.* 2006;7:519.
80. Han J-D, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature.* 2004;430(6995):88–93.
81. Lainson R, Shaw JJ. Leishmaniasis in biology and medicine and 1987, "Evolution, classification and geographical distribution". In: The leishmaniasis in biology and medicine. Orlando: Academic Press; 1986. p. 1–120.
82. Scott P. Leishmania—a parasitized parasite. *N Engl J Med.* 2011;364(18):1773–4.
83. Rogers MB, et al. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Res.* 2011;21(12):2129–42.
84. Downing T, et al. Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* 2011;21(12):2143–56.
85. Consortium U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2016;45(D1):D158–69.
86. Zhang QC, et al. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature.* 2012;490(7421):556–60.
87. Singh R, Park D, Xu J, Hosur R, Berger B. Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. *Nucleic Acids Res.* 2010;38(Web Server issue):W508–15.
88. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. *Proc Natl Acad Sci U S A.* 2000;99(9):5896–901.
89. Shah PK, Aloy P, Boik P, Russell RB. Structural similarity to bridge sequence space: finding new families on the bridges. *Protein Sci.* 2005;14(5):1305–14.
90. Fan H, Irwin JJ, Sali A. Virtual ligand screening against comparative protein structure models. *Methods Mol Biol (Clifton, N.J.).* 2010;819:105–26.
91. Dong GQ, et al. Prediction of substrates for glutathione transferases by covalent docking. *J Chem Inf Model.* 2014;54(6):1687–99.
92. Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery. *Drug Discov Today.* 2005;10(13):895–907.
93. Cherkasov A, Ho Sui SJ, Brunham RC, Jones SJM. Structural characterization of genomes by large scale sequence-structure threading: application of reliability analysis in structural genomics. *BMC Bioinformatics.* 2004;5:101.
94. Marsden RL, Lee D, Maibaum M, Yeats C, Orengo CA. Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space. *Nucleic Acids Res.* 2006;34(3):1066–80.
95. Tuncbag N, Keskin O, Nussinov R, Gursoy A. Prediction of protein interactions by structural matching: prediction of PPI networks and the effects of mutations on PPIs that combines sequence and structural information. *Methods Mol Biol (Clifton, NJ).* 2015;1558:255–70.
96. Kim PM, Lu LJ, Xia Y, Gerstein MB. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science.* 2006;314(5807):1938–41.
97. Stumpf MPH, Wiuf C, May RM. Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci U S A.* 2003; 102(12):4221–4.
98. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev Mod Phys.* 2000;74:47–97.
99. Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. *Nature.* 1999;411(6833):41–2.
100. Pavlopoulos GA, et al. Using graph theory to analyze biological networks. *BioData Min.* 2011;4:10.
101. Eisenberg E, Levanon E. Preferential attachment in the protein network evolution. *Phys Rev Lett.* 2001;91(13):1–4.
102. O. Alzate. "Neuroproteomics". In: Neuroproteomics. Boca Raton: CRC press; 2009. p. 1–16.
103. Holland DO, Shapiro BH, Xue P, Johnson ME. Protein-protein binding selectivity and network topology constrain global and local properties of interface binding networks. *Sci Rep.* 2017;7(1):5631.
104. Kafri R, Dahan O, Levy J, Pilpel Y. Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc Natl Acad Sci U S A.* 2008;105(4):1243–8.
105. T. Korcsmáros, M. S. Szalay, C. Bode, I. A. Kovács, and P. Csérmely. "How to design multi-target drugs: target search options in cellular networks". *Arxiv preprint q-bio/0703010.* 2005;2:6.
106. Farkas IJ, et al. Network-based tools for the identification of novel drug targets. *Sci Signal.* 2011;4(173):pt3.
107. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. "The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics". *PLoS Comput Biol.* 2007;4:e59.
108. Rolland T, et al. A proteome-scale map of the human interactome network. *Cell.* 2014;159(5):1212–26.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



**5 CAPÍTULO II – SYSTEMATIC *IN SILICO* EVALUATION OF *LEISHMANIA SPP.*
PROTEOMES FOR DRUG DISCOVERY**

Crhisllane Rafaele dos Santos Vasconcelos^{1,2,3} e Antonio Mauro Rezende^{1,2,3}

1Microbiology Department of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE Brazil

2Bioinformatics Plataform of Instituto Aggeu Magalhães – FIOCRUZ, Recife, PE Brazil

3Genetics Department of Universidade Federal de Pernambuco, Recife, PE Brazil

**Systematic *in silico* evaluation of *Leishmania spp.*
proteomes for drug discovery**

Artigo publicado na revista Frontiers in Chemistry -
Research Topic: Cheminformatics Approaches in Drug
Discovery for Neglected Tropical Diseases.

**Recife
2020**



Systematic *in silico* Evaluation of *Leishmania* spp. Proteomes for Drug Discovery

Chrisllane Rafael dos Santos Vasconcelos^{1,2} and Antonio Mauro Rezende^{1,2*}

¹ Bioinformatics Platform, Microbiology Department, Instituto Aggeu Magalhães, Recife, Brazil, ² Postgraduate Program in Genetics, Genetics Department, Universidade Federal de Pernambuco, Recife, Brazil

OPEN ACCESS

Edited by:

Marcus Scotti,
Federal University of Paraíba, Brazil

Reviewed by:

Rubens Monte Neto,
René Rachou Institute, Oswaldo Cruz
Foundation (FIOCRUZ), Brazil
Maria Paola Costi,
University of Modena and Reggio
Emilia, Italy

*Correspondence:

Antonio Mauro Rezende
antonio.rezende@fiofcrz.br

Specialty section:

This article was submitted to
Medicinal and Pharmaceutical
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 16 September 2020

Accepted: 24 March 2021

Published: 27 April 2021

Citation:

dos Santos Vasconcelos CR and
Rezende AM (2021) Systematic
in silico Evaluation of *Leishmania* spp.
Proteomes for Drug Discovery.
Front. Chem. 9:607139.
doi: 10.3389/fchem.2021.607139

Leishmaniasis is a group of neglected infectious diseases, with approximately 1.3 million new cases each year, for which the available therapies have serious limitations. Therefore, it is extremely important to apply efficient and low-cost methods capable of selecting the best therapeutic targets to speed up the development of new therapies against those diseases. Thus, we propose the use of integrated computational methods capable of evaluating the druggability of the predicted proteomes of *Leishmania braziliensis* and *Leishmania infantum*, species responsible for the different clinical manifestations of leishmaniasis in Brazil. The protein members of those proteomes were assessed based on their structural, chemical, and functional contexts applying methods that integrate data on molecular function, biological processes, subcellular localization, drug binding sites, druggability, and gene expression. These data were compared to those extracted from already known drug targets (BindingDB targets), which made it possible to evaluate *Leishmania* proteomes for their biological relevance and treatability. Through this methodology, we identified more than 100 proteins of each *Leishmania* species with druggability characteristics, and potential interaction with available drugs. Among those, 31 and 37 proteins of *L. braziliensis* and *L. infantum*, respectively, have never been tested as drug targets, and they have shown evidence of gene expression in the evolutionary stage of pharmacological interest. Also, some of those *Leishmania* targets showed an alignment similarity of <50% when compared to the human proteome, making these proteins pharmacologically attractive, as they present a reduced risk of side effects. The methodology used in this study also allowed the evaluation of opportunities for the repurposing of compounds as anti-leishmaniasis drugs, inferring potential interaction between *Leishmania* proteins and ~1,000 compounds, of which only 15 have already been tested as a treatment for leishmaniasis. Besides, a list of potential *Leishmania* targets to be tested using drugs described at BindingDB, such as the potential interaction of the DEAD box RNA helicase, TRYR, and PEPCK proteins with the Staurosporine compound, was made available to the public.

Keywords: drug repurposing, protein interaction network, drug targets, leishmaniasis, *Leishmania*

INTRODUCTION

Leishmaniasis is a set of diseases caused by flagellated protozoa of the genus *Leishmania*, which can infect mammals since the Mesozoic era (Akhoundi et al., 2016). They are transmitted to humans and other mammals by the sting of phlebotomine of the genus *Phlebotomus* and *Lutzomyia*. Also, those parasites have a wide range of hosts, including canids, rodents, and humans (Maroli et al., 2013). Currently, with an estimated 1.3 million annual cases and 350 million people at risk of contracting the diseases, leishmaniasis is the cause of more than 20,000 deaths per year and remains a global public health problem (World Health Organization/Country Office for India, 2017; Gradoni, 2018).

World Health Organization (WHO) recommendations used to treat leishmaniasis are dependent on geographic region, as treatment failures are related to host genetics, their nutritional states, and to the selection of resistant strains and species present in certain regions. In general, the treatment involves pentavalent antimony, amphotericin B, paromomycin, pentamidine, and miltefosine, which are used in single drug or combined treatments. However, most of these drugs have unavoidable and sometimes irreversible side effects, such as pentamidine, which can cause insulin-dependent diabetes mellitus (World Health Organization, 2010; Nagle et al., 2014). The high incidence of these side effects has led to patients abandoning treatment or using the drugs incorrectly, causing the selection of resistant parasites (Ghorbani and Farhudi, 2018). Drugs with milder side effects, such as miltefosine, have other downsides such as the long half-life that results in an extensive drug accumulation in the body and the teratogenic effects observed even when applied in dosages below the therapeutic dose (Dorlo et al., 2008, 2012; Rahman et al., 2017). In addition, other points that should be considered regarding the use of miltefosine as the main treatment are the high frequency of drug resistance, low availability and high cost, incompatible with poor endemic areas (Sunyoto et al., 2018). In this way, it is possible to understand the urgent need to apply methodologies capable of identifying new therapeutic targets and drugs with potential leishmanicide.

The process of discovering a new drug can be divided into a phenotypic screening or target-based approaches (Swinney, 2013). In phenotypic screening it is possible to select compounds with cytotoxicity against the whole cell, measure cellular function without imposing the need for prior knowledge about the targets; While in target-based approaches the process of developing a new drug has as its key point the discovery of a target or a set of targets (Gilbert, 2013; Schenone et al., 2013; Swinney, 2013). This knowledge about the target, different from what occurs in the phenotypic screening method, allows access to crucial information for the optimization of compounds, such as knowledge about the target/drug binding site (Gilbert, 2013; Li and Kang, 2017). However, target selection is often based on biological evidence which results in the analysis of a small target group, exploring only a small portion of the total proteome and neglecting potentially therapeutically important new drug targets (Patel et al., 2013; Schenone et al., 2013).

Thus, when the goal is to select new targets for drug development, where large-scale analyzes of molecules and

compounds are performed, computational methods are presented as an initial choice, providing methodologies capable of identify targets from a proteome starting from the recognition of patterns present in compounds and their already known targets (Schenone et al., 2013), being a powerful methodology for integrating data, accelerating the first stages of the study, and reducing the cost of the process (Sanseau et al., 2012; Patel et al., 2013; Schenone et al., 2013). The detection of potential pharmacological targets by computational methods can be focused on the target or the ligand using data extracted from sources such as protein interaction network, sequence or structural similarity of proteins against approved drug targets, the similarity between structures of the ligands and their side effects, druggability regarding protein structure and gene expression (Sliwoski et al., 2014). However, to exploit the potential of such data, the integration of all available information must be carried out effectively.

In this context, we applied computational analyses integrating information in biological and pharmacological contexts to infer interactions between the proteome of *Leishmania infantum* and *Leishmania braziliensis* with available drugs (BindingDB compounds). These species are responsible for the different clinical manifestations of leishmaniasis in Brazil, a country that is among the global regions with the highest numbers of cases (Ministério Da Saúde, 2006; Brasil et al., 2017, 2019; Pan American Health Organization, 2018). Thus, the possibility of drug repurposing could be verified and new information about the druggable proteome of *Leishmania* species was recovered.

MATERIALS AND METHODS

Data

L. braziliensis and *L. infantum* predicted protein sequences were obtained from the TriTrypDB database (release 9.0) (Aslett et al., 2010), while the protein tertiary structures, and data regarding protein interactions for each above proteome, were retrieved from a database created in a previous work of our group (dos Santos Vasconcelos et al., 2018).

Data about known interactions between proteins without restricting by any specific organism and drugs were obtained from the BindingDB database (Liu et al., 2007) on September 5, 2019. This repository was chosen as it integrates interaction data from several public databases and assigns an affinity value for a pair of proteins and drugs. With protein IDs available at BindingDB, it was possible to obtain the amino acid sequences and three-dimensional structures of each protein target using the Uniprot database (Bairoch et al., 2005) and *Protein Data Bank* (PDB) (Berman et al., 2000), respectively. Proteins without X-ray crystallography three-dimensional structures at PDB had their three-dimensional models retrieved from the SwissModel repository (Bienert et al., 2017), which uses automated comparative modeling techniques to predict protein structures. All PDB format files were processed to remove ligands. This was necessary once a protein can not have a bound ligand when it is used as input to tools that perform searches by regions (pockets) able to interact with drugs. Therefore, to identify pockets, the structures must be free of ligands.

Construction of Similarity Matrices

To evaluate the proteins from Binding DB regarding their chemical (the physico-chemical properties of three-dimensional structure) and biological (molecular function, biological processes in which the protein is involved and its subcellular location) contexts, a methodology similar to that described by Patel et al. was used, which consists of classifying potential targets based on the annotation of characteristics about the three-dimensional structure, druggability, functional class and subcellular location (Patel et al., 2013). However, the methodology applied here proposes a direct comparison of each characteristics annotated for a proteome with the same characteristics extracted from already known drug targets. Thus, features from structural, functional, and sequence levels were extracted from the recovered targets. The same features were extracted for the *Leishmania* species proteomes to compare them to targets recovered from BindingDB, and thus to build similarity matrices between described targets and *Leishmania* proteins.

Therefore, considering 2 proteins t_i and p_j : t_i is a protein described as a reference drug target from Binding DB; and p_j is a protein belonging to the *L. braziliensis* or *L. infantum* proteomes. We constructed 5 similarity matrices S^x where x indicates the information used to construct the matrix, and S_{ij} represents the similarity relationship between t_i and p_j . Hence, similarity matrices were built considering the similarity between the binding sites of drugs (S^b), the druggability value between the most similar binding sites (S^d), the subcellular location (S^s), the biological processes that proteins were involved (S^p), and the molecular functions of the proteins (S^m).

The first two matrices (S^b and S^d) were built using the information on ligandability at the protein structural context. Ligandability is the ability of a protein to bind with high affinity to small drug-like compounds (Hussein et al., 2016). The ligandability prediction using the three-dimensional structure is performed by evaluating physicochemical parameters such as polar and non-polar surface area, surface complexity, and pocket dimensions (drug binding site) (Le Guilloux et al., 2009; Volkamer et al., 2010). Thus, to produce both matrices cited above, the BindingDB target structures and the *L. braziliensis* and *L. infantum* protein structures were submitted to the binding site detection algorithm present at the Fpocket tool (Le Guilloux et al., 2009). This software was chosen because the results obtained are more accurate when compared to other tools, such as SiteFinder, PocketFinder, and SiteMap (Le Guilloux et al., 2009; Schmidtke et al., 2010). With the identification of each binding site, a druggability index/score is assigned to each binding site found in the protein structures (scores values ranging from 0 to 1, with 0 assigned to pockets with a low probability of binding to a drug-like molecule, and 1 assigned to pockets with a high probability). Also, a threshold (druggability score ≥ 0.5) established by the Fpocket tool was applied to ensure the druggability of the binding sites of evaluated targets. Thus, the structure of the binding sites and their druggability index was obtained.

The predicted binding sites of each *Leishmania* spp. protein was then compared to binding sites of BindingDB protein targets

using the Pocketmatch tool version 2.1 (Yeturu and Chandra, 2008). In this way, all predicted binding sites of a p_j protein were aligned with the binding sites of the t_i protein, generating similarity values. The highest similarity value is the one selected to S_{ij}^b .

In order to construct the similarity matrix S^d , the druggability values for the most similar binding sites between proteins p_j and t_i were considered. Thus, the S_{ij}^d of this matrix was calculated using the following equation:

$$S_{ij}^d = 1 - D_{ij}$$

being D_{ij} the one-dimension Euclidean distance between the t_i and p_j druggability values.

The matrices referring to Molecular Functions and Biological Processes were built using semantic similarity between Gene Ontology (GO) terms. The GO terms of each BindingDB protein target were retrieved from the Uniprot database. For those without GO annotation, as well as for the proteomes of *L. braziliensis* and *L. infantum*, the Blast2GO tool (Conesa and Götz, 2008) was used to assign GO terms. Blast2GO is a tool for automatic functional annotation based on homology transfer (Conesa and Götz, 2008). The terms retrieved from Uniprot, as well as those annotated by Blast2GO, were submitted to the *csblgo* R package (Ovaska et al., 2008) to calculate semantic similarity between proteins p_j and t_i applying the Relevance algorithm (Schlicker et al., 2006). Hence, two matrices were produced, one containing the similarity between the terms referring to Molecular Function (S^m), and the other containing the similarity between the Biological Process terms (S^p).

The subcellular location (SL) matrix consisted of a binary matrix S^s for a pair of t_i and p_j proteins, being $S_{ij}^s = 1$ when the pair of proteins shared the same subcellular location, and $S_{ij}^s = 0$ when the proteins had different subcellular locations. The data about the subcellular location were retrieved from the Uniprot database. For BindingDB protein targets and *L. braziliensis* and *L. infantum* proteins without that information, the WolfSort tool (Horton et al., 2007) was applied to predict the subcellular localization.

Selection and Evaluation of Potential Drug Targets

The members of *Leishmania* proteomes were classified according to the degree of similarity between them and BindingDB drug targets using as basis the constructed similarity matrices. Thus, the greater the similarity of a *Leishmania* protein to a BindingDB drug target, the greater the probability of that protein being a compound target. In order to consider only the proteins most likely to interact with a drug, the following criteria were used: 1 - a *Leishmania* protein must have a binding site and its druggability similar to the same BindingDB drug target; 2 - a protein that passes the previous criterion must also be similar to the same BindingDB drug target in at least two of the other three metrics evaluated, being Molecular Function, Biological Process, or Subcellular Location. Thus, it is possible to infer that the

selected *Leishmania* protein may have the potential to bind to the same compounds that interact with the BindingDB drug target.

However, it is important to remember that even if a protein has the ability to interact with a drug, for that protein to be considered truly druggable, its interaction with a drug must generate a directed effect on the organism. Thus, a protein is potentially interesting as a drug target when the maintenance of its function is essential to keep the organism survival (Schmidtke and Barril, 2010; Desaphy et al., 2012; Hussein et al., 2016). Thus, to evaluate the essentiality of the *Leishmania* species proteins, we considered their positions in the network of protein interaction.

L. braziliensis and *L. infantum* proteins had their essentiality evaluated through their positioning in protein interaction networks (PIN). Interaction data were retrieved from previous work of our group (Rezende et al., 2012; dos Santos Vasconcelos et al., 2018) integrated with interactions data retrieved from the String database (Szklarczyk et al., 2015). To analyze the essentiality of a protein, the Cytoscape tool version 3.8.0 (Shannon et al., 2003) was used to calculate the centrality indexes: Degree Centrality, Neighborhood Connectivity, and Betweenness Centrality. These metrics reflect the centrality-lethality rule, which describes high-degree proteins as essential and can be used as good predictors of biological importance (Zotenko et al., 2008).

Due to the existence of two main distinct evolutionary stages in the biological cycle of the *Leishmania* species, and only one of these stages parasitizing humans (Bates and Rogers, 2004), the levels of gene expression between the two stages were evaluated. Thus, RNA sequencing data from *L. braziliensis* species (PRJNA494068) were retrieved from the Sequence Read Archive database (SRA) (Kodama et al., 2012), mapped and aligned with reference sequences through the Bowtie2 tool (Langmead and Salzberg, 2012). The data were then formatted using the SAMtools tool (Li et al., 2009), and the reads aligned for each gene were counted using the Htseq-count tool (Anders et al., 2015). The expression analysis between the evolutionary stages was performed using the DESeq2 library (Anders and Huber, 2010) of the R programming environment. Thus, allowing the identification of genes that are expressed in the amastigote evolutionary stage.

In order to reduce the chances of side effects, we carried out an analysis of the similarity between *Leishmania* spp. proteins selected previously and the human proteome (recovered from the Uniprot database), using the alignment tool Blastp. We used a cut-off value of 50% similarity based on the analysis of Patel et al. which showed a relationship between structure, function, and sequence for sequence similarity above 50% (Patel et al., 2013). Considering similar proteins interact with similar drugs, evaluating the similarity with human proteins may reflect the possibility of side effects (Zhou et al., 2015). All data retrieved and produced during this study were stored using a relational database model through the MySQL database manager tool. The completed database scheme and all the data produced are available as **Supplementary Material** on the <https://leishtargets.github.io/> in order to facilitate the reproducibility of this research and data retrieval to guide other projects.

RESULTS

Data

Our last access to BindingDB contained 64,025 interactions described with high-affinity value (better than 10 uM) between 3,128 proteins and 29,353 compounds. The choice for interactions with purchasable compounds was made due to the facility of their acquisition for subsequent projects with biological tests. *L. braziliensis* and *L. infantum* proteomes addressed in this study contained a total of 8,357 and 8,239 proteins, respectively. None of these proteins were found in the BindingDB database as targets with high-affinity value for a purchasable compound, and only two BindingDB protein targets belonged to the same genus addressed in this study, and they are Bifunctional dihydrofolate reductase-thymidylate synthase (LmjF.06.0860) and Pteridine reductase 1 (LmjF.23.0270), both proteins of *Leishmania major*. However, there are other proteins of the *Leishmania* genus described as interacting with compounds in BindingDB. These proteins were not recovered due to a low-affinity value with the tested compounds or due to the impossibility of buying such compound. Thus, we identified 6 proteins of the genus *Leishmania* that were tested for 40 compounds: Farnesyl pyrophosphate synthase (LmjF.22.1360); Glyceraldehyde-3-phosphate dehydrogenase-like protein (LmxM.34.4750); NAD-dependent protein deacetylase (LinJ.26.0200), Putative lanosterol 14- α -demethylase (LinJ.11.1100), N(1), N(8)-bis(glutathionyl)spermidine reductase (LinJ.05.0350) and Tubulin α chain (Q1A5Y2). Twenty-five of those compounds are present in our similarity data due to their high-affinity with other BindingDB protein targets (**Supplementary Material 1**).

Construction of Similarity Matrices Druggability of Binding Sites

For the analyses that involved the identification of protein binding sites, we recovered a total of 2,641 BindingDB protein targets with structure, of which 1,592 and 1,049 were from the PDB and SwissModel, respectively. Among the BindingDB targets recovered, 14 structures predicted by SwissModel suite had <80% coverage and/or identity when compared to the protein sequence deposited in the Uniprot; 6 protein structures (3j9m, 4u4r, 5gjr, 5j4z, 5vfl, 6ek0) could not be recovered from the PDB in an automated way due to the size of the deposited structure. In the end, 468 BindingDB proteins had no structure deposited in both repositories used.

After treatment, the atomic coordinate files (PDB files) were submitted to the Fpocket tool to predict binding sites and assign a score. We were unable to identify binding sites for 17 BindingDB protein targets. Thus, further analyses were performed only to coordinate files of the binding sites that were able of being evaluated by the PocketMatch tool. Therefore, 2,057 BindingDB protein targets displayed binding sites according to these criteria.

The total protein structures of *L. braziliensis* and *L. infantum* proteomes were 681 and 708, respectively (dos Santos Vasconcelos et al., 2018). The three-dimensional conformations produced through comparative modeling did not need to be processed to remove ligands, and all recovered structures were submitted to the Fpocket tool. The Fpocket tool was unable to

TABLE 1 | Total druggable binding sites.

Species	Predicted structures	Protein structures with binding sites	Protein structures with druggable binding sites
<i>L. braziliensis</i>	681	669 (98.2%)	599 (87.9%)
<i>L. infantum</i>	708	701 (99%)	629 (88.8%)

recover binding sites for 12 and 7 proteins of *L. braziliensis* and *L. infantum*, respectively. Afterward, the atomic coordinates related to the predicted binding sites of *L. braziliensis* and *L. infantum* proteins were compared to the binding sites of BindingDB target proteins. That comparison was performed using the PocketMatch tool.

Therefore, matrices that depend on three-dimensional structure (S^b and S^d) were composed of 2,057 BindingDB protein targets and 669 and 701 proteins of *L. braziliensis* and *L. infantum*, respectively. Based on the analysis carried out with the Fpocket, it was possible to identify that almost 99% of the protein structures of the two *Leishmania* species had binding sites that could be identified by Fpocket, with a druggability score between 0.021 and 0.926 for *L. braziliensis* and between 0.042 and 0.985 for *L. infantum* (see **Table 1** and **Supplementary Material 2**). Considering the threshold indicated in the tool documentation, which reports a druggability score >0.5 for druggable proteins, we identified that almost 88% of the protein structures of both *Leishmania* species had druggable binding sites.

When we evaluated the druggability similarity matrices and binding sites between BindingDB drug targets and *L. braziliensis* and *L. infantum* proteins, only 277 and 265 proteins of *L. braziliensis* and *L. infantum*, respectively, showed similar structure of their binding sites ($S_{ij}^b \geq 0.7$). When we integrated binding sites with druggability similarity, it was possible to identify the number of *Leishmania* proteins with the structure of binding sites and druggability scores similar to the same BindingDB drug target, essential information to infer the drug repurposing to a new protein. In this way, we were able to identify that 192 proteins of *L. braziliensis* and 210 proteins of *L. infantum* have binding sites similar to known drug targets ($S_{ij}^d \geq 0.8$ and $S_{ij}^b \geq 0.7$), considering both their three-dimensional conformation with their chemical composition, and their druggability score (data on the similarity of druggability and binding sites can be retrieved from the **Supplementary Material** at <https://leishtargets.github.io/>).

Gene Ontology Terms and Subcellular Location

When we analyzed the GO annotations, most of the recovered BindingDB protein targets were involved in metabolism processes, cell communication, and development. When verifying the enrichment of the terms using the David tool version 6.8, it was possible to see that the BindingDB proteins were enriched for protein serine/threonine kinase activity, protein phosphorylation, protein kinase activity, and ATP binding (Enrichment Score: 175.78). The analysis performed with the *csblgo* package allowed us to identify significant similarity (Relevance similarity ≥ 0.8) between BindingDB

protein drug targets and approximately 92% of proteins from both *Leishmania* species proteomes (7,625 and 7,525 proteins of *L. braziliensis* and *L. infantum*, respectively), considering the GO terms related to Biological Processes. When evaluating this similarity using the matrices for GO terms related to Molecular Functions, we identified that near 82% of both *Leishmania* proteomes had similarities with BindingDB proteins (6,905 and 6,815 proteins of *L. braziliensis* and *L. infantum*, respectively).

Based on the data predicted by the WolfpsortII tool, it was possible to identify the majority (32.8%) of the BindingDB recovered drug targets have predicted location for the plasma membrane, while 22.7, 16.17, and 14.19% were related to the cytosol, nuclear, and extracellular locations, respectively. There were also predictions for chloroplast, cytoskeleton, endoplasmic reticulum, Golgi apparatus, lysosome, mitochondria, peroxisome, and vacuolar membrane, with $\sim 3.4\%$ of proteins predicted for 2 different subcellular locations.

It was possible to predict subcellular location for most proteins of *L. braziliensis* and *L. infantum*. Only the proteins LbrM.03.0710 and LbrM.09.0840 of *L. braziliensis* and LinJ.15.1300 of *L. infantum* did not return any prediction on subcellular location. All these proteins are described in the Uniprot database as uncharacterized proteins. We found that approximately 34% of both proteomes had subcellular location predicted as nuclear, while 18, 16, and 14% had predicted location as cytosol, mitochondria, and plasma membrane, respectively. There were also predictions for cytoskeleton, endoplasmic reticulum, extracellular, Golgi apparatus, lysosome, and peroxisome, with $\sim 5\%$ of both proteomes predicted for 2 different subcellular locations. When evaluating the similarity matrices, we verified all proteins of both *Leishmania* proteomes had subcellular localization similar to some drug targets.

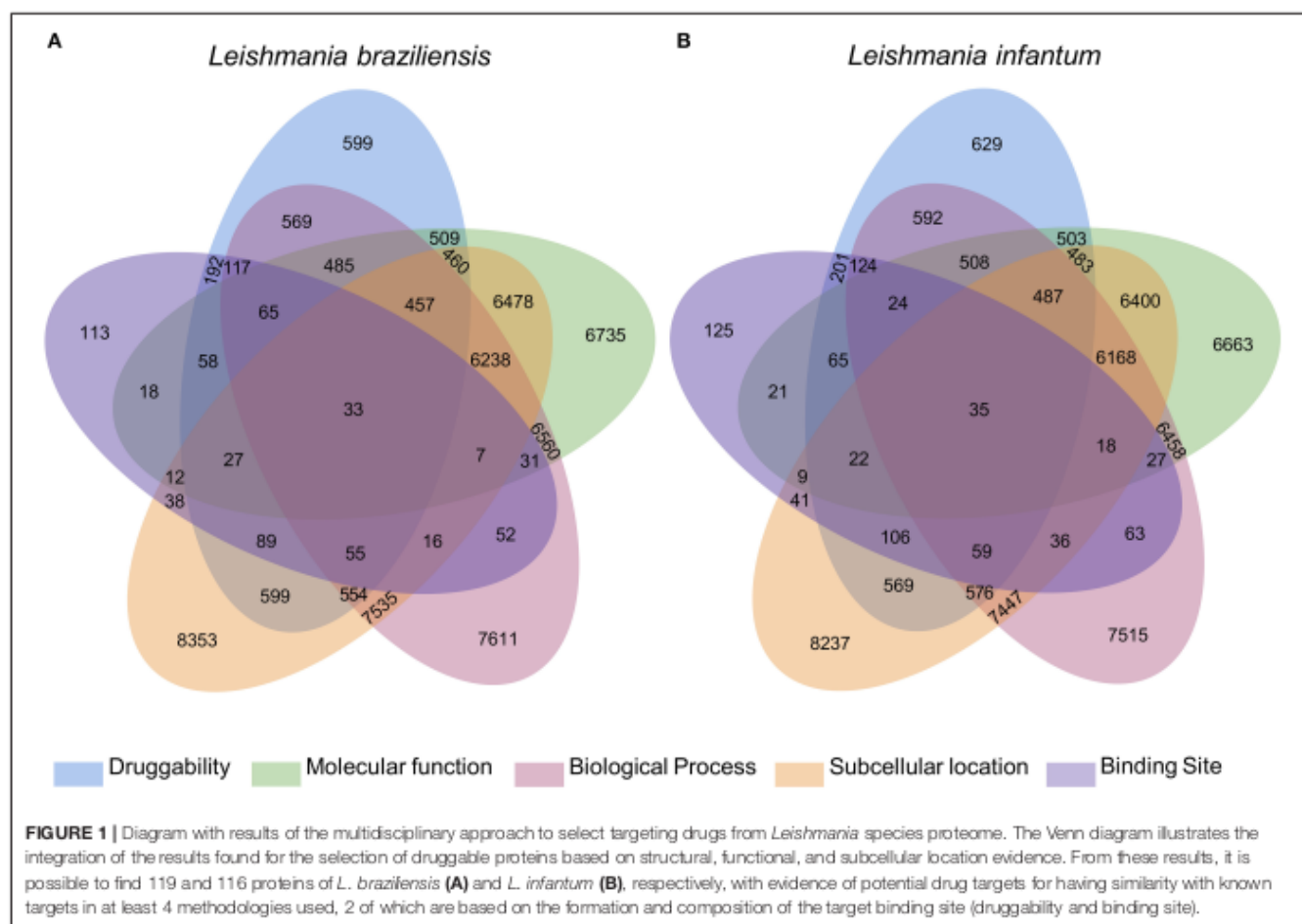
Selection of Potential *Leishmania* Drug Targets

The integration of all similarity data was used to select the most pharmacologically interesting proteins (**Figure 1**). Thus, we identified that 119 and 116 *L. braziliensis* and *L. infantum* proteins, respectively, were similar to BindingDB drug targets in at least 4 levels of similarity and that 33 and 35 of those proteins were similar to drug targets in 5 levels of similarity. We considered these last set of proteins the most promising to interact with drugs to be redirected (**Supplementary Materials 3, 4**).

These results indicate which *L. braziliensis* and *L. infantum* proteins present evidence of potential drug targets based on structural, functional, and subcellular location data. However, it was necessary to verify which of these proteins are present in the evolutionary stage of pharmacological interest, and which of these proteins would have a significant impact on the physiological stability of the organism, if used as a drug target.

Evaluation of Promising *Leishmania* Drug Targets

To analyze the presence of *Leishmania* proteins similar to drug targets at the evolutionary stage of amastigote, we analyzed the



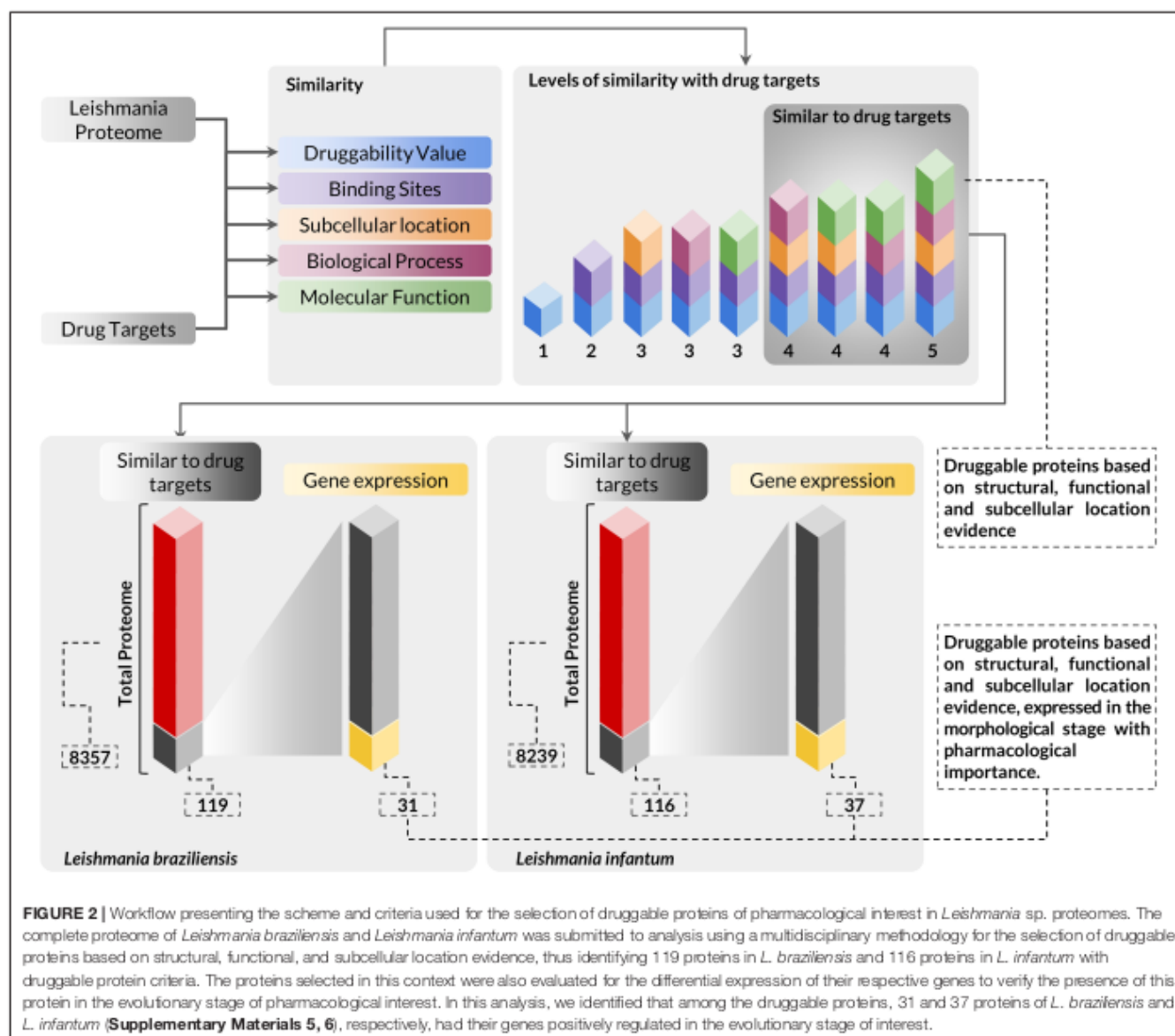
gene expression on data retrieved from SRA databases. To date, only one study deposited in the SRA (PRJNA494068) had total RNA sequencing data for the two distinct evolutionary stages (amastigote and promastigote) of wild strain of *L. braziliensis* species with no chemical interference. There was no data following such criteria for *L. infantum*. The results of expression analysis for *L. braziliensis* were used to infer gene expression for orthologs in *L. infantum*. Thus, to identify expressed genes in the evolutionary form present in the human host, the general distribution of the genes was analyzed, according to the log2 fold change between amastigote and promastigote forms. A total of 3,506 expressed genes were positively regulated with absolute values of log2 fold change > 0 and adjusted *p*-value ≤ 0.05 (Supplementary Material 1).

Integrating the results of expressed genes with the evidence for druggable targets, we identified that 31 and 37 proteins of *L. braziliensis* and *L. infantum*, respectively, had evidence of druggability and evidence of expression of its genes in amastigote, the evolutionary stage of pharmacological interest (Figure 2) (Supplementary Materials 5, 6). We evaluated the functional annotation of potential drug targets in *Leishmania* spp. proteomes using the categorization of GO terms related to biological process and molecular function

(Supplementary Material 1). From these analyses, we found that the selected targets are strongly related to metabolism, cell organization, and biogenesis, and have molecular function mainly associated with protein binding and catalytic activity.

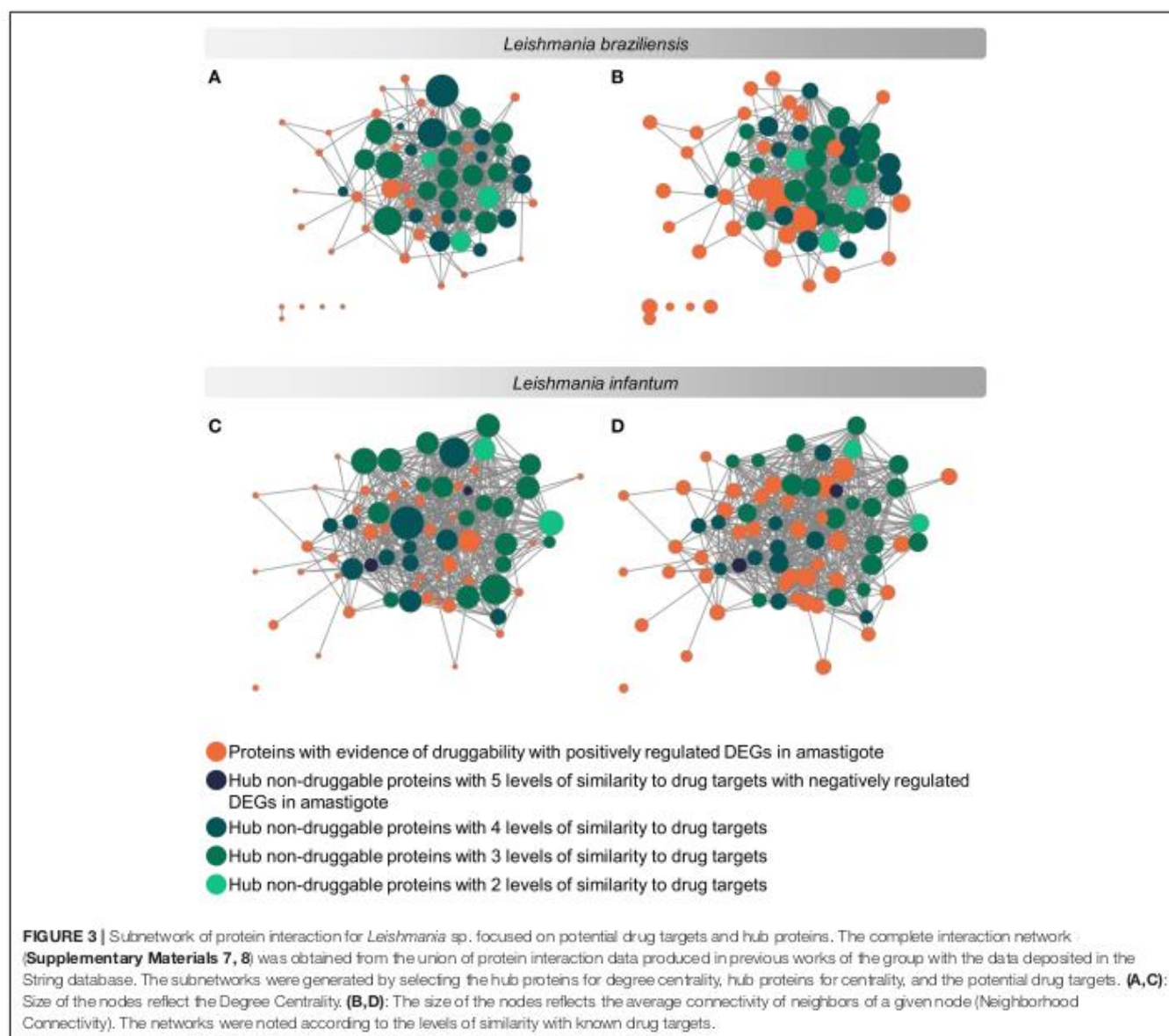
The evaluation regarding the importance of each parasite protein for the physiological stability of the organism was analyzed considering the degree of interaction of each protein in a complex of protein interaction networks (PINs). The networks used contained 3,397 proteins of *L. braziliensis* and 3,254 proteins of *L. infantum* (Supplementary Materials 7, 8 and data on protein interaction network can be retrieved from the Supplementary Material at <https://leishtargets.github.io/>). We identified that among the 31 proteins of *L. braziliensis* with evidence of druggability, 7 proteins had <20 interactions, while 7 proteins had more than 100 interactions. Considering the 37 proteins of *L. infantum*, we identified that 5 proteins had <20 interactions, while 9 proteins had more than 100 interactions (Figure 3).

According to the PINs, the potential selected targets are not the main hubs proteins in the network, however, they tend to be directly connected or to the main hub proteins. It was still possible to notice that the network hubs mostly have a high similarity with drug targets, but they were not within the criteria



established for potential drug repurposing. Thus, the network's hubs can be described as potential targets for the development of new drugs. When evaluating networks using the average connectivity of neighbors, we observed that 2 potential selected targets for each proteome, putative DEAD box RNA helicase (LbrM.35.2040, LbrM.35.2370, and LinJ.36.2260) and putative ATP-dependent RNA helicase (LinJ.28.1420) were present in highly connected neighborhoods, having a neighborhood connectivity value superior to hub proteins (with many direct interactions), thus being important for coefficient clustering and consequently important for the stability of the protein networks (Fox et al., 2011). The evaluation of similarity with the human proteome was carried out in an integrated way with the degree centrality (Figure 4), in order to weigh the importance for stability in parasite physiology and the possibility of side effects. Based on this analysis, the potential targets with

a higher degree of centrality also had a greater similarity with the human proteome, especially when considering the coverage between the proteins. Just as proteins with lower degree centrality also have less similarity, data consistent with one of the first characteristics found in biological interaction networks, which is the conservation between highly connected protein species (Sun and Kim, 2011). However, it is possible to identify 4 and 5 proteins in *L. braziliensis* and *L. infantum*, respectively (Table 2), with degree centrality > 50 and similarity against the human proteome < 50% (Figure 4). The evaluation of these characteristics makes these proteins with the greatest potential to be tested as drug targets since they have the availability of a three-dimensional conformation that can be exploited for optimization of compounds (Figure 5), they have evidence of druggability with the similarity of the binding site and druggability score with known targets, in addition to having a similar



molecular function and biological process, and share the same subcellular location.

Thus, using the cross-reference between recovered and generated data, 31 proteins from *L. braziliensis* (Supplementary Material 5) and 37 proteins from *L. infantum* (Supplementary Material 6) were identified with evidence of being potential drug targets. Among these, 2 proteins in each proteome are important for network stability due to their highly connected neighborhoods. Seven and nine proteins, in *L. braziliensis* and *L. infantum*, respectively, are relatively essential in the network as they have a degree centrality >100. While 4 and 5 proteins have similarity value < 50% and degree value > 50, based on known targets similar to these proteins, 1,105 and 145 compounds with the possibility of repurposing to be tested for their leishmanicidal potential, of which, only 9 compounds have already been tested against *Leishmania* (Supplementary Materials 5, 6).

DISCUSSION

Studies designed to select proteins for deep validation as drug targets are often carried out without assessing the whole proteome of a pathogen, and they are often focused on classic groups of proteins, neglecting new potentially druggable targets of therapeutic importance (Patel et al., 2013). The use of only traditional methods results in a lack of information about the real druggability of complete proteomes (Brown and Superti-Furga, 2003; Prinz et al., 2011). This fact becomes even more critical when talking about neglected diseases, for which the lack of information is not only concentrated on druggability, but it extends to other areas, such as the protein three-dimensional conformations and protein-drug interaction (Charlton et al., 2018). This reality turns out to be even more worrying considering the possibility of discovering new compounds by the

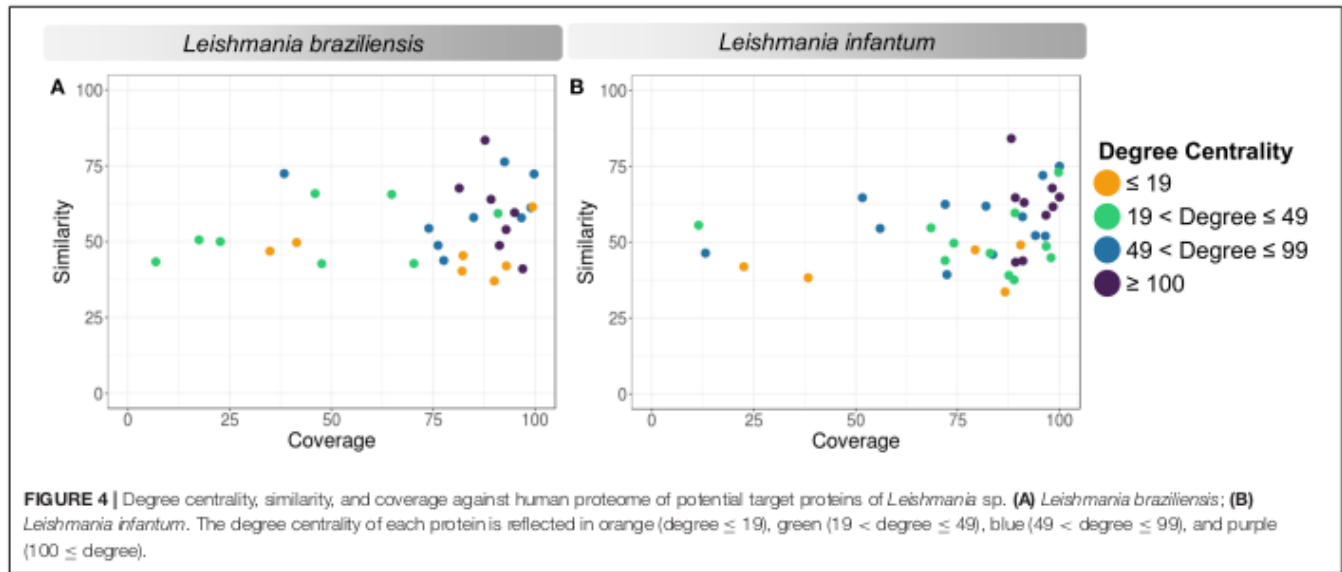


TABLE 2 | Potential drug targets with relative PIN essentiality and <50% similarity to the human proteome.

Protein	Description	Similarity*	Coverage*	Number of potential compounds to interact with <i>Leishmania</i> proteins	Number of potential compounds already tested for <i>Leishmania</i> deposited in BindingDB
LbrM.10.0640	Glycerol-3-phosphate dehydrogenase (GPHD)	48.7	91.2	63	8
LbrM.29.2360	Putative kinesin	48.7	76.1	640	0
LbrM.30.2040	Putative alcohol dehydrogenase	41.0	72.2	35	0
LbrM.35.6570	Putative small G-protein	43.7	77.5	341	1
LinJ.02.0710	Hypothetical Protein - Dipeptidylcarboxypeptidase	45.8	83.6	42	0
LinJ.10.0560	Glycerol-3-phosphate dehydrogenase (GPHD)	43.8	91.0	52	0
LinJ.27.1710	Phosphoenolpyruvate carboxykinase	46.3	13.1	22	0
LinJ.30.2100	Putative alcohol dehydrogenase	43.3	89.2	9	0
LinJ.34.3030	Putative Alpha-keto-acid decarboxylase	39.2	72.3	33	0

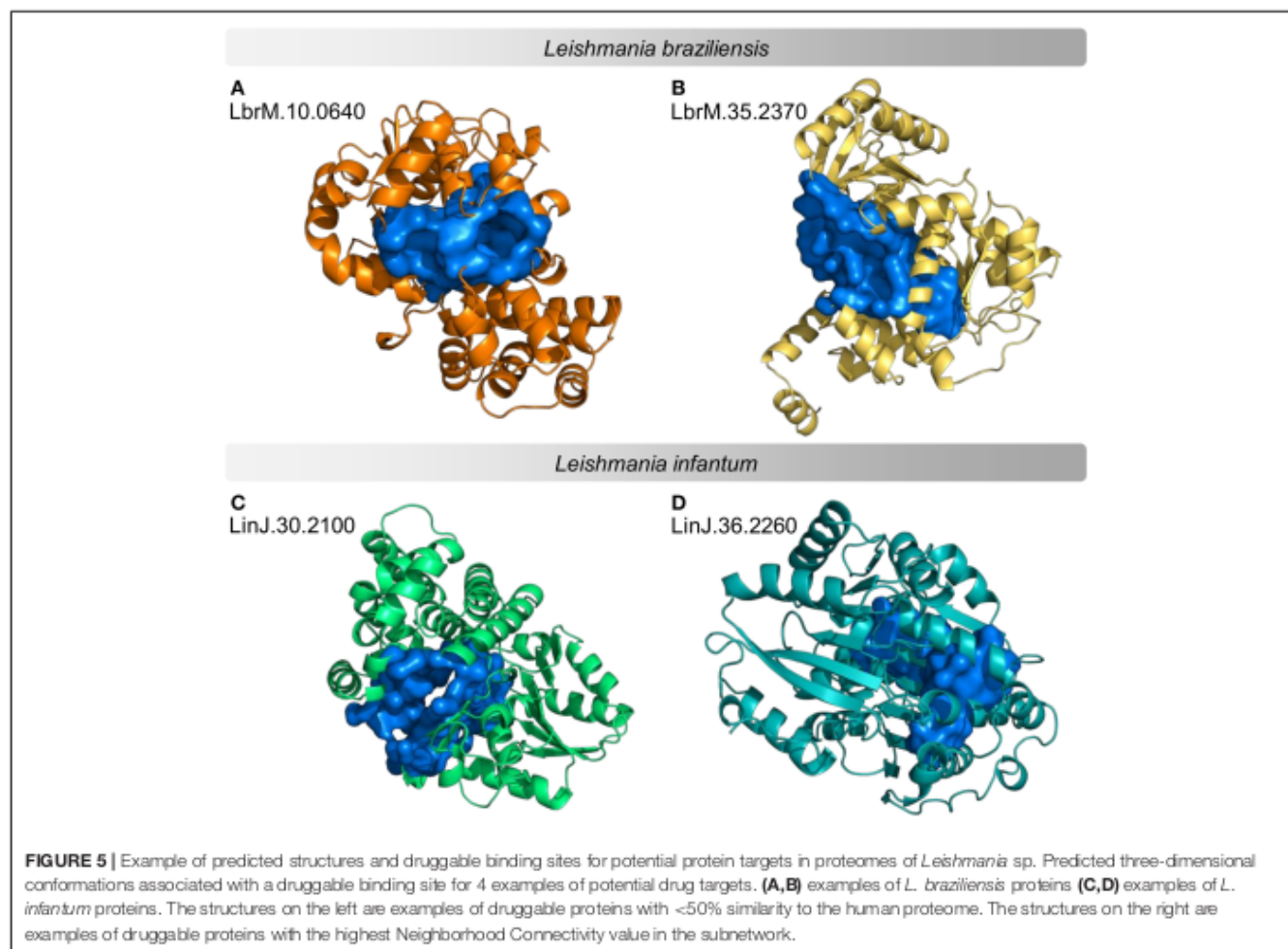
*Against the Human proteome.

pharmaceutical industry, since drugs for this group of diseases generate little financial return, thus reducing the efforts of this sector in financing research for the development of new drugs (Trouiller et al., 2002; Chatelain et al., 2013; Charlton et al., 2018). Thus, using approaches capable of analyzing complete proteomes searching for characteristics that can classify proteins as potential targets is an essential step that can help us answer questions about the druggable proteome.

In this way, we present here a modification of the methodology applied by Patel et al. (2013), capable of evaluating complete proteomes through the assessment of features about three-dimensional protein structure, druggability, functional class and subcellular location, overlapping these data with the gene expression profiles and protein topological information from interaction networks (Patel et al., 2013). In the analyses carried out by Patel et al. (2013) to identify druggable cancer

proteins, the evaluation of such characteristics cited earlier was performed to select the most promising druggable proteins, however, these characteristics were not directly compared to drug targets, there was no evaluation of similarity semantics between GO terms, and structural alignments of the binding sites were not performed, being the only direct comparison carried out through sequence alignment between the cancer proteins and drug targets (Patel et al., 2013). The absence of these analyses might not have an impact on the results of Patel et al. (2013), since the authors from that work started their study with a very defined set of oncology proteins which could be considered as primary drug targets.

In contrast, our analyses started from complete proteomes without any predefined drug candidate, so it was necessary to apply an approach comparing protein by protein between the parasite's proteome and the BindingDB database, in order to



identify potential targets. The significant difference between the approaches as well as the non-applicability of the approach carried out by Patel et al. (2013) for a complete proteome makes the results of both studies not directly comparable.

Therefore, we applied here an approach capable of evaluating the similarity of a proteome against known drug targets, considering the conformational and chemical similarity of the binding sites, the similarity of the molecular functions and biological processes those proteins are involved with, and their subcellular location. Furthermore, we integrated those data with their position in biological protein interaction networks, in order to provide a list of proteins that are important for the maintenance of the disease state and can be potential targets for drugs. This methodology also provides a list of compounds already developed with the ability to interact with the proteins selected as potential targets.

As cited in the results section, only three *L. infantum* proteins were assayed against drugs and their results are deposited in BindingDB: SIR2RP1 (NAD-dependent protein deacetylase - LinJ.26.0200), CYP51 (Putative lanosterol 14- α -demethylase - LinJ.11.1100) and TRYR [N(1),N(8)-bis (glutathionyl) spermidine reductase - LinJ.05.0350]. SIR2RP1 has already been tested as a target for the compounds Nicotinamide and analogs,

Surfactin, and Sirtinol, being described as target potential and essential for parasite survival (Vergnes et al., 2005; Zheng, 2013). This protein was not present in our results due to the absence of a predicted three-dimensional structure for it. However, it has significant similarity in terms of molecular function, biological process, and subcellular location with 234 known drug targets. CYP51 is a cytochrome 450 enzyme essential for steroidal biosynthesis that has enzyme activity inhibited by the compound VNI [(R)-N-(1-(2,4-dichlorophenyl)-2-(1H-imidazol-1-yl)ethyl)-4-(5-phenyl-1,3,4-oxadiazol-2-yl)benzamide] and analogs (Friggeri et al., 2018). TRYR is one of the enzymes involved in the process of neutralizing reactive oxygen species, being a drug target already studied in trypanosomatids due to its essentiality for parasite survival and due to its significant difference from glutathione reductase, the equivalent enzyme in humans (Schmidt and Krauth-Siegel, 2005; Rodrigues et al., 2012). In pharmacological context, LinJ.05.0350 was described as a target for 1,3,4-thiadiazolium-2-aminide, a compound with antitumor, antibacterial, antifungal, and anti-inflammatory activity already described and that presents leishmanicidal activity inhibiting the growth of *Leishmania* species *in vitro* and resulting in the decrease of lesion size and parasitic load (Da Silva et al., 2002; Rodrigues et al., 2007, 2009, 2012).

LinJ.11.1100 and LinJ.05.0350 proteins were evaluated in our study with significant similarity for the 5 metrics used against known targets of BindingDB. In our results, LinJ.11.1100 was described as similar to Steroid 21-hydroxylase (P08686), Prostaglandin G/H synthase 2 (P79208), and Thromboxane-A synthase (P49430); making it possible to infer LinJ.11.1100 as a potential target for 49 compounds, among them the inhibitor of cytochrome 450 (2S,4S)-ketoconazole and the anti-inflammatory and anticancer resveratrol, which already had antipromastigote and antiamastigote effects described in *L. major*, *Leishmania amazonensis* and *Leishmania donovani* (Kedzierski et al., 2007; Dinesh et al., 2014; Ferreira et al., 2014). LinJ.05.0350 protein, which has already been related to drug-resistant phenotype, was identified in our approach due to its similarity with Serine / threonine-protein kinase MRCK alpha (Q5VT25) and BRSK2 (Q8IWQ3), thus being inferred as a potential target for 28 compounds, including kinase inhibitors used as cancer treatment staurosporine and sunitinib, both compounds already described in the literature as potential anti-leishmanial (Dalton et al., 2010; Foucher et al., 2013; Sanderson et al., 2014; Balanco et al., 2019). Despite the high similarity with drug targets and the possibility of interaction with the compounds described above, LinJ.11.1100 and LinJ.05.0350 are not present in the final list of potential drug targets, since their genes did not show evidence of expression in the amastigote form (data on compounds with the possibility of interaction with these proteins can be retrieved from the **Supplementary Material** available at <https://leishtargets.github.io/>).

In addition to the three proteins of *L. infantum*, three other proteins of the genus *Leishmania* have the result of interaction with drugs deposited in the BindingDB, two of which belong to the proteome of *L. mexicana* and one of *L. tarentolae*. Orthologs for these proteins can be found in our results, however they are not present in the final list of potential drug targets for different reasons. The orthologous proteins of LmxM.34.4750 - Glyceraldehyde-3-phosphate dehydrogenase-like protein (LbrM.34.4710, LinJ.35.4810) were not recovered due to the absence of available three-dimensional structure. While the orthologs for LmjF.22.1360 - Farnesyl pyrophosphate synthase (LbrM.22.1240, LinJ.22.1210) and Q1A5Y2 - Tubulin alpha chain (LbrM.13.0200, LbrM.13.0190, LinJ.13.0330) were similar to BindingDB drug targets in the context chemical and biological, however they were not in accordance with criteria established in the methodology. This result allows us to assume that these proteins are potential drug targets, although it is not possible to infer interaction with the compounds covered in this study.

The multidisciplinary computational analysis allowed the identification of 68 proteins with the potential to be tested as new targets for the development or repurposing of drugs. Neither of these proteins has been described in BindingDB as a drug target. Evaluating the predicted GO terms for those 68 selected proteins, it was noticeable that we identified proteins related to functions already well-described and used by the pharmaceutical industry, such as those with kinase activity (**Supplementary Material 1**), a class of proteins that has already been used as drug targets for inflammatory, autoimmune diseases and cancer (Cohen, 2009;

Cohen and Alessi, 2013; Patterson et al., 2014; Bhullar et al., 2018). This finding is consistent with the literature since that class of proteins has also been indicated as possible drug targets for the treatment of leishmaniasis caused by *L. donovani* and *Leishmania mexicana* (Ali et al., 2012; Rachidi et al., 2014; Catta-Preta and Mottram, 2018; Raj et al., 2019). The potential of neutralizing proteins of that class has also been demonstrated by the use of the kinase inhibitor Trametinib for amastigotes and promastigotes of *L. braziliensis* and *L. infantum* (Borba et al., 2019). The same observation can be noted to targets with hydrolase activity and ATP-binding, classes of proteins that have been highlighted as potential targets (Cravatt and Lichtman, 2003; Chen et al., 2005; Soni et al., 2020), and also reported as a potential drug target in *L. infantum* (Chávez-Fumagalli et al., 2018; Sales et al., 2019).

Among the 68 potential targets it is possible to find proteins in both species that are potentially essential, but which are not currently widely explored as drug targets in *Leishmania* species, such as coproporphyrinogen III oxidase (LbrM.06.1260, LinJ.06.1330) and vacuolar protein sorting-associated protein 4 (LbrM.29.2470, LinJ.29.2610). This fact may be associated with the high similarity of these proteins with their orthologs in the human proteome. However, proteins such as small g-protein (LbrM.35.6570, LinJ.36.6510), which in our results were inferred as potential drug targets in both species, do not show high similarity with orthologs in humans, but remain unexplored as drug targets.

The applied methodology also allowed to identify proteins that until recently (in the version of the genome used in the study) were noted as hypothetical proteins, namely Major Facilitator Superfamily (LinJ.17.1550), Dephospho-CoA kinase (LinJ.18.0290), PPR repeat/Pentatricopeptide repeat domain/PPR repeat family (LinJ.36.5040), all with putative annotation; A fact that is repeated throughout the list, since among the 68 potential targets, 45 have putative annotation, thus showing that important proteins of the leishmanias druggable proteome still need further studies that reveal the full potential of these proteins to be exploited as targets drugs.

Combining the results of similarity matrices with gene expression and the protein interaction networks allowed the identification of 7 and 9 essential proteins for the amastigote evolutionary stage in *L. braziliensis* and *L. infantum*. Among these essential proteins it was possible to identify the putative DEAD box RNA helicase (LbrM.35.2040, LbrM.35.2370, and LinJ.36.2260) and putative ATP-dependent RNA helicase (LinJ.28.1420) proteins, which were reported as central by the average connectivity of neighbors, a fact that reflects their importance for parasite survival, since such proteins are essential for RNA metabolism, differentiation of amastigotes, and infectivity (Sharma and Jankowsky, 2014; Padmanabhan et al., 2016; Pandey et al., 2020). In our results, these proteins were similar to 10 drug targets of BindingDB according to the established criteria, thus we were able to infer the potential interactions, among these we had the already mentioned Staurosporine and Sunitinib, the protein kinase inhibitor Sorafenib and the antibiotic Novobiocin, both with antileishmanial activity already identified in experimental methods (Singh et al., 2005; Sanderson et al., 2014). Despite

not having a low similarity with the human proteome, being highly conserved proteins, they can be inferred as potential drug targets, since when inactivated they result in an inability to differentiate into axenic amastigotes, loss of mitochondrial membrane potential, mitochondrial fragmentation, and cell death (Padmanabhan et al., 2016; Pandey et al., 2020).

The analysis of similarity against the human proteome allowed to verify the possibility of side effects, prioritizing the targets of *Leishmania* spp. most different from the human proteome. Cross-referencing these data with data on known targets allowed us to identify a list of compounds already developed with the possibility of interacting with potential selected targets. Among more than 1,000 compounds, there are only 15 compounds already tested for *Leishmania* species with assay described in BindingDB, considering all 68 targets selected, and only 9 compounds considering targets with low similarity to the human proteome. Evaluating the list of potential targets, it was possible to identify orthologs between the species covered in this work, showing the possibility of redirecting a drug with the potential to act in different parasite species. Among these compounds inferred as antileishmanial in our results are compounds already widely known, such as the general proteinase inhibitor Pepstatin A (BindingDB ID 912) and compounds poorly associated with a potential leishmanicide, such as C20H22N8O5 (BindingDB ID 18046), Gossypol (BindingDB ID 46555) and Bosutinibe chemotherapy (BindingDB ID 4552).

In both *Leishmania* species studied, GPDH proteins (LbrM.10.0640 and LinJ.10.0560) were identified as central proteins in the PIN and potential drug targets, a fact that is compatible with literature data (Suresh et al., 2000; Ogungbe and Setzer, 2013; Passalacqua et al., 2015) since such proteins are essential for parasite survival, and present structural characteristics consistent with the requirements for intelligent drug design (Cruz et al., 2012; da Silva et al., 2012; Belluti et al., 2014; Costa et al., 2020), being drug targets already studied in other trypanosomatids such as *Trypanosoma cruzi* (Leite et al., 2009; Maluf et al., 2013) and *Trypanosoma brucei* (Haanstra et al., 2017). In addition to having a key role in the glycolytic pathway that is related to the energy production of the parasite (Naderer et al., 2010; Zhao et al., 2018), those proteins participate indirectly in the pentose phosphate pathway, therefore they are related to the maintenance of the disease state, since that pathway represents an important process to parasitic defense against reactive oxygen species produced by hosts (Maugeri et al., 2003; Costa et al., 2020).

Costa et al. (2020) also reported GPDHs as potential drug targets describing regions of the proteins, which are essential for their molecular function in parasites, however, those segments are not conserved compared to mammalian species, highlighting the importance of these regions for drug development, a result that is also consistent with the data presented by Haanstra et al. (2017), who describes GPDH inhibitors in *T. brucei* that result in the death of trypanosomes without injury to host cells (Haanstra et al., 2017). In our results, those proteins were inferred as potential targets for more than 50 drugs, including Methotrexate, Pyrimethamine, and Quinoline compounds. Methotrexate shows an antileishmanial effect with a significant decrease in the average infection rate when associated with meglumine antimoniate

(Mahmoudvand et al., 2017). However, the main use of this drug is currently related to the investigation of metabolic pathways in *Leishmania* species due to its inhibition function of folate metabolic pathways (Vickers and Beverley, 2011). The pyrimethamine has been evaluated for decades regarding its activity in leishmaniasis, and an antileishmanial effect has already been reported when in its combined forms (Neal, 1976; Serban, 2019). Quinoline compounds have already shown potent antileishmanial activity *in vitro* and *in vivo*, thus different formulations of these compounds have been tested as a treatment for leishmaniasis, due to their high efficiency when compared to the standard drug pentamidine, and due to the absence of toxicity to mammalian cells in the concentrations necessary for leishmanicidal action (Sahu et al., 2002; Tempone et al., 2005; Gopinath et al., 2014; Upadhyay et al., 2018).

Alcohol dehydrogenase proteins have been associated with miltefosine resistance in *Leishmania* parasites, as they are also involved with the reduction of reactive oxygen species and protection against oxidative stress (Camielli et al., 2014; Hefnawy et al., 2017). This class of proteins has also been described as a potential pharmacological target in analyzes performed using the *T. cruzi* proteome (Alves-Ferreira et al., 2009) and in our analyzes, these proteins (LbrM.30.2040, LinJ.30.2100) were inferred as potential targets for formamide compounds, which are known to inhibit their activity (Gibbons and Hurley, 2004; Plapp, 2010; Espuelas et al., 2012), and targets for benzimidazole-based compounds, drugs for which leishmanicidal activity has been reported and used as antiparasitic agents that act in inhibiting tubulin polymerization (Mota et al., 2014; De Luca et al., 2018; Sánchez-Salgado et al., 2018).

Phosphoenolpyruvate carboxykinase protein (PEPCK) is described as a key factor in parasite gluconeogenesis, being expressed constitutively in both main evolutionary stages, and reported as a highly immunogenic antigen capable of triggering a robust T-cell-mediated immunity in humans (Rodríguez-Contreras and Hamilton, 2014; Mou et al., 2015). In our results, PEPCK (LinJ.27.1710) was identified as a potential target for Staurosporine compounds, protein kinase inhibitors also previously inferred with potential interaction for other proteins in the dataset of potential targets in *Leishmania* species, that have an effect on leishmaniasis preventing the growth of parasitic cells and reducing their virulence (Becker and Jaffe, 1997; Basmacıyan et al., 2018; Omura et al., 2018).

Although new analyses, both computational and using traditional methods, are necessary for the identification of targets and the development of new drugs, our study shows there are many druggable and pharmacological proteins capable of interacting with already developed compounds, and they can provide essential information for future studies to test the interaction between these targets and drugs, reducing the cost of developing new drugs and the time to discover new therapeutic approaches (DiMasi et al., 2003, 2010; Basak, 2012).

We present a modified approach to the methods developed by Patel et al. (2013) where new data has been integrated for the selection of potential targets. Thus, the methodology presented is a computational approach capable of selecting a list of proteins of pharmacological interest in a complete proteome, with evidence of interaction with drugs already available. It was also possible

to provide a list of already available compounds with the ability to interact with the identified proteins. Thus, the application of this approach can be used as a first step in the intelligent selection of targets for drug repurposing, helping to prioritize proteins and compounds to be validated experimentally. Finally, we emphasize that the approach described here does not replace traditional methods capable of identifying details of the target/drug interaction, however, it can be used in conjunction allowing a wide evaluation of the possibilities within a proteome.

DATA AVAILABILITY STATEMENT

The datasets produced in this study can be found in **Supplementary Materials** and at the dedicated website: <https://leishtargets.github.io/>.

AUTHOR CONTRIBUTIONS

AR: conceptualization, project administration, resources, and supervision. CS: data curation and validation. CS and AR: formal

analysis, methodology, writing – original draft, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Fundação Oswaldo Cruz (FIOCRUZ). The computational platform used for applying the methodology described in this study was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.607139/full#supplementary-material>

REFERENCES

- Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., et al. (2016). A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Negl. Trop. Dis.* 10, 1–40. doi: 10.1371/journal.pntd.0004349
- Ali, N. O. M., Ibrahim, M. E., Grant, K. M., and Mottram, J. C. (2012). *Leishmania mexicana*: expression, characterization and activity assessment of E. coli-expressed recombinant CRK3. *Eur. Rev. Med. Pharmacol. Sci.* 16, 1338–1345. Available online at: <https://www.europeanreview.org/article/1211> (accessed January 15, 2021).
- Alves-Ferreira, M., Guimarães, A. C. R., Capriles, P. V. S. Z., Dardenne, L. E., and Degraeve, W. M. (2009). A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Mem. Inst. Oswaldo Cruz* 104, 1100–1110. doi: 10.1590/S0074-02762009000800006
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, 1–12. doi: 10.1186/gb-2010-11-10-r106
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Aslett, M., Aurrecochea, C., Berriman, M., and Al, E. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 38, D457–D462. doi: 10.1093/nar/gkp851
- Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2005). The universal protein resource (UniProt). *Nucleic Acids Res.* 33, D154–D159. doi: 10.1093/nar/gki070
- Balanco, J. M. F., Sussmann, R. A. C., Verdager, I. B., Gabriel, H. B., Kimura, E. A., and Katzin, A. M. (2019). Tocopherol biosynthesis in *Leishmania* (L.) amazonensis promastigotes. *FEBS Open Bio* 9, 743–754. doi: 10.1002/2211-5463.12613
- Basak, S. C. (2012). Chemobioinformatics: the advancing frontier of computer-aided drug design in the post-genomic era. *Curr. Comput. Aided. Drug Des.* 8, 1–2. doi: 10.2174/157340912799218507
- Basmacıyan, L., Berry, L., Gros, J., Azas, N., and Casanova, M. (2018). Temporal analysis of the autophagic and apoptotic phenotypes in *Leishmania parasites*. *Microb. Cell* 5, 404–417. doi: 10.15698/mic2018.09.646
- Bates, P. A., and Rogers, M. E. (2004). New insights into the developmental biology and transmission mechanisms of *Leishmania*. *Curr. Mol. Med.* 4, 601–609. doi: 10.2174/1566524043360285
- Becker, S., and Jaffe, C. L. (1997). Effect of protein kinase inhibitors on the growth, morphology, and infectivity of *Leishmania promastigotes*. *Parasitol. Res.* 83, 273–280. doi: 10.1007/s004360050246
- Belluti, F., Uliassi, E., Veronesi, G., Bergamini, C., Kaiser, M., Brun, R., et al. (2014). Toward the development of dual-targeted glyceraldehyde-3-phosphate dehydrogenase/trypanothione reductase inhibitors against trypanosoma brucei and trypanosoma cruzi. *ChemMedChem* 9, 371–382. doi: 10.1002/cmdc.201300399
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Bhullar, K. S., Lagarón, N. O., McGowan, E. M., Parmar, I., Jha, A., Hubbard, B. P., et al. (2018). Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol. Cancer* 17, 1–20. doi: 10.1186/s12943-018-0804-2
- Bienert, S., Waterhouse, A., De Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., et al. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* 45, D313–D319. doi: 10.1093/nar/gkw1132
- Borba, J. V. B., Silva, A. C., Ramos, P. I. P., Grazia, N., Miguel, D. C., Muratov, E. N., et al. (2019). Unveiling the kinomes of *Leishmania infantum* and *L. braziliensis* empowers the discovery of new kinase targets and antileishmanial compounds. *Comput. Struct. Biotechnol. J.* 17, 352–361. doi: 10.1016/j.csbj.2019.02.005
- Brasil, Ministério da Saúde, Secretaria de Vigilância em Saúde, Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços (2019). *Guia de Vigilância em Saúde*. 3rd ed. Ministério da Saúde Brasília: Ministério da Saúde. Available online at: http://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_saude_3ed.pdf (accessed December 5, 2020).
- Brasil, Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância das Doenças Transmissíveis (2017). *Manual De Vigilância Da Leishmaniose Tegumentar*. 1st ed. Ministério da Saúde. Available online at: http://bvsms.saude.gov.br/bvs/publicacoes/manual_vigilancia_leishmaniose_tegumentar.pdf (accessed December 5, 2020).
- Brown, D., and Superti-Furga, G. (2003). Rediscovering the sweet spot in drug discovery. *Drug Discov. Today* 8, 1067–1077. doi: 10.1016/S1359-6446(03)02902-7
- Carnielli, J. B. T., de Andrade, H. M., Pires, S. F., Chapeaurouge, A. D., Perales, J., Monti-Rocha, R., et al. (2014). Proteomic analysis of the soluble proteomes of miltefosine-sensitive and -resistant *Leishmania infantum* chagasi isolates obtained from Brazilian patients with different treatment outcomes. *J. Proteomics* 108, 198–208. doi: 10.1016/j.jprot.2014.05.010

- Catta-Preta, C. M. C., and Mottram, J. C. (2018). Drug candidate and target for leishmaniasis. *Nature* 560, 171–172. doi: 10.1038/d41586-018-05765-y
- Charlton, R. L., Rossi-Bergmann, B., Denny, P. W., and Steel, P. G. (2018). Repurposing as a strategy for the discovery of new anti-leishmanials: the state-of-the-art. *Parasitology* 145, 219–236. doi: 10.1017/S0031182017000993
- Chatelain, E., Chemistry, M., Nwaka, S., Hudson, A., Pink, R., Hudson, A., et al. (2013). Innovative lead discovery strategies for tropical diseases. *Nat. Rev. Drug Discov.* 20, 727–740.
- Chávez-Fumagalli, M. A., Schneider, M. S., Lage, D. P., Tavares, G. D. S. V., Mendonça, D. V. C., Santos, T. T. D. O., et al. (2018). A computational approach using bioinformatics to screening drug targets for *Leishmania infantum* species. *Evid. Based Complement. Altern. Med.* 2018, 1–9. doi: 10.1155/2018/6813467
- Chen, X., Wang, S., Wu, N., and Yang, C. (2005). Leukotriene A4 hydrolase as a target for cancer prevention and therapy. *Curr. Cancer Drug Targets* 4, 267–283. doi: 10.2174/1568009043333041
- Cohen, P. (2009). Targeting protein kinases for the development of anti-inflammatory drugs. *Curr. Opin. Cell Biol.* 21, 317–324. doi: 10.1016/j.cob.2009.01.015
- Cohen, P., and Alessi, D. R. (2013). Kinase drug discovery - What's next in the field? *ACS Chem. Biol.* 8, 96–104. doi: 10.1021/cb300610s
- Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 1–12. doi: 10.1155/2008/619832
- Costa, C. H. S., da Bichara, T. W., Gomes, G. C., dos Santos, A. M., da Costa, K. S., Lima, A. H. L., et al. (2020). Unraveling the conformational dynamics of glycerol 3-phosphate dehydrogenase, a nicotinamide adenine dinucleotide-dependent enzyme of *Leishmania mexicana*. *J. Biomol. Struct. Dyn.* 39, 1–12. doi: 10.1080/07391102.2020.1778538
- Cravatt, B. F., and Lichtman, A. H. (2003). Fatty acid amide hydrolase: an emerging therapeutic target in the endocannabinoid system. *Curr. Opin. Chem. Biol.* 7, 469–475. doi: 10.1016/S1367-5931(03)00079-6
- Cruz, M. C., Souza-Melo, N., da Silva, C. V., DaRocha, W. D., Bahia, D., Araújo, P. R., et al. (2012). Trypanosoma cruzi: role of δ -amastin on extracellular amastigote cell invasion and differentiation. *PLoS ONE* 7:e51804. doi: 10.1371/journal.pone.0051804
- Da Silva, E. F., Canto-Cavaleiro, M. M., Braz, V. R., Cysne-Finkelstein, L., Leon, L. L., and Echevarria, A. (2002). Synthesis, and biological evaluation of new 1,3,4-thiadiazolium-2-phenylamine derivatives against *Leishmania amazonensis* promastigotes and amastigotes. *Eur. J. Med. Chem.* 37, 979–984. doi: 10.1016/S0223-5234(02)01401-0
- da Silva, M. F. L., Zampieri, R. A., Muxel, S. M., Beverley, S. M., and Floeter-Winter, L. M. (2012). *Leishmania amazonensis* arginase compartmentalization in the lysosome is important for parasite infectivity. *PLoS ONE* 7:e34022. doi: 10.1371/journal.pone.0034022
- Dalton, J. E., Maroof, A., Owens, B. M. J., Narang, P., Johnson, K., Brown, N., et al. (2010). Inhibition of receptor tyrosine kinases restores immunocompetence and improves immune-dependent chemotherapy against experimental leishmaniasis in mice. *J. Clin. Invest.* 120, 1204–1216. doi: 10.1172/JCI41281
- De Luca, L., Ferro, S., Buemi, M. R., Monforte, A. M., Gitto, R., Schirmeister, T., et al. (2018). Discovery of benzimidazole-based *Leishmania mexicana* cysteine protease CPB2.8 Δ CTE inhibitors as potential therapeutics for leishmaniasis. *Chem. Biol. Drug Des.* 92, 1585–1596. doi: 10.1111/cbdd.13326
- Desaphy, J., Azdimousa, K., Kellenberger, E., and Rognan, D. (2012). Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J. Chem. Inf. Model.* 52, 2287–2299. doi: 10.1021/ci300184x
- DiMasi, J. A., Feldman, L., Seckler, A., and Wilson, A. (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clin. Pharmacol. Ther.* 87, 272–277. doi: 10.1038/clpt.2009.295
- DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22, 151–185. doi: 10.1016/S0167-6296(02)00126-1
- Dinesh, N., Pallerla, D. S. R., Kaur, P. K., Kishore Babu, N., and Singh, S. (2014). Exploring *Leishmania donovani* 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) as a potential drug target by biochemical, biophysical and inhibition studies. *Microb. Pathog.* 66, 14–23. doi: 10.1016/j.micpath.2013.11.001
- Dorlo, T. P. C., Balasegaram, M., Beijnen, J. H., and de Vries, P. J. (2012). Miltefosine: a review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *J. Antimicrob. Chemother.* 67, 2576–2597. doi: 10.1093/jac/dks275
- Dorlo, T. P. C., Van Thiel, P. P. A. M., Huitema, A. D. R., Keizer, R. J., De Vries, H. J. C., Beijnen, J. H., et al. (2008). Pharmacokinetics of miltefosine in old world cutaneous leishmaniasis patients. *Antimicrob. Agents Chemother.* 52, 2855–2860. doi: 10.1128/AAC.00014-08
- dos Santos Vasconcelos, C. R., de Lima Campos, T., and Rezende, A. M. (2018). Building protein-protein interaction networks for *Leishmania* species through protein structural information. *BMC Bioinformatics* 19, 1–13. doi: 10.1186/s12859-018-2105-6
- Espuelas, S., Plano, D., Nguema, P., Font, M., Palop, J. A., Irache, J. M., et al. (2012). Innovative lead compounds and formulation strategies as newer kinetoplastid therapies. *Curr. Med. Chem.* 19, 4259–4288. doi: 10.2174/092986712802884222
- Ferreira, C., Soares, D. C., Cunha Do Nascimento, M. T., Pinto-da-Silva, L. H., Sarzedas, C. G., Tinoco, L. W., et al. (2014). Resveratrol is active against *Leishmania amazonensis*: in vitro effect of its association with amphotericin B. *Antimicrob. Agents Chemother.* 58, 6197–6208. doi: 10.1128/AAC.00093-14
- Foucher, A. L., Rachidi, N., Gharbi, S., Blisnick, T., Bastin, P., Pemberton, I. K., et al. (2013). Apoptotic marker expression in the absence of cell death in staurosporine-treated *Leishmania donovani*. *Antimicrob. Agents Chemother.* 57, 1252–1261. doi: 10.1128/AAC.01983-12
- Fox, A. D., Hescott, B. J., Blumer, A. C., and Slonim, D. K. (2011). Connectedness of PPI network neighborhoods identifies regulatory hub proteins. *Bioinformatics* 27, 1135–1142. doi: 10.1093/bioinformatics/btr099
- Friggeri, L., Hargrove, T. Y., Rachakonda, G., Blobaum, A. L., Fisher, P., De Oliveira, G. M., et al. (2018). Sterol 14 α -demethylase structure-based optimization of drug candidates for human infections with the protozoan Trypanosomatidae. *J. Med. Chem.* 61, 10910–10921. doi: 10.1021/acs.jmedchem.8b01671
- Ghorbani, M., and Farhoudi, R. (2018). Leishmaniasis in humans: drug or vaccine therapy? *Drug Des. Devel. Ther.* 12, 25–40. doi: 10.2147/DDDT.S146521
- Gibbons, B. J., and Hurley, T. D. (2004). Structure of three class I human alcohol dehydrogenases complexed with isoenzyme specific formamide inhibitors. *Biochemistry* 43, 12555–12562. doi: 10.1021/bi0489107
- Gilbert, I. H. (2013). Drug discovery for neglected diseases: molecular target-based and phenotypic approaches. *J. Med. Chem.* 56, 7719–7726. doi: 10.1021/jm400362b
- Gopinath, V. S., Rao, M., Shivhare, R., Vishwakarma, P., Ghose, S., Pradhan, A., et al. (2014). Design, synthesis, ADME characterization and antileishmanial evaluation of novel substituted quinoline analogs. *Bioorganic Med. Chem. Lett.* 24, 2046–2052. doi: 10.1016/j.bmcl.2014.03.065
- Gradoni, L. (2018). "A brief introduction to leishmaniasis epidemiology," in *The Leishmaniasis: Old Neglected Tropical Diseases*, eds F. Bruschi and L. Gradoni (Berlin, Germany: Springer), 1–13.
- Haanstra, J. R., Gerding, A., Dolga, A. M., Sorgdrager, F. J. H., Buist-Homan, M., Du Toit, F., et al. (2017). Targeting pathogen metabolism without collateral damage to the host. *Sci. Rep.* 7, 1–15. doi: 10.1038/srep40406
- Hefnawy, A., Berg, M., Dujardin, J. C., and De Muylder, G. (2017). Exploiting knowledge on *Leishmania* drug resistance to support the quest for new drugs. *Trends Parasitol.* 33, 162–174. doi: 10.1016/j.pt.2016.11.003
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Hussein, H. A., Geneix, C., Petitjean, M., Borrel, A., Flatters, D., and Camproux, A.-C. (2016). Global vision of druggability issues: applications and perspectives. *Drug Discov. Today* 22, 404–415. doi: 10.1016/j.drudis.2016.11.021
- Kedzierski, L., Curtis, J. M., Kaminska, M., Jodynis-Liebert, J., and Murias, M. (2007). In vitro antileishmanial activity of resveratrol and its hydroxylated analogues against *Leishmania major* promastigotes and amastigotes. *Parasitol. Res.* 102, 91–97. doi: 10.1007/s00436-007-0729-y

- Kodama, Y., Shumway, M., and Leinonen, R. (2012). The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40, D54–D56. doi: 10.1093/nar/gkr854
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 10, 1–11. doi: 10.1186/1471-2105-10-168
- Leite, A. C., Ambrozio, A. R. P., Castilho, M. S., Vieira, P. C., Fernandes, J. B., Oliva, G., et al. (2009). Screening of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase enzyme inhibitors. *Brazilian J. Pharmacogn.* 19, 1–6. doi: 10.1590/S0102-695X2009000100002
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, Y., and Kang, C. (2017). Solution NMR spectroscopy in target-based drug discovery. *Molecules* 22:1399. doi: 10.3390/molecules22091399
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, D198–D201. doi: 10.1093/nar/gkl999
- Mahmoudvand, H., Kheirandish, F., Mirbadie, S. R., Kayedi, M. H., Rezaei Riabi, T., Ghasemi, A. A., et al. (2017). The potential use of methotrexate in the treatment of cutaneous leishmaniasis: *in vitro* assays against sensitive and meglumine antimoniate-resistant strains of *Leishmania tropica*. *Iran. J. Parasitol.* 12, 339–347.
- Maluf, F. V., Andricopulo, A. D., Oliva, G., and Guido, R. V. (2013). A pharmacophore-based virtual screening approach for the discovery of *Trypanosoma cruzi* GAPDH inhibitors. *Future Med. Chem.* 5, 2019–2035. doi: 10.4155/fmc.13.166
- Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N., and Gradoni, L. (2013). Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. *Med. Vet. Entomol.* 27, 123–147. doi: 10.1111/j.1365-2915.2012.01034.x
- Maugeri, D. A., Cazzulo, J. J., Burchmore, R. J. S., Barrett, M. P., and Ogbunode, P. O. J. (2003). Pentose phosphate metabolism in *Leishmania mexicana*. *Mol. Biochem. Parasitol.* 130, 117–125. doi: 10.1016/S0166-6851(03)00173-7
- Ministério Da Saúde (2006). *Manual de Vigilância e Controle da Leishmaniose Visceral*. Brasília: Ministério Da Saúde.
- Mota, V. Z., De Carvalho, G. S. G., Da Silva, A. D., Costa, L. A. S., De Almeida Machado, P., Coimbra, E. S., et al. (2014). Gold complexes with benzimidazole derivatives: Synthesis, characterization and biological studies. *BioMetals* 27, 183–194. doi: 10.1007/s10534-014-9703-1
- Mou, Z., Li, J., Boussoffara, T., Kishi, H., Hamana, H., Ezzati, P., et al. (2015). Identification of broadly conserved cross-species protective *Leishmania* antigen and its responding CD4⁺ T cells. *Sci. Transl. Med.* 7, 310ra167. doi: 10.1126/scitranslmed.aac5477
- Naderer, T., Heng, J., and McConville, M. J. (2010). Evidence that intracellular stages of *Leishmania major* utilize amino sugars as a major carbon source. *PLoS Pathog.* 6:e1001245. doi: 10.1371/journal.ppat.1001245
- Nagle, A. S., Khare, S., Kumar, A. B., Supek, F., Buchynsky, A., Mathison, C. J. N., et al. (2014). Recent developments in drug discovery for leishmaniasis and human african trypanosomiasis. *Chem. Rev.* 114, 11305–11347. doi: 10.1021/cr500365f
- Neal, R. A. (1976). Effect of sodium stibogluconate and pyrimethamine on mouse infections with *Leishmania mexicana*. *Ann. Trop. Med. Parasitol.* 70:252. doi: 10.1080/00034983.1976.11687120
- Ogunbe, I. V., and Setzer, W. N. (2013). In-silico *Leishmania* target selectivity of antiparasitic terpenoids. *Molecules* 18, 7761–7847. doi: 10.3390/molecules18077761
- Omura, S., Asami, Y., and Crump, A. (2018). Staurosporine: new lease of life for parent compound of today's novel and highly successful anti-cancer drugs. *J. Antibiot.* 71, 688–701. doi: 10.1038/s41429-018-0029-z
- Ovaska, K., Laakso, M., and Hautaniemi, S. (2008). Fast gene ontology based clustering for microarray experiments. *BioData Min.* 1, 1–8. doi: 10.1186/1756-0381-1-11
- Padmanabhan, P. K., Zghidi-Abouzid, O., Samant, M., Dumas, C., Aguiar, B. G., Estaquier, J., et al. (2016). DDX3 DEAD-box RNA helicase plays a central role in mitochondrial protein quality control in *Leishmania*. *Cell Death Dis.* 7, e2406–e2406. doi: 10.1038/cddis.2016.315
- Pan American Health Organization (2018). Epidemiological report of the Americas. *Rep. Leishmaniasis* 6:7. Available online at: <https://iris.paho.org/handle/10665.2/34856> (accessed December 5, 2021)
- Pandey, S. C., Pande, V., and Samant, M. (2020). DDX3 DEAD-box RNA helicase (Hel67) gene disruption impairs infectivity of *Leishmania donovani* and induces protective immunity against visceral leishmaniasis. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-75420-y
- Passalacqua, T. G., Torres, F. A. E., Nogueira, C. T., De Almeida, L., Del Cistia, M. L., Dos Santos, M. B., et al. (2015). The 2',4'-dihydroxychalcone could be explored to develop new inhibitors against the glycerol-3-phosphate dehydrogenase from *Leishmania* species. *Bioorg. Med. Chem. Lett.* 25, 3564–3568. doi: 10.1016/j.bmcl.2015.06.085
- Patel, M. N., Halling-Brown, M. D., Tym, J. E., Workman, P., and Al-Lazikani, B. (2013). Objective assessment of cancer genes for drug discovery. *Nat. Rev. Drug Discov.* 12, 35–50. doi: 10.1038/nrd3913
- Patterson, H., Nibbs, R., McInnes, I., and Siebert, S. (2014). Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. *Clin. Exp. Immunol.* 176, 1–10. doi: 10.1111/cei.12248
- Plapp, B. V. (2010). Conformational changes and catalysis by alcohol dehydrogenase. *Arch. Biochem. Biophys.* 493, 3–12. doi: 10.1016/j.abb.2009.07.001
- Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10, 712–712. doi: 10.1038/nrd3439-c1
- Rachidi, N., Taly, J. F., Durieu, E., Ledercq, O., Aulner, N., Prina, E., et al. (2014). Pharmacological assessment defines *Leishmania donovani* casein kinase 1 as a drug target and reveals important functions in parasite viability and intracellular infection. *Antimicrob. Agents Chemother.* 58, 1501–1515. doi: 10.1128/AAC.02022-13
- Rahman, R., Goyal, V., Haque, R., Jamil, K., Faiz, A., Samad, R., et al. (2017). Safety and efficacy of short course combination regimens with AmBisome, miltefosine and paromomycin for the treatment of visceral leishmaniasis (VL) in Bangladesh. *PLoS Negl. Trop. Dis.* 11:e0005635. doi: 10.1371/journal.pntd.0005635
- Raj, S., Saha, G., Sasidharan, S., Dubey, V. K., and Saudagar, P. (2019). Biochemical characterization and chemical validation of *Leishmania* MAP Kinase-3 as a potential drug target. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-52774-6
- Rezende, A. M., Folador, E. L., Resende, D., de, M., and Ruiz, J. C. (2012). Computational prediction of protein-protein interactions in *Leishmania* predicted proteomes. *PLoS ONE* 7:e51304. doi: 10.1371/journal.pone.0051304
- Rodrigues, R. F., Castro-Pinto, D., Echevarria, A., Dos Reis, C. M., Del Cistia, C. N., Sant'Anna, C. M. R., et al. (2012). Investigation of trypanothione reductase inhibitory activity by 1,3,4-thiadiazolium-2-aminide derivatives and molecular docking studies. *Bioorganic Med. Chem.* 20, 1760–1766. doi: 10.1016/j.bmc.2012.01.009
- Rodrigues, R. F., Charret, K. S., Da Silva, E. F., Echevarria, Á., Amaral, V. F., Leon, L. L., et al. (2009). Antileishmanial activity of 1,3,4-thiadiazolium-2-aminide in mice infected with *Leishmania amazonensis*. *Antimicrob. Agents Chemother.* 53, 839–842. doi: 10.1128/AAC.00062-08
- Rodrigues, R. F., da Silva, E. F., Áurea Echevarria, Fajardo-Bonin, R., Amaral, V. F., Leon, L. L., et al. (2007). A comparative study of mesoionic compounds in *Leishmania* sp. and toxicity evaluation. *Eur. J. Med. Chem.* 42, 1039–1043. doi: 10.1016/j.ejmech.2006.12.026
- Rodriguez-Contreras, D., and Hamilton, N. (2014). Gluconeogenesis in *Leishmania mexicana*: contribution of glycerol kinase, phosphoenolpyruvate carboxykinase, and pyruvate phosphate dikinase. *J. Biol. Chem.* 289, 32989–33000. doi: 10.1074/jbc.M114.569434
- Sahu, N. P., Pal, C., Mandal, N. B., Banerjee, S., Raha, M., Kundu, A. P., et al. (2002). Synthesis of a novel quinoline derivative, 2-(2-methylquinolin-4-ylamino)-N-phenylacetamide - A potential antileishmanial agent. *Bioorg. Med. Chem.* 10, 1687–1693. doi: 10.1016/S0968-0896(02)00046-9
- Sales, E. M., Sousa, G. S., Belouezane, C., Almeida, F. C. L., and Figueroa-Villar, J. D. (2019). Expression, purification and spectrophotometric analysis of nucleoside hydrolase from *Leishmania chagasi* (LcNH). *Protein Expr. Purif.* 161, 40–48. doi: 10.1016/j.pep.2019.04.009

- Sánchez-Salgado, J. C., Bilbao-Ramos, P., Dea-Ayuela, M. A., Hernández-Luis, F., Bolás-Fernández, F., Medina-Franco, J. L., et al. (2018). Systematic search for benzimidazole compounds and derivatives with antileishmanial effects. *Mol. Divers.* 22, 779–790. doi: 10.1007/s11030-018-9830-7
- Sanderson, L., Yardley, V., and Croft, S. L. (2014). Activity of anti-cancer protein kinase inhibitors against *Leishmania* spp. *J. Antimicrob. Chemother.* 69, 1888–1891. doi: 10.1093/jac/dku069
- Sanseau, P., Agarwal, P., Barnes, M. R., Pastinen, T., Richards, J. B., Cardon, L. R., et al. (2012). Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* 30, 317–320. doi: 10.1038/nbt.2151
- Schenone, M., Dančik, V., Wagner, B. K., and Clemons, P. A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9, 232–240. doi: 10.1038/nchembio.1199
- Schlicker, A., Domingues, F. S., Rahnenführer, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics* 7, 1–16. doi: 10.1186/1471-2105-7-302
- Schmidt, A., and Krauth-Siegel, R. (2005). Enzymes of the trypanothione metabolism as targets for antitrypanosomal drug development. *Curr. Top. Med. Chem.* 2, 1239–1259. doi: 10.2174/156802602393048
- Schmidtke, P., and Barril, X. (2010). Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* 53, 5858–5867. doi: 10.1021/jm100574m
- Schmidtke, P., Souaille, C., Estienne, F., Baurin, N., and Kroemer, R. T. (2010). Large-scale comparison of four binding site detection algorithms. *J. Chem. Inf. Model.* 50, 2191–2200. doi: 10.1021/ci1000289
- Serban, G. (2019). Future prospects in the treatment of parasitic diseases: 2-amino-1,3,4-thiadiazoles in leishmaniasis. *Molecules* 24:1557. doi: 10.3390/molecules24081557
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharma, D., and Jankowsky, E. (2014). The Ded1/DDX3 subfamily of DEAD-box RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* 49, 343–360. doi: 10.3109/10409238.2014.931339
- Singh, G., Jayanarayan, K. G., and Dey, C. S. (2005). Novobiocin induces apoptosis-like cell death in topoisomerase II over-expressing arsenite resistant *Leishmania donovani*. *Mol. Biochem. Parasitol.* 141, 57–69. doi: 10.1016/j.molbiopara.2005.01.014
- Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336
- Soni, D. K., Dubey, S. K., and Bhatnagar, R. (2020). ATP-binding cassette (ABC) import systems of *Mycobacterium tuberculosis*: target for drug and vaccine development. *Emerg. Microbes Infect.* 9, 207–220. doi: 10.1080/22221751.2020.1714488
- Sun, M. G. F., and Kim, P. M. (2011). Evolution of biological interaction networks: from models to real data. *Genome Biol.* 12, 1–11. doi: 10.1186/gb-2011-12-12-235
- Sunyoto, T., Potet, J., and Boelaert, M. (2018). Why miltefosine—a life-saving drug for leishmaniasis—is unavailable to people who need it the most. *BMJ Glob. Heal.* 3, 1–10. doi: 10.1136/bmjgh-2018-000709
- Suresh, S., Turley, S., Oppendoerfer, F. R., Michels, P. A. M., and Hol, W. G. J. (2000). A potential target enzyme for trypanocidal drugs revealed by the crystal structure of NAD-dependent glycerol-3-phosphate dehydrogenase from *Leishmania mexicana*. *Structure* 8, 541–552. doi: 10.1016/S0969-2126(00)00135-0
- Swinney, D. C. (2013). Phenotypic vs. Target-based drug discovery for first-in-class medicines. *Clin. Pharmacol. Ther.* 93, 299–301. doi: 10.1038/clpt.2012.236
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Tempone, A. G., Melo Pompeu Da Silva, A. C., Brandt, C. A., Scalzaretto Martinez, F., Borborema, S. E. T., Barata Da Silveira, M. A., et al. (2005). Synthesis and antileishmanial activities of novel 3-substituted quinolines. *Antimicrob. Agents Chemother.* 49, 1076–1080. doi: 10.1128/AAC.49.3.1076-1080.2005
- Trouiller, P., Oliaro, P., Torreale, E., Orbinski, J., Laing, R., and Ford, N. (2002). Drug development for neglected diseases: a deficient market and a public-health policy failure. *Lancet* 359, 2188–2194. doi: 10.1016/S0140-6736(02)09096-7
- Upadhyay, A., Kushwaha, P., Gupta, S., Dodda, R. P., Ramalingam, K., Kant, R., et al. (2018). Synthesis and evaluation of novel triazolyl quinoline derivatives as potential antileishmanial agents. *Eur. J. Med. Chem.* 154, 172–181. doi: 10.1016/j.ejmech.2018.05.014
- Vergnes, B., Sereno, D., Tavares, J., Cordeiro-Da-Silva, A., Vanhille, L., Madjidian-Sereno, N., et al. (2005). Targeted disruption of cytosolic SIR2 deacetylase discloses its essential role in *Leishmania* survival and proliferation. *Gene* 363, 85–96. doi: 10.1016/j.gene.2005.06.047
- Vickers, T. J., and Beverley, S. M. (2011). Folate metabolic pathways in *Leishmania*. *Essays Biochem.* 51, 63–80. doi: 10.1042/bse0510063
- Volkamer, A., Griewel, A., Grömbacher, T., and Rarey, M. (2010). Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J. Chem. Inf. Model.* 50, 2041–2052. doi: 10.1021/ci100241y
- World Health Organization (2010). *Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis*. Geneva: World Health Organization.
- World Health Organization/Country Office for India (2017). *Accelerated plan for kala-azar elimination*. National Vector Borne Disease Control Programme. Available online at: https://www.who.int/leishmaniasis/resources/Accelerated-Plan-Kala-azar1-Feb2017_light.pdf?ua=1 (accessed August 1, 2020).
- Yeturu, K., and Chandra, N. (2008). PocketMatch : a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics* 17, 1–17. doi: 10.1186/1471-2105-9-543
- Zhao, Y., Li, X., Wang, F., Zhao, X., Gao, Y., Zhao, C., et al. (2018). Glycerol-3-phosphate dehydrogenase (GPDH) gene family in *Zea mays* L.: identification, subcellular localization, and transcriptional responses to abiotic stresses. *PLoS ONE* 13:e0200357. doi: 10.1371/journal.pone.0200357
- Zheng, W. (2013). Sirtuins as emerging anti-parasitic targets. *Eur. J. Med. Chem.* 59, 132–140. doi: 10.1016/j.ejmech.2012.11.014
- Zhou, H., Gao, M., and Skolnick, J. (2015). Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.* 5, 1–13. doi: 10.1038/srep11090
- Zotenko, E., Mestre, J., O'Leary, D. P., and Przytycka, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* 4:e1000140. doi: 10.1371/journal.pcbi.1000140

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 dos Santos Vasconcelos and Rezende. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

6 DISCUSSÃO GERAL

A previsão de interações entre moléculas e medicamentos é de extrema importância no desenvolvimento racional de uma terapêutica, uma vez que tal informação indica novos candidatos a medicamentos e infere redirecionamento de fármacos já disponíveis. Apesar dos avanços tecnológicos, disponibilizar a informação acerca da interação molécula-droga em larga escala ainda é inviável em termos de custo e laboriosidade. Concomitante a essa problemática, o desenvolvimento computacional e o acesso aberto a dados públicos têm tornado disponível uma crescente quantidade de dados relacionados alvos biológicos, criando assim oportunidades para utilização de métodos *in silico* com o objetivo de predição de interação entre alvos moleculares e drogas. Dessa forma, as alternativas computacionais se apresentam como metodologias eficazes e de baixo custo, sendo assim uma necessidade para acelerar a descoberta de novas linhas terapêuticas (CHENG et al., 2017).

A utilização dos métodos computacionais apresentados ao longo do trabalho aplicado aos proteomas de espécies de *Leishmania* destacou a falta de informação disponível para doenças tropicais negligenciadas (VANDERELST; SPEYBROECK, 2010; VERREST; DORLO, 2017), uma vez que houve a necessidade de predição computacional da maioria dos dados utilizados para construção das redes de interação e matrizes de similaridade. No entanto, mesmo com a ausência de informações experimentais, foi possível aplicar a metodologia e selecionar potenciais alvos, demonstrando assim que a possibilidade de aplicação dessa abordagem pode beneficiar a busca de potenciais alvos em outras doenças.

A integração dos dados utilizados para selecionar alvos potenciais com base na similaridade entre os proteomas de interesse e alvos conhecidos considerando também dados de expressão gênica, permitiu selecionar alvos potenciais sem negligenciar as informações que podem ser extraídas de um proteoma, aproveitando a riqueza de dados que podem ser gerados a partir das técnicas de bioinformática. Isso permitiu a identificação de 7 e 9 proteínas essenciais para o estágio morfológico de amastigotas em *L. braziliensis* e *L. infantum*. Além disso, a análise da métrica de conectividade média com proteínas vizinhas permitiu a identificação de 4 proteínas com um valor de conectividade superior ao das proteínas hub. Destacando assim a essencialidade dessas proteínas para a estabilidade da rede de interação, uma vez

que fazem parte do fluxo de informações entre proteínas com alto grau de conectividade (FOX et al., 2011).

Esses dados ainda foram integrados com a análise de similaridade com o proteoma humano, permitindo assim verificar a possibilidade de efeitos colaterais, priorizando os alvos de *Leishmania spp.* que apresentam baixa similaridade com o proteoma humano. Integrar essa informação com os dados de fármacos disponíveis para alvos já conhecidos permitiu inferir uma lista de mais de 1000 compostos já desenvolvidos com a possibilidade de interação com os potenciais alvos selecionados. Considerando todos os 68 alvos selecionados, apenas 15 compostos, entre os mais 1000 identificados, já foram testados como antileishmanial. Ao considerar somente os compostos com possibilidade de redirecionamento para os alvos com baixa similaridade ao proteoma humano, apenas 9 compostos já foram testados. Dentre estes, podemos destacar os fármacos preditos para o alvo LbrM.10.0640: Metotrexato, que apresenta efeito anti-leishmania com diminuição significativa da taxa média de infecção quando associado ao antimoniato de meglumina (MAHMOUDVAND et al., 2017), e Pirimetamina, que tem sido avaliada há décadas quanto à sua atividade nas leishmanioses, sendo seu efeito anti-leishmania já identificado nas suas formas combinadas (NEAL, 1976; SERBAN, 2019); e a droga prevista para o alvo LbrM.35.6570: Podofilotoxina que tem um efeito anti-leishmania por inibição da polimerização da tubulina (ESCUDERO-MARTÍNEZ et al., 2017).

Embora novas análises, tanto computacionais quanto por métodos tradicionais, sejam necessárias para a identificação de alvos e o desenvolvimento de novos fármacos, este estudo mostra que existem muitas proteínas de importância biológica e farmacológica capazes de interagir com compostos já desenvolvidos, fornecendo assim informações para estudos futuros com testes de interação entre esses alvos e medicamentos, reduzindo o custo de desenvolvimento de novos medicamentos e o tempo necessário para identificar novas abordagens terapêuticas (BASAK, 2012; DIMASI et al., 2010; DIMASI; HANSEN; GRABOWSKI, 2003).

7 CONCLUSÃO

Apresentamos uma abordagem modificada dos métodos desenvolvidos por Patel et. al., (PATEL et al., 2013), onde novos dados foram integrados para a seleção de alvos potenciais. Assim, a metodologia apresentada é uma abordagem computacional capaz de selecionar uma lista de proteínas de interesse farmacológico em um proteoma completo, disponibilizando evidências de interação com fármacos já disponíveis. Essa metodologia foi aplicada ao proteoma de *Leishmania sp.* de interesse biológico e patogênico no Brasil, fornecendo uma lista de proteínas importantes para a manutenção do estado da doença, as quais apresentam evidências de viabilidade para serem utilizadas como alvos de medicamentos. A partir dessa abordagem também foi possível fornecer uma lista de compostos já desenvolvidos com capacidade de interagir com as proteínas identificadas. Dessa forma, a aplicação dessa abordagem pode ser utilizada como um primeiro passo na seleção racional de alvos para redirecionamento de fármacos, ajudando a priorizar proteínas e compostos a serem validados experimentalmente.

Por fim, ressaltamos que a abordagem aqui descrita não substitui os métodos tradicionais capazes de identificar detalhadamente a interação alvo-fármaco, mas pode ser utilizada em conjunto com os métodos experimentais para selecionar alvos além daqueles tradicionalmente já conhecidos nas espécies, permitindo assim uma ampla avaliação das possibilidades dentro de um proteoma.

REFERÊNCIAS

- AKHOUNDI, M. et al. A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. **PLoS Neglected Tropical Diseases**, v. 10, n. 3, p. 1–40, 2016.
- AKHOUNDI, M. et al. Leishmania infections: Molecular targets and diagnosis. **Molecular Aspects of Medicine**, v. 57, p. 1–29, 2017.
- ANDRADE, E. L. et al. Non-clinical studies required for new drug development – Part I: Early in silico and in vitro studies, new target discovery and validation, proof of principles and robustness of animal studies. **Brazilian Journal of Medical and Biological Research**, v. 49, n. 11, p. e5644, 2016.
- ANDREWS, K. T.; FISHER, G.; SKINNER-ADAMS, T. S. Drug repurposing and human parasitic protozoan diseases. **International Journal for Parasitology: Drugs and Drug Resistance**, 2014.
- ANVERSA, L. et al. Human leishmaniasis in Brazil: A general review. **Revista da Associação Médica Brasileira**, v. 64, n. 3, p. 281–289, 2018.
- APIC, G. et al. Illuminating drug discovery with biological pathways. **FEBS Letters**, v. 579, n. 8, p. 18872–18877, 2005.
- ASHBURNER, M. et al. **Gene ontology: Tool for the unification of biology** *Nature Genetics*, 2000.
- ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic acids research**, v. 38, n. Database issue, p. D457–62, 2010.
- AURRECOECHEA, C. et al. EuPathDB: The eukaryotic pathogen genomics database resource. **Nucleic Acids Research**, v. 45, n. D1, p. D581–D591, 2017.
- BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics**, v. 5, n. 2, p. 101–113, 2004.
- BASAK, S. C. Chemobioinformatics: the Advancing Frontier of Computer-Aided Drug Design in the Post-Genomic Era. **Current Computer-Aided Drug Design**, v. 8, n. 1, p. 1–2, 2012.
- BATES, P. A.; ROGERS, M. E. New Insights into the Developmental Biology and Transmission Mechanisms of Leishmania. **Current Molecular Medicine**, v. 4, n. 6, p. 601–609, 2004.
- BELEHU, A.; NAAFS, B. Diabetes mellitus associated with pentamidine mesylate. **The Lancet**, v. 319, n. 8287, p. 1463–1464, 1982.
- BRADY, G. P.; STOUTEN, P. F. W. Fast prediction and visualization of protein binding pockets with PASS. **Journal of Computer-Aided Molecular Design**, v. 14, n. 4, p. 383–401, 2000.

BRASIL et al. **Guia de Vigilância em Saúde**. 3. ed. Brasília: Ministério da Saúde, 2019.

BROWN, D.; SUPERTI-FURGA, G. Rediscovering the sweet spot in drug discovery. **Drug Discovery Today**, v. 8, n. 23, p. 1067–1077, 2003.

CARBON, S. et al. The Gene Ontology Resource: 20 years and still GOing strong. **Nucleic Acids Research**, 2019.

CARNIELLI, J. B. T. et al. Natural resistance of leishmania infantum to miltefosine contributes to the low efficacy in the treatment of visceral leishmaniasis in Brazil. **American Journal of Tropical Medicine and Hygiene**, v. 101, n. 4, p. 789–794, 2019.

CHARLTON, R. L. et al. Repurposing as a strategy for the discovery of new anti-leishmanials: The-state-of-the-art. **Parasitology**, v. 145, n. 2, p. 219–236, 2018.

CHENG, T. et al. Large-Scale Prediction of Drug-Target Interaction: a Data-Centric Review. **AAPS Journal**, 2017.

CLAUS, B. L.; UNDERWOOD, D. J. Discovery informatics: Its evolving role in drug discovery. **Drug Discovery Today**, v. 7, p. 957–966, 2002.

CONESA, A. et al. A survey of best practices for RNA-seq data analysis. **Genome Biology**, v. 17, n. 1, p. 13, 2016.

CROFT, S. L.; SNOWDON, D.; YARDLEY, V. The activities of four anticancer alkyllysophospholipids against *Leishmania donovani*, *Trypanosoma cruzi* and *Trypanosoma brucei*. **Journal of Antimicrobial Chemotherapy**, v. 38, n. 6, p. 1041–1047, 1996.

CSERMELY, P. et al. **Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review** *Pharmacology and Therapeutics*, 2013.

DESAPHY, J. et al. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. **Journal of Chemical Information and Modeling**, v. 52, n. 8, p. 2287–2299, 2012.

DIMASI, J. A. et al. Trends in Risks Associated With New Drug Development: Success Rates for Investigational Drugs. **Clinical Pharmacology & Therapeutics**, v. 87, n. 3, p. 272–277, 2010.

DIMASI, J. A.; HANSEN, R. W.; GRABOWSKI, H. G. The price of innovation: New estimates of drug development costs. **Journal of Health Economics**, v. 22, n. 2, p. 151–185, 2003.

DUDLEY, J. T.; DESHPANDE, T.; BUTTE, A. J. Exploiting drug-disease relationships for computational drug repositioning. **Briefings in Bioinformatics**, v. 12, n. 4, p. 303–311, 2011.

EDELSBRUNNER, H.; MÜCKE, E. P. Three-dimensional alpha shapes. **ACM**

Transactions on Graphics, v. 13, n. 1, p. 43–72, 1994.

ESCUDERO-MARTÍNEZ, J. M. et al. Antileishmanial activity and tubulin polymerization inhibition of podophyllotoxin derivatives on *Leishmania infantum*. **International Journal for Parasitology: Drugs and Drug Resistance**, 2017.

FOX, A. D. et al. Connectedness of PPI network neighborhoods identifies regulatory hub proteins. **Bioinformatics**, v. 27, n. 8, p. 1135–1142, 2011.

GAO, M.; SKOLNICK, J. A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. **PLoS Computational Biology**, v. 9, n. 10, p. 1–12, 2013.

GEILING, E. M. K.; CANNON, P. R. Pathologic effects of elixir of sulfanilamide (diethylene glycol) poisoning: A clinical and experimental correlation: Final report. **Journal of the American Medical Association**, v. 111, n. 10, p. 919–926, 1938.

GELB, M. H. et al. Protein farnesyl and N-myristoyl transferases: Piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. **Molecular and Biochemical Parasitology**, v. 126, n. 2, p. 155–163, 2003.

GHORBANI, M.; FARHOUDI, R. Leishmaniasis in humans : drug or vaccine therapy ? **Drug Design, Development and Therapy**, v. 12, p. 25–40, 2018.

GILBERT, I. H. Drug discovery for neglected diseases: Molecular target-based and phenotypic approaches. **Journal of Medicinal Chemistry**, v. 56, n. 20, p. 7719–7726, 2013.

GRADONI, L. A brief introduction to leishmaniasis epidemiology. In: BRUSCHI, F.; GRADONI, L. (Eds.). . **The Leishmaniasis: Old Neglected Tropical Diseases**. Berlin, Germany: Springer, 2018. p. 1–13.

GREAVES, P.; WILLIAMS, A.; EVE, M. First dose of potential new medicines to humans: How animals help. **Nature Reviews Drug Discovery**, v. 3, p. 226–236, 2004.

GUPTA, G.; OGHUMU, S.; SATOSKAR, A. R. Mechanisms of Immune Evasion in Leishmaniasis. **Advances in Applied Microbiology**, v. 82, p. 155–184, 2013.

HAILU, A. et al. Geographical variation in the response of visceral leishmaniasis to paromomycin in East Africa: A multicentre, open-label, randomized trial. **PLoS Neglected Tropical Diseases**, v. 4, n. 10, p. e709, 2010.

HALAKOU, F. et al. Enriching Traditional Protein-protein Interaction Networks with Alternative Conformations of Proteins. **Scientific Reports**, 2017.

HASTINGS, J. et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. **Nucleic Acids Research**, v. 44, p. D1214–D1219, 2016.

HAZARIKA, A. N. Treatment of kala-azar with pentamidine isothionate; a study of 55. **Indian medical gazette**, v. 84, n. 4, p. 140–145, 1949.

HEMMINGS, H. C.; EGAN, T. D. **Pharmacology and Physiology for Anesthesia:**

Foundations and Clinical Application. Second ed. Philadelphia: [s.n.].

HENDLICH, M.; RIPPMANN, F.; BARNICKEL, G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. **Journal of Molecular Graphics and Modelling**, v. 15, n. 6, p. 359–363, 1997.

HODOS, R. A. et al. In silico methods for drug repurposing and pharmacology. **Wiley Interdisciplinary Reviews: Systems Biology and Medicine**, v. 8, p. 186–211, 2016.

HOPKINS, A. L.; GROOM, C. R. The druggable genome. **Nature Reviews Drug Discovery**, v. 1, n. 9, p. 727–730, 2002.

HOPKINS, A. L.; MASON, J. S.; OVERINGTON, J. P. **Can we rationally design promiscuous drugs?** **Current Opinion in Structural Biology**, 2006.

HUSSEIN, H. A. et al. Global vision of druggability issues: applications and perspectives. **Drug Discovery Today**, v. 22, n. 2, p. 404–415, 2016.

ISIK, Z. et al. Drug target prioritization by perturbed gene expression and network information. **Scientific Reports**, v. 5, p. 17417, 2015.

JEONG, H. et al. Lethality and centrality in protein networks. **Nature**, v. 411, n. 6833, p. 41–42, 2001.

KAITIN, K. I. Deconstructing the drug development process: The new face of innovation. **Clinical Pharmacology and Therapeutics**, v. 87, n. 3, p. 356–361, 2010.

KAR, G.; GURSOY, A.; KESKIN, O. Human cancer protein-protein interaction network: A structural perspective. **PLoS Computational Biology**, v. 5, n. 12, 2009.

KAWABATA, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. **Proteins: Structure, Function and Bioinformatics**, v. 78, n. 5, p. 1195–1211, 2010.

KHANNA, I. Drug discovery in pharmaceutical industry: Productivity challenges and trends. **Drug Discovery Today**, v. 17, p. 1088–1102, 2012.

KIM, S. et al. PubChem substance and compound databases. **Nucleic Acids Research**, v. 44, p. D1202–D1213, 2016.

KLUG, D. M.; GELB, M. H.; POLLASTRI, M. P. **Repurposing strategies for tropical disease drug discovery** **Bioorganic and Medicinal Chemistry Letters**, 2016.

KNOWLES, J.; GROMO, G. Target selection in drug discovery. **Nature Reviews Drug Discovery**, v. 2, n. 1, p. 63–69, 2003.

KOCH, M. A.; WALDMANN, H. Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. **Drug Discovery Today**, v. 10, n. 7, p. 471–483, 2005.

KOUTSOUKAS, A. et al. From in silico target prediction to multi-target drug design: Current databases, methods and applications. **Journal of Proteomics**, v. 74, n. 12, p.

2554–2574, 2011.

KOZAKOV, D. et al. New Frontiers in Druggability. **Journal of Medicinal Chemistry**, v. 58, n. 23, p. 9063–9088, 2015.

KUHN, M. et al. Large-scale prediction of drug-target relationships. **FEBS letters**, v. 582, n. 8, p. 1283–1290, 2008.

LASKOWSKI, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. **Journal of Molecular Graphics**, v. 13, n. 5, p. 323–330, 1995.

LAW, V. et al. DrugBank 4.0: Shedding new light on drug metabolism. **Nucleic Acids Research**, v. 42, n. D1, 2014.

LE GUILLOUX, V.; SCHMIDTKE, P.; TUFFERY, P. Fpocket: An open source platform for ligand pocket detection. **BMC Bioinformatics**, v. 10, n. 168, p. 1–11, 2009.

LÉON, D.; MARKEL, S. **In Silico Technologies in Drug Target Identification and Validation**. New York: CRC Press, 2006.

LESLIE, C.; ESKIN, E.; NOBLE, W. S. **The spectrum kernel: a string kernel for SVM protein classification**. Biocomputing 2002. **Anais...2002**

LESLIE, C. S. et al. Mismatch string kernels for discriminative protein classification. **Bioinformatics**, v. 20, n. 4, p. 467–476, 2004.

LESTINOVA, T. et al. Insights into the sand fly saliva: Blood-feeding and immune interactions between sand flies, hosts, and Leishmania. **PLoS Neglected Tropical Diseases**, v. 11, n. 7, p. 1–26, 2017.

LIANG, J.; EDELSBRUNNER, H.; WOODWARD, C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. **Protein Science**, v. 7, n. 9, p. 1884–1897, 1998.

LIN, D. **An Information-Theoretic Definition of Similarity**. ICML. **Anais...1998**

LIU, T. et al. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. **Nucleic Acids Research**, v. 35, n. suppl_1, p. D198–D201, 2007.

LUO, F. et al. Modular organization of protein interaction networks. **Bioinformatics**, 2007.

MAHMOUDVAND, H. et al. The potential use of methotrexate in the treatment of cutaneous leishmaniasis: In vitro assays against sensitive and meglumine antimoniate-resistant strains of leishmania tropica. **Iranian Journal of Parasitology**, v. 12, n. 3, p. 339–347, 2017.

MAROLI, M. et al. Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. **Medical and Veterinary Entomology**, v. 27, n. 2, p. 123–147, 2013.

MATSUMORI, N. et al. Direct interaction between amphotericin B and ergosterol in lipid bilayers as revealed by ^2H NMR spectroscopy. **Journal of the American Chemical Society**, 2009.

MENG, X.-Y. et al. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. **Current Computer Aided-Drug Design**, 2012.

MINIE, M. et al. CANDO and the infinite drug discovery frontier. **Drug Discovery Today**, v. 19, n. 9, p. 1353–1363, 2014.

MITCHELL, J. B. O. The Relationship between the Sequence Identities of Alpha Helical Proteins in the PDB and the Molecular Similarities of Their Ligands. **Journal of Chemical Information and Computer Sciences**, v. 41, n. 6, p. 1617–1622, 2001.

MORIZOT, G. et al. Antimony to Cure Visceral Leishmaniasis Unresponsive to Liposomal Amphotericin B. **PLoS Neglected Tropical Diseases**, v. 10, n. 1, p. 1–7, 2016.

NAGLE, A. S. et al. Recent developments in drug discovery for leishmaniasis and human african trypanosomiasis. **Chemical Reviews**, v. 114, n. 22, p. 11305–11347, 2014.

NALLAN, L. et al. Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. **Journal of Medicinal Chemistry**, v. 48, n. 11, p. 3704–3713, 2005.

NAPOLITANO, F. et al. Drug repositioning: A machine-learning approach through data integration. **Journal of Cheminformatics**, v. 5, n. 6, p. 1–9, 2013.

NASCIMENTO, A. C. A.; PRUDÊNCIO, R. B. C.; COSTA, I. G. A multiple kernel learning algorithm for drug-target interaction prediction. **BMC Bioinformatics**, v. 17, n. 1, p. 1–16, 2016.

NEAL, R. A. Effect of sodium stibogluconate and pyrimethamine on mouse infections with *Leishmania mexicana*. **Annals of Tropical Medicine and Parasitology**, v. 70, n. 2, p. 252, 1976.

NICOLAOU, K. C. Advancing the drug discovery and development process. **Angewandte Chemie - International Edition**, v. 53, n. 35, p. 9128–9140, 2014.

NWAKA, S.; HUDSON, A. Innovative lead discovery strategies for tropical diseases. **Nature Reviews Drug Discovery**, v. 5, n. 11, p. 941–955, 2006a.

NWAKA, S.; HUDSON, A. Innovative lead discovery strategies for tropical diseases **Nature Reviews Drug Discovery**, 2006b.

OVASKA, K.; LAAKSO, M.; HAUTANIEMI, S. Fast Gene Ontology based clustering for microarray experiments. **BioData Mining**, v. 1, n. 11, p. 1–8, 2008.

OVERINGTON, J. P.; AL-LAZIKANI, B.; HOPKINS, A. L. How many drug targets are there? **Nature Reviews Drug Discovery**, 2006.

PALME, J.; HOCHREITER, S.; BODENHOFER, U. KeBABS: An R package for kernel-

based analysis of biological sequences. **Bioinformatics**, v. 31, n. 15, p. 2574–2576, 2015.

PAN AMERICAN HEALTH ORGANIZATION. Epidemiological Report of the Americas. **Report Leishmaniases**, v. 6, p. 7, 2018.

PATEL, M. N. et al. Objective assessment of cancer genes for drug discovery. **Nature Reviews Drug Discovery**, v. 12, n. 1, p. 35–50, 2013.

PURKAIT, B. et al. Mechanism of amphotericin B resistance in clinical isolates of *Leishmania donovani*. **Antimicrobial Agents and Chemotherapy**, v. 56, n. 2, p. 1031–1041, 2012.

RAMACHANDRAIAH, G.; CHANDRA, N. R. Sequence and structural determinants of mannose recognition. **Proteins: Structure, Function and Genetics**, v. 39, n. 4, p. 358–364, 2000.

RAVASZ, E. et al. Hierarchical organization of modularity in metabolic networks. **Science (New York, N.Y.)**, v. 297, n. 5586, p. 1551–5, 2002.

RESNIK, P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. **Journal of Artificial Intelligence Research**, p. 95–130, 1999.

REYNAUD, E. Protein misfolding and degenerative diseases. **Nature Education**, v. 3, n. 9, p. 28, 2010.

RIJAL, S. et al. Increasing failure of miltefosine in the treatment of kala-azar in nepal and the potential role of parasite drug resistance, reinfection, or noncompliance. **Clinical Infectious Diseases**, v. 56, n. 11, p. 1530–1538, 2013.

RUAL, J. F. et al. Towards a proteome-scale map of the human protein-protein interaction network. **Nature**, v. 437, n. 7062, p. 1173–1178, 2005.

SAIGO, H. et al. Protein homology detection using string alignment kernels. **Bioinformatics**, v. 20, n. 11, p. 1682–1689, 2004.

SAMS-DODD, F. Target-based drug discovery: Is something wrong? **Drug Discovery Today**, v. 10, n. 2, p. 139–147, 2005.

SANSEAU, P. et al. Use of genome-wide association studies for drug repositioning. **Nature Biotechnology**, v. 30, p. 317–320, 2012.

SCHENONE, M. et al. Target identification and mechanism of action in chemical biology and drug discovery. **Nature Chemical Biology**, v. 9, n. 4, p. 232–240, 2013.

SCHLICKER, A. et al. A new measure for functional similarity of gene products based on gene ontology. **BMC Bioinformatics**, v. 7, n. 1, p. 1–16, 2006.

SCHMIDTKE, P.; BARRIL, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. **Journal of Medicinal Chemistry**, v. 53, n. 15, p. 5858–5867, 2010.

SERBAN, G. Future prospects in the treatment of parasitic diseases: 2-amino-1,3,4-thiadiazoles in leishmaniasis. **Molecules**, v. 24, n. 8, p. 1557, 2019.

SKOLNICK, J. et al. Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function. **Bioorganic and Medicinal Chemistry Letters**, v. 25, n. 6, p. 1163–1170, 2015.

SLIWOSKI, G. et al. Computational methods in drug discovery. **Pharmacological reviews**, v. 66, n. 1, p. 334–95, 2014.

SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **Journal of Molecular Biology**, v. 147, n. 1, p. 195–197, 1981.

SOUZA DO LAGO, A. et al. The elderly respond to antimony therapy for cutaneous leishmaniasis similarly to young patients but have severe adverse reactions. **American Journal of Tropical Medicine and Hygiene**, v. 98, n. 5, p. 1317–1324, 2018.

SRIVASTAVA, P. et al. Unusual case of resistance to amphotericin B in visceral leishmaniasis in a region in India where leishmaniasis is not endemic. **Journal of Clinical Microbiology**, v. 49, n. 8, p. 3088–3091, 2011.

STEVERDING, D. The history of leishmaniasis. **Parasites and Vectors**, v. 10, n. 1, p. 1–10, 2017.

SUNDAR, S. et al. Oral Miltefosine for Indian Visceral Leishmaniasis. **New England Journal of Medicine**, v. 347, n. 22, p. 1739–1746, 2002.

SUNDAR, S. et al. Injectable paromomycin for visceral leishmaniasis in India. **New England Journal of Medicine**, v. 356, n. 25, p. 2571–2581, 2007.

SUNDAR, S. et al. Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use. **Clinical Infectious Diseases**, v. 55, n. 4, p. 543–550, 2012.

SUNDAR, S.; SINGH, A. Recent developments and future prospects in the treatment of visceral leishmaniasis. **Therapeutic Advances in Infectious Disease**, v. 3, n. 3–4, p. 98–109, 2016.

SUNYOTO, T.; POTET, J.; BOELAERT, M. Why miltefosine—a life-saving drug for leishmaniasis—is unavailable to people who need it the most. **BMJ Global Health**, v. 3, n. 3, p. 1–10, 2018.

SWINNEY, D. C. Phenotypic vs. Target-based drug discovery for first-in-class medicines. **Clinical Pharmacology and Therapeutics**, v. 93, n. 4, p. 299–301, 2013.

SZKLARCZYK, D. et al. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. **Nucleic Acids Research**, v. 44, p. D380–D384, 2016.

TAN, F. et al. Drug repositioning by applying ‘expression profiles’ generated by integrating chemical structure similarity and gene semantic similarity. **Molecular BioSystems**, v. 10, n. 5, p. 1126–1138, 2014.

TARI, L. W. The Utility of Structural Biology in Drug Discovery. In: TARI, L. W. (Ed.). . **Structure-Based Drug Discovery**. San Diego: Humana Press, 2012. p. 385.

TEIXEIRA, D. E. et al. The Cell Biology of Leishmania: How to Teach Using Animations. **PLoS Pathogens**, 2013.

THIMM, M. et al. Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. **Journal of Chemical Information and Computer Sciences**, v. 44, n. 5, p. 1816–1822, 2004.

TROUILLER, P. et al. Drug development for neglected diseases: A deficient market and a public-health policy failure. **Lancet**, v. 359, p. 2188–2194, 2002.

VANAERSCHOT, M. et al. Treatment failure in leishmaniasis: drug-resistance or another (epi-) phenotype ? **Expert Review of Anti-infective Therapy**, v. 12, n. 8, p. 937–946, 2014.

VANDERELST, D.; SPEYBROECK, N. **Quantifying the lack of scientific interest in neglected tropical diseases** **PLoS Neglected Tropical Diseases**, 2010.

VANHAELEN, Q. et al. Design of efficient computational workflows for in silico drug repurposing. **Drug Discovery Today**, v. 22, n. 2, p. 210–222, 2017.

VERLI, H. **Bioinformática da Biologia à Flexibilidade Molecular**. [s.l: s.n.]. v. 53

VERREST, L.; DORLO, T. P. C. **Lack of Clinical Pharmacokinetic Studies to Optimize the Treatment of Neglected Tropical Diseases: A Systematic Review** **Clinical Pharmacokinetics**, 2017.

VICENS, Q.; WESTHOF, E. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. **Structure**, v. 9, n. 8, p. 647–658, 2001.

VOLKAMER, A. et al. Analyzing the topology of active sites: On the prediction of pockets and subpockets. **Journal of Chemical Information and Modeling**, v. 50, n. 11, p. 2041–2052, 2010.

VOLKAMER, A. et al. **Combining global and local measures for structure-based druggability predictions**. *Journal of Chemical Information and Modeling*. **Anais...**2012

WALTERS, W. P. et al. What do medicinal chemists actually make? A 50-year retrospective. **Journal of Medicinal Chemistry**, v. 54, n. 19, p. 6405–6416, 2011.

WEISEL, M.; PROSCHAK, E.; SCHNEIDER, G. PocketPicker: Analysis of ligand binding-sites with shape descriptors. **Chemistry Central Journal**, v. 1, n. 7, p. 1–17, 2007.

WHO/DEPARTMENT OF CONTROL OF NEGLECTED TROPICAL DISEASES. **Global leishmaniasis update, 2006–2015: a turning point in leishmaniasis surveillance**. Geneva: [s.n.].

WILLETT, P.; BARNARD, J. M.; DOWNS, G. M. Chemical similarity searching.

Journal of Chemical Information and Computer Sciences, v. 38, p. 983–996, 1998.

WILLIAMS, K. et al. Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. **Journal of the Royal Society Interface**, v. 12, p. 1–9, 2015.

WORLD HEALTH ORGANIZATION/COUNTRY OFFICE FOR INDIA. **Accelerated plan for kala-azar elimination**. [s.l: s.n.].

WORLD HEALTH ORGANIZATION. **Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniases**. World Health Organization Technical Report Series. Geneva: [s.n.].

XU, X.; HUANG, M.; ZOU, X. Docking-based inverse virtual screening: methods, applications, and challenges. **Biophysics Reports**, v. 4, n. 1, p. 1–16, 2018.

YAMANISHI, Y. et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. **Bioinformatics**, v. 24, n. 13, p. 232–240, 2008.

YETURU, K.; CHANDRA, N. PocketMatch : A new algorithm to compare binding sites in protein structures. **BMC Bioinformatics**, v. 17, p. 1–17, 2008.

YU, G. et al. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. **Bioinformatics**, v. 26, p. 976–978, 2010a.

YU, H. et al. Genomic analysis of essentiality within protein networks. **Trends in Genetics**, v. 20, n. 6, p. 227–231, 2004.

YU, H. et al. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. **PLoS Computational Biology**, v. 3, n. 4, p. 713–720, 2007.

YU, J. et al. Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. **Bioinformatics**, v. 26, n. 1, p. 46–52, 2010b.

ZHANG, W. et al. Polypharmacology in Drug Discovery: A Review from Systems Pharmacology Perspective. **Current Pharmaceutical Design**, 2016.

ZHANG, Z. et al. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. **Bioinformatics (Oxford, England)**, v. 27, n. 15, p. 2083–8, 2011.

ZHAO, C.; WANG, Z. GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. **Scientific Reports**, v. 8, n. 1, p. 1–10, 2018.

ZHENG, X. S.; CHAN, T. F.; ZHOU, H. H. Genetic and genomic approaches to identify and study the targets of bioactive small molecules. **Chemistry and Biology**, v. 11, n. 5, p. 609–618, 2004.

ZHOU, W.; YAN, H. Alpha shape and delaunay triangulation in studies of protein-related interactions. **Briefings in Bioinformatics**, v. 15, n. 1, p. 54–64, 2014.

ANEXOS

ANEXO A - Primeira Folha dos Artigos Publicados Durante o Período do Doutorado
Não Diretamente Relacionados a Tese.



AMERICAN
SOCIETY FOR
MICROBIOLOGY

Antimicrobial Agents
and Chemotherapy®

MECHANISMS OF RESISTANCE



High Frequency of OXA-253-Producing *Acinetobacter baumannii* in Different Hospitals in Recife, Brazil

Felipe Lira de Sá Cavalcanti,^{a,b} Carina Lucena Mendes-Marques,^b
Chrisllane Rafael dos Santos Vasconcelos,^{b,c} Túlio de Lima Campos,^b
Antonio Mauro Rezende,^b Danilo Elias Xavier,^b Nilma Cintra Leal,^b
Oswaldo Pompilio de-Melo-Neto,^b Marcia Maria Camargo de Moraes,^a
Tereza Cristina Leal-Balbino^b

Laboratório de Resistência Microbiana, Instituto de Ciências Biológicas, Universidade de Pernambuco, Recife, Brazil^a; Centro de Pesquisas Aggeu Magalhães/Fiocruz, Recife, Pernambuco, Brazil^b; Departamento de Genética, Universidade Federal de Pernambuco, Recife, Brazil^c

ABSTRACT Here, we report the isolation of 31 *Acinetobacter baumannii* strains producing OXA-253 in a single large Brazilian city. These strains belonged to five different sequence types (STs), including 4 STs not previously associated with *bla*_{OXA-253}. In all strains, the *bla*_{OXA-253} gene was located in a plasmid within a genetic environment similar to what was found previously in Brazil and Italy. The reported data emphasize the successful transmission of the *bla*_{OXA-253} gene through a large area and the tendency for this resistance determinant to remain in the *A. baumannii* population.

KEYWORDS *Acinetobacter*, MLST, antibiotic resistance, genome analysis, oxacillinase

Acinetobacter baumannii is an opportunistic Gram-negative pathogen responsible for a large number of outbreaks of hospital-acquired infections. This bacterium has been linked with serious infections affecting mainly debilitated patients in intensive care units (1), and it can become resistant to a wide range of antibiotics, leading to serious impediments to treatment. Carbapenems, such as imipenem and meropenem, have been often used as drugs of last resort, but outbreaks of carbapenem-resistant *A. baumannii* have been reported in several countries, including Brazil (2).

Resistance to carbapenems in *A. baumannii* may be due to various mechanisms, including the loss of outer membrane protein (OMP), overexpression of efflux pumps, or through alterations in the penicillin-binding protein (3). An increase in reported carbapenem resistance by *A. baumannii* isolates, however, has been mostly attributed to yet another mechanism, the production of carbapenemases, such as those belonging to the class D OXA type (4).

Here, we report OXA-253-producing *A. baumannii* strains isolated from different hospitals localized within the city of Recife, state of Pernambuco, in northeastern Brazil. This variant of the OXA-143-like carbapenemase belongs to the Ambler class D of β -lactamases (5), which has previously been the subject of studies discussing its incidence in Brazil. So far, however, the *bla*_{OXA-143} gene has only been reported for samples from the states of São Paulo and Rio de Janeiro, over 1,800 km to the south of Recife (2, 6–8).

Identification of the OXA-253-producing strains occurred after sequencing of the genomes of 45 *A. baumannii* strains displaying an extensively drug-resistant (XDR) phenotype and isolated from different individuals hospitalized in five public hospitals in Recife between 2010 and 2014. These strains were considered resistant to the two carbapenems tested (imipenem and meropenem) according to the Brazilian Committee

Received 19 June 2016 Returned for
modification 21 July 2016 Accepted 3
November 2016

Accepted manuscript posted online 14
November 2016

Citation de Sá Cavalcanti FL, Mendes-Marques CL, Vasconcelos CRDS, de Lima Campos T, Rezende AM, Xavier DE, Leal NC, de-Melo-Neto OP, de Moraes MMC, Leal-Balbino TC. 2017. High frequency of OXA-253-producing *Acinetobacter baumannii* in different hospitals in Recife, Brazil. Antimicrob Agents Chemother 61:e01309-16. <https://doi.org/10.1128/AAC.01309-16>.

Copyright © 2016 American Society for Microbiology. All Rights Reserved.

Address correspondence to Tereza Cristina Leal-Balbino, cristina@cpqam.fiocruz.br.



N-glycosylation influences the catalytic activity of mosquito α -glucosidases associated with susceptibility or refractoriness to *Lysinibacillus sphaericus*

Nathaly Alexandre do Nascimento^a, Lígia Maria Ferreira^a, Tatiany Patrícia Romão^a, Darleide Maria da Conceição Correia^a, Crhisllane Rafael dos Santos Vasconcelos^b, Antônio Mauro Rezende^b, Samara Graciane Costa^c, Fernando Ariel Genta^{c,d}, Osvaldo Pompílio de-Melo-Neto^b, Maria Helena Neves Lobo Silva-Filha^{a,*}

^a Department of Entomology, Centro de Pesquisas Aggeu Magalhães/FIOCRUZ, Recife, PE 50740-465, Brazil

^b Department of Microbiology, Centro de Pesquisas Aggeu Magalhães/FIOCRUZ, Recife, PE 50740-465, Brazil

^c Instituto Oswaldo Cruz/FIOCRUZ, Rio de Janeiro, RJ 21045-900, Brazil

^d Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular, Rio de Janeiro, RJ 21941-902, Brazil

ARTICLE INFO

Article history:

Received 23 September 2016

Received in revised form

1 December 2016

Accepted 20 December 2016

Available online 23 December 2016

Keywords:

Cqm1

Aam1

Culex quinquefasciatus

Aedes aegypti

N-glycosylation

Glucosidases

ABSTRACT

Cqm1 and Aam1 are α -glucosidases (EC 3.2.1.20) expressed in *Culex quinquefasciatus* and *Aedes aegypti* larvae midgut, respectively. These orthologs share high sequence similarity but while Cqm1 acts as a receptor for the Binary (Bin) insecticidal toxin from *Lysinibacillus sphaericus*, Aam1 does not bind the toxin, rendering *Ae. aegypti* refractory to this bacterium. Aam1 is heavily glycosylated, contrasting to Cqm1, but little is known regarding how glycosylation impacts on its function. This study aimed to compare the N-glycosylation patterns and the catalytic activities of Aam1 and Cqm1. Mutant proteins were generated where predicted Aam1 N-glycosylation sites (N-PGS) were either inserted into Cqm1 or abrogated in Aam1. The mutants validated four N-PGS which were found to localize externally on the Aam1 structure. These Aam1 and Cqm1 mutants maintained their Bin binding properties, confirming that glycosylation has no role in this interaction. The α -glucosidase activity of both proteins was next investigated, with Aam1 having a remarkably higher catalytic efficiency, influenced by changes in glycosylation. Molecular dynamics showed that glycosylated and nonglycosylated Aam1 models displayed distinct patterns that could influence their catalytic activity. Differential N-glycosylation may then be associated with higher catalytic efficiency in Aam1, enhancing the functional diversity of related orthologs.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Culex quinquefasciatus and *Aedes aegypti* play important roles as vectors for etiological agents of expanding human diseases, being responsible for a significant burden to public health services worldwide (Ottesen, 2000; Bhatt et al., 2013; Musso, 2015). Mosquito molecules, such as digestive enzymes, are promising candidate targets for interventions aimed at mosquito control; however

they require further investigation. The α -glucosidases (EC 3.2.1.20) are exo-glycoside hydrolases that act on α -1,4-glycosidic bonds in order to release glucose from oligosaccharides (Krasikov et al., 2001), therefore playing a critical role during carbohydrate digestion. These enzymes are classified among the glycoside hydrolase families (GHFs) 4, 13, 31, 63, 97 and 122, but insect α -glucosidases belong mainly to GH13, according to the CAZy classification (www.cazy.org). GHF13 members are characterized by a $(\beta/\alpha)_8$ barrel 3-D structure as well as other features (Janecek, 1997; Krasikov et al., 2001).

Mosquito α -glucosidases have been characterized from species of *Anopheles*, *Culex* and *Aedes*, most of them derived from adult salivary glands (Marinotti and James, 1990; Zheng et al., 1995; Marinotti et al.,

* Corresponding author. Departamento de Entomologia, Centro de Pesquisas Aggeu Magalhães/FIOCRUZ, Av. Moraes Rego s/n, Cidade Universitária, Recife, PE 50740-465, Brazil.

E-mail address: mhneves@cpqam.fiocruz.br (M.H.N.L. Silva-Filha).

RESEARCH

Open Access



A differential transcriptional profile by *Culex quinquefasciatus* larvae resistant to *Lysinibacillus sphaericus* IAB59 highlights genes and pathways associated with the resistance phenotype

Tatiana Maria Teodoro Rezende, Antonio Mauro Rezende, Gabriel Luz Wallau, Crhisllane Rafael Santos Vasconcelos, Osvaldo Pompílio de-Melo-Neto, Maria Helena Neves Lobo Silva-Filha* and Tatiary Patrícia Romão*

Abstract

Background: The study of the mechanisms by which larvae of the *Culex quinquefasciatus* mosquito survive exposure to the entomopathogen *Lysinibacillus sphaericus* has benefited substantially from the generation of laboratory-selected colonies resistant to this bacterium. One such colony, RIAB59, was selected after regular long-term exposure of larvae to the *L. sphaericus* IAB59 strain. This strain is characterized by its ability to produce the well known Binary (Bin) toxin, and the recently characterized Cry48Aa/Cry49Aa toxin, able to kill Bin-resistant larvae. Resistance to Bin is associated with the depletion of its receptor, Cqm1 α -glucosidase, from the larvae midgut. This study aimed to identify novel molecules and pathways associated with survival of the RIAB59 larvae and the resistance phenotype.

Methods: A transcriptomic approach and bioinformatic tools were used to compare the profiles derived from the midguts of larvae resistant and susceptible to *L. sphaericus* IAB59.

Results: The RNA-seq profiles identified 1355 differentially expressed genes (DEGs), with 673 down- and 682 upregulated transcripts. One of the most downregulated DEGs was *cqm1*, which validates the approach. Other strongly downregulated mRNAs encode the enzyme pantetheinase, apolipoprotein D, lipases, heat-shock proteins and a number of lesser known and hypothetical polypeptides. Among the upregulated DEGs, the top most encodes a peroxisomal enzyme involved in lipid metabolism, while others encode enzymes associated with juvenile hormone synthesis, ion channels, DNA binding proteins and defense polypeptides. Further analyses confirmed a strong downregulation of several enzymes involved in lipid catabolism while the assignment of DEGs into metabolic pathways highlighted the upregulation of those related to DNA synthesis and maintenance, confirmed by their clustering into related protein networks. Several other pathways were also identified with mixed profiles of down- and upregulated transcripts. Quantitative RT-PCR confirmed the changes in levels seen for selected mRNAs.

Conclusions: Our transcriptome-wide dataset revealed that the RIAB59 colony, found to be substantially more resistant to Bin than to the Cry48Aa/Cry49Aa toxin, developed a differential expression profile as well as metabolic features co-selected during the long-term adaptation to IAB59 and that are most likely linked to Bin resistance.

Keywords: Biolarvicides, Binary toxin, Cry48Aa/Cry49Aa, Receptors, Cqm1, Transcriptome

*Correspondence: mhneves@cpqam.fiocruz.br; taty@cpqam.fiocruz.br
 Instituto Aggeu Magalhães-FIOCRUZ, Av. Moraes Rego s/n Cidade
 Universitária, Recife, PE 50740-465, Brazil



© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.



Contents lists available at ScienceDirect

Journal of Global Antimicrobial Resistance

journal homepage: www.elsevier.com/locate/jgar

Letter to the Editor

Multidrug-resistant *Klebsiella aerogenes* clinical isolates from Brazil carrying IncQ1 plasmids containing the *bla*_{KPC-2} gene associated with non-Tn4401 elements (NTE_{KPC-Ild})



Sir,

Klebsiella aerogenes are involved in healthcare-associated infections, causing various infections including sepsis and infections of the urinary tract and eye. They frequently have high levels of resistance to different classes of drugs, including carbapenems, and a high capacity for clonal dissemination. Although *K. aerogenes* isolates may have a clonal relationship, they may harbour different plasmids and have different resistance profiles, carrying several antimicrobial resistance genes such as *bla*_{KPC}, *bla*_{VIM}, *bla*_{IMP} and *bla*_{CTX-M} [1]. The *bla*_{KPC} gene was first reported in Brazil in the city of Recife-PE in *Klebsiella pneumoniae* strains [2]. However, a consensus has not yet been reached on the plasmid location of the *bla*_{KPC-2} gene, the transposon variants that can carry this gene and which incompatibility (Inc) groups are present in *K. aerogenes* carrying the *bla*_{KPC} gene. Here we report the first description of the non-Tn4401 element (NTE_{KPC-Ild}) carrying the *bla*_{KPC-2} and *aph(3')-Vli* genes in IncQ1 plasmids in *K. aerogenes* as well as the first report of the *bla*_{SCO-1} gene in *K. aerogenes*.

Two multidrug-resistant clinical *K. aerogenes* isolates (Ea18A and Ea33A) carrying the *bla*_{KPC-2} gene from blood samples of two patients admitted to the intensive care unit of a public hospital in Recife-PE, Brazil, in 2012 were selected. The isolates presented a clonal relationship by enterobacterial repetitive intergenic consensus (ERIC) PCR as determined in a previous study by our group [1]. The isolates were identified using a BD BACTEC™ 9120/Phoenix automated system. Data were interpreted according to Clinical and Laboratory Standards Institute (CLSI) 2018 guidelines.

Plasmid DNA was extracted using a Plasmid Mini Kit (QIAGEN) according to the manufacturer's instructions. Isolates were characterised using an Illumina MiSeq sequencer with Nextera XT libraries (Illumina Inc.).

Data were processed to remove low-quality reads using Trimmomatic. The filtered reads were then de novo assembled using the Velvet tool, whose parameters were optimised through the Velvet Optimiser programme. The Velvet results were also used as input for another assembly programme (CAP3) to improve the assemblies. Prediction and gene annotation were performed using Prokka. Nucleotide sequences of the isolates were deposited in GenBank under accession nos. **MG786907** and **MH000708**.

Analysis using ResFinder and GenBank databases showed 100% identity for the *bla*_{KPC-2}, *bla*_{SCO-1}, *sul2* and *aph(3')-Vli* genes in the plasmid DNA of both isolates.

The *bla*_{KPC-2} gene was found inserted between a partial ISKpn6 insertion sequence (Δ ISKpn6) with an associated left inverted repeat (IRL) and *tnpR* resolvase. A 21-bp fragment corresponding to the *bla*_{TEM} gene was found upstream of the *bla*_{KPC-2} gene (Fig. 1). Comparing the genetic environment of the *bla*_{KPC-2} gene with sequences deposited in GenBank, the variant NTE_{KPC-Ild} (GenBank accession no. **MH595533**) with ~100% identity for the isolates was observed.

In isolate Ea18A, the replicon corresponding to IncQ1 was found in the same contig as the *aph(3')-Vli* gene, whereas in isolate Ea33A the replicon corresponding to IncQ1 was found in the same contig as the *bla*_{KPC-2} gene (Fig. 1).

Moreover, all Inc groups described by Caratolli et al. [3] were investigated on the PlasmidFinder platform, however only IncA/C₂ complex, IncL/M and IncQ1 were detected (Fig. 1).

The general data from plasmid sequencing are described in Supplementary Tables S1–S4.

We suggest that NTE_{KPC-Ild} is the variant circulating in Recife-PE, Brazil, which can probably be found in other species of enterobacteria, as has already been found in *K. pneumoniae* [4]. In addition, there is the probable emergence of new variants of NTE_{KPC} that can lose or potentiate the transposition capacity of the *bla*_{KPC-2} gene, for which further studies on NTE_{KPC} are required. However, this does not prevent this gene from being transferred by conjugative and non-conjugative plasmids.

Although non-conjugative, the small, promiscuous IncQ plasmids have the ability to be mobilisable. IncQ1 plasmids bearing the *bla*_{KPC} gene have been reported in *K. pneumoniae* and *Pseudomonas aeruginosa* in Brazil [4,5]. This fact is an indication that IncQ was the plasmid responsible for disseminating the *bla*_{KPC} gene in Brazil. In addition, both sequenced isolates showed resistance to amikacin as well as carbapenems. This can be explained by the association of *bla*_{KPC-2} and *aph(3')-Vli* in the same IncQ plasmid in the two sequenced isolates, since they were present in the same contig, thus showing the association of relevant resistance mechanisms in plasmids of *K. aerogenes*.

The *bla*_{SCO-1} gene confers resistance to penicillins and, to a lesser extent, cephalosporins and carbapenems [6]. *bla*_{SCO-1} has been found in different species, however this is the first report of this gene in *K. aerogenes*.

These findings demonstrate the ability of *K. aerogenes* to accumulate resistance determinants using plasmids, such as IncA/C₂ and IncL/M, as well as the presence of a mobilisable and promiscuous plasmid, such as IncQ1. In addition, the presence of different resistance genes inserted in the *K. aerogenes* plasmids observed in this study emphasises the importance of monitoring the circulation of plasmids and this bacterial species in the hospital environment. This is still neglected and little investigated. The observed changes emphasise the continuous recombination and

<http://dx.doi.org/10.1016/j.jgar.2020.05.001>

2213-7165/© 2020 Published by Elsevier Ltd on behalf of International Society for Antimicrobial Chemotherapy. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

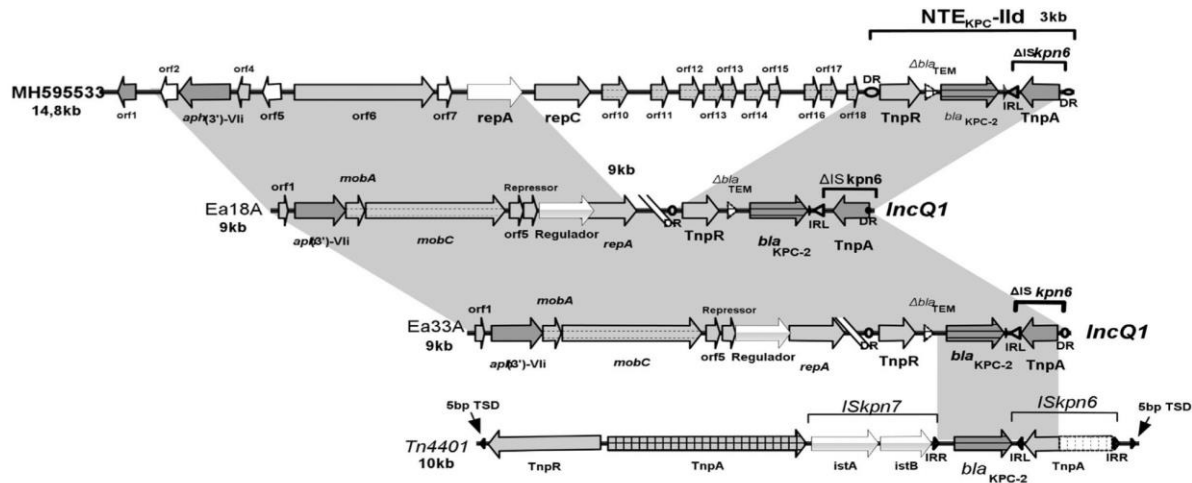


Fig. 1. Comparison of non-Tn4401 (NTE_{KPC}-Ild) elements carrying the *bla*_{KPC-2} gene in *Klebsiella aerogenes* isolates Ea18A and Ea33A and reference sequences deposited in GenBank (Tn4401, accession no. **KJ933392**; Tn3, accession no. **FJ628167**; and NTE_{KPC}-Ild, accession no. **MH595533**). Protein coding sequences are represented as arrows and are labelled with the name of the gene. Vertical lines represent gaps schematising the termination of one contig and the beginning of another contig. Grey shading represents homologous shared regions (>95%). The direct repeat sequence of NTE_{KPC}-Ild is represented by a circle.

evolution of the non-Tn4401 element (NTE_{KPC}) that can carry the *bla*_{KPC-2} gene in Brazil.

This study has some limitations regarding methodology. Hybridisation of resistance genes, circularisation of plasmids by sequencing of long sequences and conjugation could answer other questions about the plasmids and resistance genes found. However, for the purpose of the present work, with the methodology used it was possible to investigate the location of the *bla*_{KPC-2} and *aph*(3')-VII genes in an IncQ1 plasmid and to characterise the element NTE_{KPC}-Ild as well as the discovery of the *bla*_{SCO-1} and *sul*2 genes in the plasmid DNA of the isolates.

Funding

None declared.

Competing Interests

None declared.

Ethical approval

Not required.

Acknowledgments

The authors are grateful to the Central Laboratory of the CCB/UFPE-LABCE (Brazil) for sequencing, especially Prof. Dr Marcos Antonio de Moraes and Dr Heidi Lacerda.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jgar.2020.05.001>.

References

- [1] Cabral AB, Maciel MAV, Barros JF, Antunes MM, Barbosa CMM, Lopes ACS. Clonal spread and accumulation of β -lactam resistance determinants in *Enterobacter aerogenes* and *Enterobacter cloacae* complex isolates from infection and

colonization in patients at a public hospital in Recife, Pernambuco, Brazil. *J Med Microbiol* 2017;66:70–7, doi:<http://dx.doi.org/10.1099/jmm.0.000398>.

- [2] Monteiro J, Santos AF, Asensi MD, Peirano G, Gales AC. First report of KPC-2-producing *Klebsiella pneumoniae* strains in Brazil. *Antimicrob Agents Chemother* 2009;53:333–4, doi:<http://dx.doi.org/10.1128/AAC.00736-08>.
- [3] Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods* 2005;63:219–28, doi:<http://dx.doi.org/10.1016/j.mimet.2005.03.018>.
- [4] Cerdeira LT, Lam MMC, Wyres KL, Wick RR, Judd LM, Lopes R, et al. Small IncQ1 and Col-like plasmids harboring *bla*_{KPC-2} and non-Tn4401 elements (NTE_{KPC}-Ild) in high-risk lineages of *Klebsiella pneumoniae* CG258. *Antimicrob Agents Chemother* 2019;63, doi:<http://dx.doi.org/10.1128/AAC.02140-18> pii: e02140-18.
- [5] Jeannot K, Danassie M, Triponney P, Bour M, Gueudet T, Beyrouthy R, et al. A novel IncQ plasmid carrying gene *bla*_{CTX-M-3} in *Pseudomonas aeruginosa*. *J Antimicrob Chemother* 2019;74:823–5, doi:<http://dx.doi.org/10.1093/jac/dky454>.
- [6] Poirel L, Corvec S, Rapoport M, Mugnier P, Petroni A, Pasteran F, et al. Identification of the novel narrow-spectrum β -lactamase SCO-1 in *Acinetobacter* spp. from Argentina. *Antimicrob Agents Chemother* 2007;51:2179–84, doi:<http://dx.doi.org/10.1128/AAC.01600-06>.

Elizabeth Maria Bispo Beltrão

Érica Maria de Oliveira

Chrisllane Rafael dos Santos Vasconcelos

Universidade Federal de Pernambuco (UFPE), Recife-PE, Brazil

Adriane Borges Cabral

Universidade Estadual de Ciências da Saúde de Alagoas (UNCISAL),

Maceió-AL, Brazil

Antônio Mauro Rezende

Instituto Aggeu Magalhães (IAM), FIOCRUZ-PE, Recife-PE, Brazil

Ana Catarina Souza Lopes*

Universidade Federal de Pernambuco (UFPE), Recife-PE, Brazil

* Corresponding author. Present address: Centro de Ciências Médicas - CCM, Universidade Federal de Pernambuco (UFPE), Prof. Moraes Rego Ave., 50732-970 Recife-PE, Brazil.

E-mail addresses: ana.lopes.ufpe@gmail.com, ana.slopes@ufpe.br (A.C.S. Lopes).

Received 17 December 2019

Available online 20 May 2020



Received: 15 November 2020 | Revised: 24 January 2021 | Accepted: 25 January 2021

DOI: 10.1111/mmi.14688

MICROREVIEW

WILEY

Epitranscriptome machinery in Trypanosomatids: New players on the table?

Suellen Rodrigues Maran¹ | João Luiz de Lemos Padilha Pitta² | Crisllane Rafael dos Santos Vasconcelos² | Suzanne M. McDermott³ | Antonio Mauro Rezende² | Nilmar Silvio Moretti³

¹Laboratory of Molecular Biology of Pathogens, Department of Microbiology, Immunology and Parasitology, Federal University of São Paulo, São Paulo, Brazil

²Department of Microbiology, Aggeu Magalhães Institute/FIOCRUZ, Recife, Brazil

³Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA

Correspondence

Antonio Mauro Rezende, Department of Microbiology, Aggeu Magalhães Institute/FIOCRUZ, Recife, PE, Brazil.
Email: antonio.rezende@cpqam.fiocruz.br

Nilmar Silvio Moretti, Department of Microbiology, Immunology and Parasitology, Federal University of São Paulo, São Paulo, SP, Brazil.
Email: nilmar.moretti@unifesp.br

Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2018/09948-0 and 2019/13765-1; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 424729/2018-0; Fundação Oswaldo Cruz, Grant/Award Number: 400739/2019-4

Abstract

Trypanosoma and *Leishmania* parasites cause devastating tropical diseases resulting in serious global health consequences. These organisms have complex life cycles with mammalian hosts and insect vectors. The parasites must, therefore, survive in different environments, demanding rapid physiological and metabolic changes. These responses depend upon regulation of gene expression, which primarily occurs posttranscriptionally. Altering the composition or conformation of RNA through nucleotide modifications is one posttranscriptional mechanism of regulating RNA fate and function, and modifications including N6-methyladenosine (m6A), N1-methyladenosine (m1A), N5-methylcytidine (m5C), N4-acetylcytidine (ac4C), and pseudouridine (Ψ), dynamically regulate RNA stability and translation in diverse organisms. Little is known about RNA modifications and their machinery in Trypanosomatids, but we hypothesize that they regulate parasite gene expression and are vital for survival. Here, we identified Trypanosomatid homologs for writers of m1A, m5C, ac4C, and Ψ and analyze their evolutionary relationships. We systematically review the evidence for their functions and assess their potential use as therapeutic targets. This work provides new insights into the roles of these proteins in Trypanosomatid parasite biology and treatment of the diseases they cause and illustrates that Trypanosomatids provide an excellent model system to study RNA modifications, their molecular, cellular, and biological consequences, and their regulation and interplay.

KEYWORDS

ac4C, m1A, m5C, pseudouridine, RNA modification, Trypanosomatids

1 | INTRODUCTION

Trypanosomatids are unicellular protozoan organisms in the class Kinetoplastida. Most of the known species are parasites belonging to the *Trypanosoma* and *Leishmania* genera that have a huge public health and economic impact. The diseases that these parasites

cause are considered some of the most important neglected tropical diseases, affecting nearly 22 million people globally (Altamura et al., 2020; Buscher et al., 2017; Filardy et al., 2018; Okwor & Uzonna, 2016; Perez-Molina & Molina, 2018).

Trypanosomatid parasites have complex life cycles with mammalian hosts and insect vectors, meaning that the parasites must be

Suellen Rodrigues Maran and João Luiz de Lemos Padilha Pitta contributed equally to this work.



In and Outs of Chuviridae Endogenous Viral Elements: Origin of a Potentially New Retrovirus and Signature of Ancient and Ongoing Arms Race in Mosquito Genomes

Filipe Zimmer Dezordi¹, Chrisllane Rafael dos Santos Vasconcelos², Antonio Mauro Rezende² and Gabriel Luz Wallau^{1*}

¹ Departamento de Entomologia, Instituto Aggeu Magalhães (IAM), Fundação Oswaldo Cruz (FIOCRUZ), Recife, Brazil

² Departamento de Microbiologia, Instituto Aggeu Magalhães (IAM), Fundação Oswaldo Cruz (FIOCRUZ), Recife, Brazil

OPEN ACCESS

Edited by:

Horacio Naveira,
University of A Coruña, Spain

Reviewed by:

Jean-Michel Drezen,
Université de Tours, France
Carol D. Blair,
Colorado State University,
United States

*Correspondence:

Gabriel Luz Wallau
gabriel.wallau@cpqam.fiocruz.br;
gabriel.wallau@fiocruz.br

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 12 March 2020

Accepted: 28 September 2020

Published: 22 October 2020

Citation:

Dezordi FZ, Vasconcelos CRS,
Rezende AM and Wallau GL (2020) In
and Outs of Chuviridae Endogenous
Viral Elements: Origin of a Potentially
New Retrovirus and Signature
of Ancient and Ongoing Arms Race
in Mosquito Genomes.
Front. Genet. 11:542437.
doi: 10.3389/fgene.2020.542437

Background: Endogenous viral elements (EVEs) are sequences of viral origin integrated into the host genome. EVEs have been characterized in various insect genomes, including mosquitoes. A large EVE content has been found in *Aedes aegypti* and *Aedes albopictus* genomes among which a recently described *Chuviridae* viral family is of particular interest, owing to the abundance of EVEs derived from it, the discrepancy among the chuvirus endogenized gene regions and the frequent association with retrotransposons from the BEL-Pao superfamily. In order to better understand the endogenization process of chuviruses and the association between chuvirus glycoproteins and BEL-Pao retrotransposons, we performed a comparative genomics and evolutionary analysis of chuvirus-derived EVEs found in 37 mosquito genomes.

Results: We identified 428 EVEs belonging to the *Chuviridae* family confirming the wide discrepancy among the chuvirus genomic regions endogenized: 409 glycoproteins, 18 RNA-dependent RNA polymerases and one nucleoprotein region. Most of the glycoproteins (263 out of 409) are associated specifically with retroelements from the Pao family. Focusing only on well-assembled Pao retroelement copies, we estimated that 263 out of 379 Pao elements are associated with chuvirus-derived glycoproteins. Seventy-three potentially active Pao copies were found to contain glycoproteins into their LTR boundaries. Thirteen out of these were classified as complete and likely autonomous copies, with a full LTR structure and protein domains. We also found 116 Pao copies with no trace of glycoproteins and 37 solo glycoproteins. All potential autonomous Pao copies, contained highly similar LTRs, suggesting a recent/current activity of these elements in the mosquito genomes.

Conclusion: Evolutionary analysis revealed that most of the glycoproteins found are likely derived from a single or few glycoprotein endogenization events associated with a recombination event with a Pao ancestral element. A potential functional Pao-chuvirus hybrid (named Anakin) emerged and the glycoprotein was further replicated through

ANEXO B - Artigos Submetidos Para Publicação Durante o Período do Doutorado

Não Diretamente Relacionados a Tese.

***In silico* predictions of protein interactions between Zika virus and human host**

João Luiz de Lemos Padilha Pitta^{Corresp., Equal first author, 1}, Crhisllane Rafaele dos Santos Vasconcelos^{Equal first author, 2}, Gabriel da Luz Wallau³, Túlio de Lima Campos⁴, Antonio Mauro Rezende^{Corresp. 1}

¹ Microbiology Department, Aggeu Magalhães Institute - FIOCRUZ/PE, Recife, PE, Brasil

² Genetics Department, Federal University of Pernambuco, Recife, PE, Brasil

³ Entomology Department, Aggeu Magalhães Institute - FIOCRUZ/PE, Recife, PE, Brasil

⁴ Bioinformatics Platform, Aggeu Magalhães Institute - FIOCRUZ/PE, Recife, PE, Brasil

Corresponding Authors: João Luiz de Lemos Padilha Pitta, Antonio Mauro Rezende
Email address: joao.pitta@tjpe.jus.br, antonio.rezende@cpqam.fiocruz.br

Background. The ZIKA virus (ZIKV) belongs to the Flaviviridae family and was first isolated in the 40's, remained underreported until its global threat in 2016, where drastic consequences were reported as Guillan-Barre syndrome and microcephaly in newborns. Understanding molecular interactions of ZIKV proteins during host's infection is important to develop treatments and prophylactic measures. Large-scale experimental approaches normally used to detect protein-protein interaction (PPI) are onerous and labor intensive. On the other hands, computational methods may overcome these challenges and guide traditional approaches on one or few protein molecules. The prediction of PPIs can be used to study the host-parasite interactions at the protein level and reveal key pathways that allows viral infection.

Results. Applying Random Forest and Support Vector Machine (SVM) algorithms, we performed predictions of PPI between two ZIKV strains and human proteomes. The consensus number of predictions of both algorithms was 17,223 pairs of proteins. Functional enrichment analyses were executed with the predicted networks in order to access the biological meanings on the protein interactions. Some pathways related to viral infection and neurological development were found for both ZIKV strains on the enrichment analysis, but JAK-STAT pathway was observed only for strain PE243 when compared with FSS13025 strain.

Conclusions. The consensus network of PPI predictions made by Random Forest and SVM algorithms allowed an enrichment analysis that corroborates with many aspects of ZIKV infection. The enrichment results are mainly related with viral infection, neuronal development and immune response and presented differences among the two ZIKV strains compared. Strain PE243 presented more predicted interactions between proteins from JAK-STAT signaling pathway, which could lead to a more intensive immune response when compared with strain FSS13025. These results show that the methodology employed here can reveals potentially new interactions between the ZIKV and humans' cells.

ANEXO C - Atividade de Divulgação Científica Realizada Durante o Período do Doutorado.

Relatório VPEIC-004-FIO-17-2-57 de Divulgação Científica - Edital 2018

Título: Scicast – porque a ciência tem que ser divertida

Responsável: Antonio Mauro Rezende,
antonio.rezende@cpqam.fiocruz.br,
contato@scicast.com.br,
81-991223436

1- Resumo

O desenvolvimento tecnológico juntamente com os avanços da internet permitiu o surgimento de um conceito novo de aprendizagem nomeado de *mobile learning* (m-learning). Uma das ferramentas desta nova área é o *podcasting*, um arquivo multimídia que vem sendo utilizado em universidades para possibilitar aos alunos o acesso a revisão de aulas [1], no meio artístico, para disseminação de ideias, músicas, discussões de filmes, e no meio científico para discussão de temas e/ou artigos (<https://www.nature.com/nature/articles?type=nature-podcast>).

Esta última possibilidade de utilização do podcast mostra a capacidade de aplicar esta ferramenta para estreitar os laços entre a comunidade e o meio científico. Tal alcance é de grande valia para ambas as partes, apresentando novas descobertas e o pensamento científico para o público e permitindo que cientistas pensem sobre as implicações sociais do seu tema.

Assim, o presente trabalho divulgou por meio de podcasts, as áreas de pesquisa da Fundação Oswaldo Cruz (FIOCRUZ), explanando cada tema com pesquisadores especializados na área juntamente com a equipe já especializada em podcasts. As gravações seguiram pautas desenvolvidas por especialistas, revisadas por pares, com participação de pesquisadores dos centros de pesquisa da FIOCRUZ. Dessa forma, o projeto contou com a participação de pesquisadores, alunos e funcionários da FIOCRUZ, permitindo que esses colocassem em prática a divulgação científica ligada a seus projetos, levando assim o pensamento crítico científico a mais de 100 mil ouvintes.

2- Principais resultados

O principal objetivo do projeto consistiu em apresentar de maneira acessível a não especialistas algumas áreas de estudo da FIOCRUZ, resultando em 11 diferentes podcasts, sendo cada um referente a um tema específico.

O primeiro podcast, (disponível em: <https://www.deviantart.com.br/podcasts/scicast-268/>), abordou o tema de vacinas (Figura 1), explanando acerca da história sobre essa temática, o desenvolvimento das primeiras vacinas, as técnicas científicas utilizadas para descoberta e desenvolvimento de vacinas e sua importância. Esse episódio contou com a participação do pesquisador Lindomar José Pena do Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) foi ouvido por 122.643 pessoas.



Figura 1: Capa do episódio de vacinas, produzida pela equipe de arte do Scicast podcast.

O segundo podcast, (disponível em: <https://www.deviantart.com.br/podcasts/scicast-274/>), abordou a educação em saúde (Figura 2), explanando acerca do tema saúde/doença, educação sanitária, as práticas e articulações sociais, e discutindo ainda sobre as metodologias e políticas envolvidas. Esse episódio contou com a participação do pesquisador convidado Marcelo Henrique Santos Paiva do Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 106.713 *downloads*.



Figura 2: Capa do episódio de Educação em Saúde, produzida pela equipe de arte do Scicast podcast.

O terceiro podcast, (disponível em: <https://www.deviantart.com.br/podcasts/scicast-277/>), abordou as doenças negligenciadas (Figura 3), onde foi discutido quais são essas doenças, porque são negligenciadas, impacto das mesmas na saúde das populações, entraves para

superação de sua incidência e a relação entre doenças negligenciadas e pobreza. Esse episódio contou com a participação do pesquisador Antonio Mauro Rezende e da funcionária e aluna de Doutorado Michelle da Silva Barros, ambos vinculados ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 114.068 *downloads*.



Figura 3: Capa do episódio Doenças Negligenciadas, comprada pela equipe de arte do Scicast podcast.

O quarto podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-285/>), abordou o Zika Vírus (Figura 4), o que foi o surto do Zika, quais seus impactos à população, porque virou um surto, como a doença evoluiu no Brasil e no mundo, as chances de um novo surto e as possibilidades de vacina. Esse episódio contou com a participação do pesquisador convidado Marcelo Henrique Santos Paiva do Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 106.296 *downloads*.



Figura 4: Capa do episódio de Zika vírus, produzida pela equipe de arte do Scicast podcast.

O quinto podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-292/>), abordou a violência e saúde (Figura 5), onde foi realizada uma reflexão sobre a violência como questão de saúde pública, o que é violência, diferentes formas de incidência, qual o tamanho do impacto da violência na sociedade, como preveni-la e como remediá-la. Esse episódio contou com a participação do doutorando Moezio Vasconcellos vinculado ao programa de pós

graduação do Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 100.417 *downloads*.



Figura 5: Capa do episódio de Violência e Saúde, adquirida pela equipe de arte do Scicast podcast.

O sexto podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-304/>), abordou o Sistema Único de Saúde (SUS) (Figura 6), onde foi discutido acerca do histórico e criação do SUS, princípios e estrutura, impacto do investimento, como funciona, comparação com outros sistemas, problemas e as possíveis soluções. Esse episódio contou com a participação do pesquisador em saúde pública Rafael Moreira do Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 84.760 *downloads*.



Figura 6: Capa do episódio do Sistema Único de Saúde, adquirida pela equipe de arte do Scicast podcast.

O sétimo podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-311/>), abordou a saúde mental (Figura 7), onde foi discutido as principais doenças, possíveis causas, formas de tratamento, transtornos relacionados a profissões e a saúde mental no meio acadêmico. Esse episódio contou com a participação da doutoranda Crhisllane Vasconcelos vinculada ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 81.296 *downloads*.



Figura 7: Capa do episódio de Saúde Mental, produzida pela equipe de arte do Scicast podcast.

O oitavo podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-318/>), abordou a poluição (Figura 8), onde foi discutido o histórico de como a humanidade começou a focar nesse tema, principais causas, principais consequências, tecnologias disponíveis para reverter a situação e o que esperar do nosso futuro. Esse episódio contou com a participação da doutoranda Crhisllane Vasconcelos vinculada ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 30.700 *downloads*.



Figura 8: Capa do episódio de Poluição, adquirida pela equipe de arte do Scicast podcast.

O nono podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-325/>), abordou a área da epidemiologia (Figura 9), onde foi discutido o início da epidemiologia e sua importância, áreas de atuação de um epidemiologista, epidemiologia e saúde pública, causalidade, bioestatística, tipos de estudos e epidemiologia na prevenção de doenças. Esse episódio contou com a participação do doutorando Rodrigo Moraes Loyo Arcoverde e da pesquisadora Dra Elaine Christine de Souza Gomes, ambos vinculados ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 57.072 *downloads*.



Figura 9: Capa do episódio de Epidemiologia, produzida pela equipe de arte do Scicast podcast.

O décimo podcast, (disponível em: <https://www.deviant.com.br/podcasts/scicast-328/>), abordou os animais peçonhentos (Figura 10), onde foi discutido quais são, a química e fisiologia relacionada aos venenos, principais utilizações do veneno, como são desenvolvidos os antídotos, como proceder em situações de risco. Esse episódio contou com a participação da doutoranda Crhisllane Vasconcelos vinculada ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 59.105 *downloads*



Figura 10: Capa do episódio de Animais peçonhentos, produzida pela equipe de arte do Scicast podcast.

O décimo primeiro podcast não teve o apoio financeiro do edital, sendo assim um episódio extra disponibilizado pela equipe do scicast podcast em comemoração aos demais episódios, (disponível em: <https://www.deviant.com.br/podcasts/scicast-348/>), abordou o pesquisador brasileiro Oswaldo Cruz (Figura 11), onde foi discutido a vida/biografia de Oswaldo Cruz, contando seu histórico de pesquisador, suas áreas de pesquisa, os demais pesquisadores que trabalharam em parceria com ele e a criação do instituto hoje nomeado de Fiocruz. Esse episódio contou com a participação da doutoranda Crhisllane Vasconcelos vinculada ao Instituto Aggeu Magalhães, Fiocruz-PE e até o momento (Outubro de 2020) conta com 48.560 *downloads*.

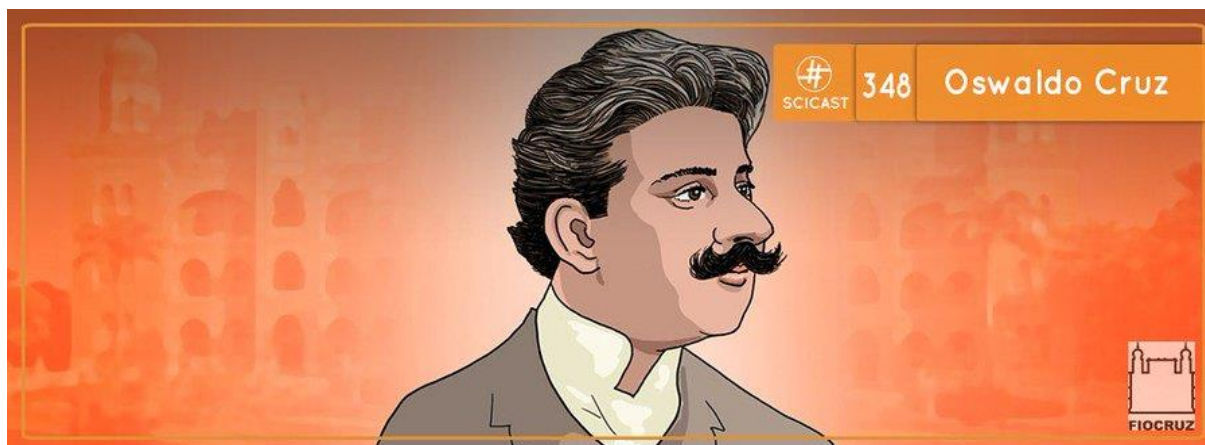


Figura 11: Capa do episódio sobre Oswaldo Cruz, produzida pela equipe de arte do Scicast podcast.

Deste modo, o projeto finalizou com a produção de 11 episódios, com participação de 9 pesquisadores/alunos do Instituto Aggeu Magalhães/Fiocruz-PE e um total de *downloads* até esse momento de 911.603, com uma média de ouvintes de aproximadamente 83.000 pessoas por episódio. Todos os episódios produzidos também estão disponíveis no *Spotify*: (https://open.spotify.com/show/0qfFcilKpNKkXy8TbZ4moP?si=1YkbfjasSPSBiyStgbF_GQ) ,

3- Desafios importantes encontrados

Por contar com uma empresa e equipe já experientes para trabalhar com esse tipo de mídia, não foi enfrentado nenhum problema técnico, sendo o maior desafio vinculado a encontrar pesquisadores disponíveis para participar do projeto e prepará-los para apresentar sua pesquisa para um público de não especialistas.

4-Texto em linguagem de acesso e compreensão para pessoas não especializadas

O *Podcasting* é uma das ferramentas do *mobile learning* onde se utiliza um dispositivo eletrônico, como um celular, para ouvir ou assistir a transmissão de mídia audiovisual, podendo esta ser somente áudio, com inegável analogia a programas de rádio gravados. As transmissões são disponibilizadas na Internet e podem ser baixadas no dispositivo com o auxílio de aplicativos agregadores ou simplesmente acessando o *site* onde a mídia está armazenada [1].

Por mais que sua ampla utilização seja algo novo, o conceito de educação através de gravações de áudio já possui mais de 50 anos, com relatos na literatura de sua utilização na área médica nos anos de 1960 [2]. Com os avanços tecnológicos e o acesso mais amplo a tais mídias, hoje, esta metodologia vem sendo relatada como facilitadora de aprendizagem, sendo empregada desde ensino de uma segunda língua [3], até na aproximação de neurocientistas com a comunidade [4].

A ascensão dos *podcasts* fez com que esta metodologia atingisse a divulgação científica mais ampla, sendo hoje produzido por muitas revistas científicas importantes, onde os artigos com mais destaque do mês são discutidos de forma mais simples permitindo uma compreensão

maior dos ouvintes. Grandes exemplos são os podcasts disponibilizados pela revista Nature (<https://www.nature.com/nature/articles?type=nature-podcast>), Cell (<https://www.cell.com/podcast>) e Science (<http://www.sciencemag.org/podcasts>).

No Brasil, um dos principais podcasts de divulgação científica é o Scicast, que possui como premissa: A ciência tem que ser divertida. Neste projeto, especialistas de diversas áreas desenvolvem programas que abordam temas científicos voltados para compreensão do público não especialista.

Dessa forma, o presente estudo contou com a parceria entre o Scicast, com experiência na mídia *podcast*, e pesquisadores do Instituto Fiocruz para produção de episódios com aproximadamente 1 hora e 40 minutos cada, abordando áreas como: Educação em Saúde, Doenças negligenciadas, Zika e atualizações, Violência e Saúde, Sistema Único de Saúde, Poluição, Saúde mental, Vacinas, Animais Peçonhentos, Epidemiologia e a vida de Oswaldo Cruz. Assim, áreas de estudo da Fiocruz puderam ser divulgadas de maneira acessível a todos, resultando em 911.603 *downloads*, com uma média de 83.000 ouvintes por episódio.

Referências:

- [1] C. Evans, “The effectiveness of m-learning in the form of podcast revision lectures in higher education,” *Computers & Education*, vol. 50, no. 2, Feb. 2008.
- [2] S. Fletcher and A. A. Watson, “Magnetic tape recording in the teaching of histopathology.,” *British journal of medical education*, vol. 2, no. 4, pp. 283–292, Dec. 1968.
- [3] M. Rahimi and M. Katal, “The role of metacognitive listening strategies awareness and podcast-use readiness in using podcasting for learning English as a foreign language,” *Computers in Human Behavior*, vol. 28, no. 4, Jul. 2012.
- [4] J. Illes *et al.*, “Neurotalk: improving the communication of neuroscience research.,” *Nature reviews. Neuroscience*, vol. 11, no. 1, pp. 61–69, Jan. 2010.

ANEXO D - Curriculum vitae (Lattes)

Crhisllane Rafaela dos Santos Vasconcelos

Curriculum Vitae

Nome civil

Nome Crhisllane Rafaela dos Santos Vasconcelos

Dados pessoais

Nascimento 28/12/1992 - Brasil
Carteira de Identidade 8795741 SDS - PE - 16/11/2010
CPF 095.999.454-84

Formação acadêmica/titulação

- 2016 - 2020** Doutorado em Genética.
 Universidade Federal de Pernambuco, UFPE, Recife, Brasil
 Título: Avaliação in silico do proteoma predito de Leishmania sp. para seleção de alvos terapêuticos
 Orientador: Antonio Mauro Rezende
 Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- 2014 - 2016** Mestrado em Genética.
 Universidade Federal de Pernambuco, UFPE, Recife, Brasil
 Título: Predição de redes de interação protéica a partir de informações estruturais de proteínas preditas em genomas de espécies de Leishmania, Ano de obtenção: 2016
 Orientador: Antonio Mauro Rezende
 Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- 2013 - 2014** Especialização em Diagnóstico Molecular.
 Faculdade Pernambucana de Saúde, FPS, Recife, Brasil
 Título: Avaliação de NEK quinases como alvo farmacológico contra leishmanioses por métodos em bioinformática
 Orientador: Rafael de Freitas e Silva
- 2010 - 2013** Graduação em Biomedicina.
 Associação Caruaruense de Ensino Superior, ASCES, Caruaru, Brasil
 Título: Perfil hepático de dependentes da substância simpaticomimética cocaína, em processo terapêutico.
 Orientador: Franklin Barbalho Magalhães

Formação complementar

- 2019 - 2019** Curso de curta duração em Identifying Gene Regulatory Networks from Gene Expression Data. (Carga horária: 40h).
 Instituto René Rachou - Fiocruz Minas, FIOCRUZ MINAS, Brasil
- 2018 - 2018** Eukaryotic Pathogen Genomics Resource Workshop, . .
 University of Georgia, UGA, Athens, Estados Unidos
- 2018 - 2018** Molecular Biophysical Chemistry. . (Carga horária: 60h).
 Instituto Aggeu Magalhães - Fiocruz PE, IAM-FIOCRUZ, Brasil

- 2016 - 2016** Ferramentas de bioinformática aplicadas à citogenética moderna. .
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
- 2015 - 2015** Curso de curta duração em Análise de dados NGS. (Carga horária: 12h).
Genomika Diagnósticos, GENOMIKA , Recife, Brasil
- 2014 - 2014** Escola de Modelagem Molecular em Sistemas Biologic. .
Laboratório Nacional de Computação Científica, LNCC, Petropolis, Brasil
- 2014 - 2014** Curso de curta duração em Predição de Estrutura de Proteínas. (Carga horária: 6h).
Laboratório Nacional de Computação Científica, LNCC, Petropolis, Brasil
- 2012 - 2012** Curso de curta duração em Educadores Sociais. (Carga horária: 16h).
Amigos do Transplante de Medula Óssea, ATMO, Brasil
- 2012 - 2012** Curso de curta duração em Oncohematologia (Teórico-prático). (Carga horária: 4h).
Associação Caruaruense de Ensino Superior, ASCES, Caruaru, Brasil
- 2012 - 2012** Curso de curta duração em Desenvolvimento de Marcadores de Microssatélite. (Carga horária: 4h).
Associação Caruaruense de Ensino Superior, ASCES, Caruaru, Brasil
- 2011 - 2011** Curso de curta duração em Biologia Molecular no Diagnóstico de Doenças. (Carga horária: 20h).
Aprimora Cursos, A C, Brasil
- 2011 - 2011** Curso de curta duração em Parasitologia Clínica. (Carga horária: 80h).
Portal Educação, PE, Brasil
- 2010 - 2010** Curso de curta duração em Hemograma: Tecnologia e Interpretação. (Carga horária: 20h).
Associação Caruaruense de Ensino Superior, ASCES, Caruaru, Brasil

Atuação profissional

1. Instituto Aggeu Magalhães - Fiocruz PE - IAM-FIOCRUZ

Vínculo institucional

- 2018 - 2018** Vínculo: Professor Convidado , Enquadramento funcional: Docência -
Montagem e Anotação de Genomas, Regime: Parcial
- 2018 - 2018** Vínculo: Professor Convidado , Enquadramento funcional: Docência -
Engenharia Genética, Regime: Parcial
- 2017 - 2017** Vínculo: Professor Convidado , Enquadramento funcional: Docência -
Engenharia Genética, Regime: Parcial
- 2017 - 2017** Vínculo: Professor Convidado , Enquadramento funcional: Docência -
Montagem e Anotação de Genomas, Regime: Parcial

2. Universidade Estadual de Campinas - UNICAMP

Vínculo institucional

- 2020 - Atual** Vínculo: Bolsista , Enquadramento funcional: Bioinformatics Scientist at the
GCCRC , Carga horária: 40, Regime: Integral

3. Universidade Federal de Pernambuco - UFPE

Vínculo institucional

2017 - 2017 Vínculo: Estágio , Enquadramento funcional: Docência - Genética Humana I ,
Carga horária: 30, Regime: Parcial

Revisor de periódico

1. MOLECULAR BIOLOGY REPORTS -

Vínculo

2019 - Atual Regime: Parcial

Áreas de atuação

1. Bioinformática
2. Genética Molecular e de Microorganismos

Idiomas

Inglês	Compreende Razoavelmente , Fala Razoavelmente , Escreve Razoavelmente , Lê Razoavelmente
Francês	Compreende Pouco , Fala Pouco , Escreve Pouco , Lê Pouco
Português	Compreende Bem , Fala Bem , Escreve Bem , Lê Bem

Prêmios e títulos

2015	ISMB 2015 Travel Fellowship Award, International Society for Computational Biology
2015	1º Lugar Projeto de Mestrado - V Jornada de Pós-Graduação em Genética, Universidade Federal de Pernambuco
2014	Projeto de Pesquisa - 1º lugar, Associação Caruaruense de Ensino Superior
2014	Projeto de Pesquisa - 2º Lugar, Associação Caruaruense de Ensino Superior
2013	Relatos de Estágios, Associação Caruaruense de Ensino Superior

Produção

Produção bibliográfica

Artigos completos publicados em periódicos

1. BISPO BELTRÃO, ELIZABETH MARIA; DE OLIVEIRA, ÉRICA MARIA; **DOS SANTOS VASCONCELOS, CRHISLLANE RAFAELE**; CABRAL, ADRIANE BORGES; REZENDE, ANTÔNIO MAURO; SOUZA LOPES, ANA CATARINA
Multidrug-resistant *Klebsiella aerogenes* clinical isolates from Brazil carrying IncQ1 plasmids containing the blaKPC-2 gene associated with non-Tn4401 elements (NTEKPC-IId). Journal of Global Antimicrobial Resistance. , v.22, p.43 - 44, 2020.

2. REZENDE, TATIANA MARIA TEODORO; REZENDE, ANTONIO MAURO; LUZ WALLAU, GABRIEL; **SANTOS VASCONCELOS, CRHISLLANE RAFAELE**; DE-MELO-NETO, OSVALDO POMPÍLIO; SILVA-FILHA, MARIA HELENA NEVES LOBO; ROMÃO, TATIANY PATRÍCIA
A differential transcriptional profile by *Culex quinquefasciatus* larvae resistant to *Lysinibacillus sphaericus* IAB59 highlights genes and pathways associated with the resistance phenotype. *Parasites & Vectors.* , v.12, p.1/407 - 17, 2019.
3. **DOS SANTOS VASCONCELOS, CRHISLLANE RAFAELE**; DE LIMA CAMPOS, TÚLIO; REZENDE, ANTONIO MAURO
Building protein-protein interaction networks for *Leishmania* species through protein structural information. *BMC BIOINFORMATICS.* , v.19, p.85 - 97, 2018.
4. DE SÁ CAVALCANTI, FELIPE LIRA; MENDES-MARQUES, CARINA LUCENA; **VASCONCELOS, CRHISLLANE RAFAELE DOS SANTOS**; DE LIMA CAMPOS, TÚLIO; REZENDE, ANTONIO MAURO; XAVIER, DANILO ELIAS; LEAL, NILMA CINTRA; DE-MELO-NETO, OSVALDO POMPÍLIO; DE MORAIS, MARCIA MARIA CAMARGO; LEAL-BALBINO, TEREZA CRISTINA
High frequency of OXA-253-producing *Acinetobacter baumannii* in different hospitals in Recife, Brazil: a new threat?. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY.* , v.61, p.AAC.01309-16 - , 2017.
5. DO NASCIMENTO, NATHALY ALEXANDRE; FERREIRA, LÍGIA MARIA; ROMÃO, TATIANY PATRÍCIA; DA CONCEIÇÃO CORREIA, DARLEIDE MARIA; **DOS SANTOS VASCONCELOS, CRHISLLANE RAFAELE**; REZENDE, ANTONIO MAURO; COSTA, SAMARA GRACIANE; GENTA, FERNANDO ARIEL; DE-MELO-NETO, OSVALDO POMPÍLIO; NEVES LOBO SILVA-FILHA, MARIA HELENA
N-glycosylation influences the catalytic activity of mosquito α -glucosidases associated with susceptibility or refractoriness to *Lysinibacillus sphaericus*. *INSECT BIOCHEMISTRY AND MOLECULAR BIOLOGY.* , v.1, p.1 - 8, 2016.

Artigos aceitos para publicação

1. DEZORDI, F. Z.; **VASCONCELOS, C. R. S.**; REZENDE, A. M.; LUZ WALLAU, GABRIEL
In and outs of *Chuviridae* endogenous viral elements: origin of a potentially new retrovirus and signature of ancient and ongoing arms race in mosquito genomes. *Frontiers in Genetics.* , 2020.

Trabalhos publicados em anais de eventos (resumo)

1. VASCONCELOS, C. R. S.; REZENDE, A. M.
Predição de redes de interação proteica a partir de informações estruturais de proteínas preditas em genomas de espécies de *Leishmania* In: XXI Encontro de Genética do Nordeste, 2016, RECIFE.
XXI Encontro de Genética do Nordeste - ENGGEN - Anais 2016. , 2016.

Apresentação de trabalho e palestra

1. **VASCONCELOS, C. R. S.**; REZENDE, A. M.
Avaliação In Silico Do Proteoma Predito De Leishmania Sp. Para Seleção De Alvos Terapêuticos, 2018.
(Outra, Apresentação de Trabalho)
2. OLIVEIRA, E. M.; MEDEIROS, R. D. A.; BELTRAO, E. M. B.; **VASCONCELOS, C. R. S.**; FIRMO, E. F.; REZENDE, A. M.; LOPES, A. C. S.
***Klebsiella pneumoniae* portadora de blaKPC-2, aph(3) e VIF, plasmídeos e genes de virulência, isolada de infecção urinária com desfecho de óbito em Recife-PE, Brasil.**, 2018.
(Congresso, Apresentação de Trabalho)
3. **VASCONCELOS, C. R. S.**; REZENDE, A. M.
Systematic In Silico Evaluation Of Leishmania Sp. Proteome For Drug Discovery, 2018. (Conferência ou

palestra,Apresentação de Trabalho)

4. VASCONCELOS, C. R. S.

Bioinformática aplicada a seleção de alvos para fármacos, 2017. (Conferência ou palestra,Apresentação de Trabalho)

Produção técnica

Trabalhos técnicos

1. VASCONCELOS, C. R. S.

Avaliação de trabalhos/poster - V SEMANA DE BIOCIEÊNCIAS E BIOTECNOLOGIA EM SAÚDE: Ciência a Serviço da Sociedade, 2017

Entrevistas, mesas redondas, programas e comentários na mídia

1. VASCONCELOS, C. R. S.; SANTOS, M. V. C.

Divulgação Científica e o mundo do Podcast, 2019

Educação e Popularização de C&T

Apresentação de trabalho e palestra

1. VASCONCELOS, C. R. S.; REZENDE, A. M.

Avaliação In Silico Do Proteoma Predito De Leishmania Sp. Para Seleção De Alvos Terapêuticos, 2018. (Outra,Apresentação de Trabalho)

2. VASCONCELOS, C. R. S.; REZENDE, A. M.

Systematic In Silico Evaluation Of Leishmania Sp. Proteome For Drug Discovery, 2018. (Conferência ou palestra,Apresentação de Trabalho)

Curso de curta duração ministrado

1. VASCONCELOS, C. R. S.; TRAJANO-SILVA, L. A. M.

Bioinformática Estrutural para Iniciantes, 2017. (Outro, Curso de curta duração ministrado)

Redes sociais, websites e blogs

1. VASCONCELOS, C. R. S.

Contrafactual #140: E se o acidente de Chernobyl não tivesse sido contido?, 2019. (Site, Mídias sociais, websites, blogs)

2. VASCONCELOS, C. R. S.

O que são os híbridos Humano-Animal, 2019. (Site, Mídias sociais, websites, blogs)

3. VASCONCELOS, C. R. S.; SILVA, I. G. S.

Reimagine o Câncer #01: O que é Câncer?, 2019. (Site, Mídias sociais, websites, blogs)

4. VASCONCELOS, C. R. S.

Reimagine o Câncer #03: Fisiopatologia do Câncer, 2019. (Site, Mídias sociais, websites, blogs)

5. VASCONCELOS, C. R. S.; SILVA, I. G. S.

SciCast #304: Sistema Único de Saúde – SUS, 2019. (Site, Mídias sociais, websites, blogs)

6. VASCONCELOS, C. R. S.; BACCHI, A. D.
SciCast #311: Saúde Mental, 2019. (Site, Mídias sociais, websites, blogs)
7. VASCONCELOS, C. R. S.; WARD, F.; KROHLING, W.
SciCast #318 - Poluição, 2019. (Site, Mídias sociais, websites, blogs)
8. VASCONCELOS, C. R. S.
SciCast #335: Chernobyl – Parte I, 2019. (Site, Mídias sociais, websites, blogs)
9. VASCONCELOS, C. R. S.
SciCast #336: Chernobyl – Parte II, 2019. (Site, Mídias sociais, websites, blogs)
10. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #591 - Vírus tem vida social?, 2019. (Site, Mídias sociais, websites, blogs)
11. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #487 : Há cura para o HIV?, 2019. (Site, Mídias sociais, websites, blogs)
12. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #538 - É possível uma IA transformar sinais cerebrais em fala?, 2019. (Site, Mídias sociais, websites, blogs)
13. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #550 - Como a terapia gênica poderia salvar o “Jimmy bolha”?, 2019. (Site, Mídias sociais, websites, blogs)
14. VASCONCELOS, C. R. S.
A amizade entre humanos e um café, 2018. (Site, Mídias sociais, websites, blogs)
15. VASCONCELOS, C. R. S.; BACCHI, A. D.
Desmistificando os alimentos transgênicos [Synapsando #5], 2018. (Site, Mídias sociais, websites, blogs)
16. VASCONCELOS, C. R. S.
Imortalidade – Como tentamos virar a ampulheta da vida?, 2018. (Site, Mídias sociais, websites, blogs)
17. VASCONCELOS, C. R. S.
SciCast #239: Química Orgânica, 2018. (Site, Mídias sociais, websites, blogs)
18. VASCONCELOS, C. R. S.
SciCast #240: Cronofarmacologia, 2018. (Site, Mídias sociais, websites, blogs)
19. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #241: A História do DNA, 2018. (Site, Mídias sociais, websites, blogs)
20. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #246: A História do DNA 2, 2018. (Site, Mídias sociais, websites, blogs)
21. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #252: A Ciência da Felicidade, 2018. (Site, Mídias sociais, websites, blogs)
22. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #264: Imunologia, 2018. (Site, Mídias sociais, websites, blogs)
23. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #268: Vacinas, 2018. (Site, Mídias sociais, websites, blogs)
24. VASCONCELOS, C. R. S.; COSTA, T. B.; PAIVA, M. H. S.
SciCast #274: Educação em Saúde, 2018. (Site, Mídias sociais, websites, blogs)

25. VASCONCELOS, C. R. S.; COSTA, T. B.; REZENDE, A. M.; BARROS, M. S.
SciCast #277: Doenças Negligenciadas, 2018. (Site, Mídias sociais, websites, blogs)
26. VASCONCELOS, C. R. S.; COSTA, T. B.
SciCast #283: Antibióticos, 2018. (Site, Mídias sociais, websites, blogs)
27. VASCONCELOS, C. R. S.; COSTA, T. B.; PAIVA, M. H. S.
SciCast #285: Zika, 2018. (Site, Mídias sociais, websites, blogs)
28. VASCONCELOS, C. R. S.
Semana da saúde 2018 – O Origami de DNA, 2018. (Site, Mídias sociais, websites, blogs)
29. VASCONCELOS, C. R. S.
Séries no Lab: Genética e Psicologia em Black Mirror, 2018. (Site, Mídias sociais, websites, blogs)
30. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #114: Nanopartículas, Neurônios, Vírus e Alzheimer, 2018. (Site, Mídias sociais, websites, blogs)
31. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #126: Gripe Espanhola e o Blues do DNA, 2018. (Site, Mídias sociais, websites, blogs)
32. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #147: A Ciência da Fake News, 2018. (Site, Mídias sociais, websites, blogs)
33. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #165: Sanidade Mental na Academia e Humor no laboratório, 2018. (Site, Mídias sociais, websites, blogs)
34. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #179: Sequenciamento genético e Bactéria comedora de carne humana, 2018. (Site, Mídias sociais, websites, blogs)
35. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #197: Neurônios, CRISPR e Vírus, 2018. (Site, Mídias sociais, websites, blogs)
36. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #211: Bactérias, Fezes, Câncer, 2018. (Site, Mídias sociais, websites, blogs)
37. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin de Notícias #83: Edição genética, Bactéria mutante e Internet das Células, 2018. (Site, Mídias sociais, websites, blogs)
38. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #285: Novos Medicamentos e Zika vírus, 2018. (Site, Mídias sociais, websites, blogs)
39. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #322: Aspirina e Células-tronco, 2018. (Site, Mídias sociais, websites, blogs)
40. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #332: Anticorpos e Bactérias, 2018. (Site, Mídias sociais, websites, blogs)
41. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #367: A Genética de todo o mundo e Expressão Gênica, 2018. (Site, Mídias sociais, websites, blogs)
42. VASCONCELOS, C. R. S.; COSTA, T. B.
Spin #377: Machine learning e Células tronco, 2018. (Site, Mídias sociais, websites, blogs)
43. VASCONCELOS, C. R. S.
Como as descobertas feitas pelos cientistas chegam até você, leitor?, 2017. (Site, Mídias sociais, websites, blogs)

blogs)

44. VASCONCELOS, C. R. S.; COSTA, T. B.

SciCast #226: Sexo, 2017. (Site, Mídias sociais, websites, blogs)

45. VASCONCELOS, C. R. S.; COSTA, T. B.

Spin de Notícias #24: CRISPR e seus usos, Cardiomiopatia Hipertrófica, a falta do Nobel e Gifs, 2017. (Site, Mídias sociais, websites, blogs)

46. VASCONCELOS, C. R. S.; COSTA, T. B.

Spin de Notícias #38: Cobertura Vacinal contra Gripe no Brasil e nova Cepa de H7N9, 2017. (Site, Mídias sociais, websites, blogs)

47. VASCONCELOS, C. R. S.; COSTA, T. B.

Spin de Notícias #56: Erros em publicações científica e Retratações, 2017. (Site, Mídias sociais, websites, blogs)

Demais produções técnicas

1. VASCONCELOS, C. R. S.; TRAJANO-SILVA, L. A. M.

Bioinformática Estrutural para Iniciantes, 2017. (Outro, Curso de curta duração ministrado)

Orientações e supervisões concluídas

Iniciação científica

1. Raul Santos Silva. **Avaliação In Silico Do Proteoma Predito De Leishmania Braziliensis Para Seleção De Alvos Terapêuticos**. 2016. Iniciação científica - Universidade Federal de Pernambuco
Inst. financiadora: Conselho Nacional de Desenvolvimento Científico e Tecnológico

Eventos

Eventos

Participação em eventos

1. **54 Congresso Da Sociedade Brasileira De Medicina Tropical**, 2018. (Congresso)
Variabilidade Da Região Intergenic Spacer Do Rna Ribossomal De 11 Diferentes Espécies De Leishmania.
2. **VIII Jornada de Pós-Graduação em Genética**, 2018. (Outra)
Avaliação In Silico Do Proteoma Predito De Leishmania Sp. Para Seleção De Alvos Terapêuticos.

Organização de evento

1. CAMPOS, T. L.; VASCONCELOS, C. R. S.; PITTA, J. L. L. P.
I workshop em Bioinformática do CAV, 2017. (Outro, Organização de evento)

Bancas

Bancas

Participação em banca de trabalhos de conclusão

Graduação

1. VASCONCELOS, C. R. S.

Participação em banca de Ana Paula Mariano Ramos. **Avaliação do teste molecular rápido GENEXPERT MTB/RIF na rotina do diagnóstico da tuberculose de uma laboratório municipal de saúde pública do Recife**, 2017

(Biomedicina) Universidade Federal de Pernambuco

Totais de produção

Produção bibliográfica

Artigos completos publicados em	
periódico.....	5
Artigos aceitos para	
publicação.....	1
Revistas	
(Magazines).....	1
Trabalhos publicados em anais de	
eventos.....	6
Apresentações de trabalhos (Conferência ou	
palestra).....	2
Apresentações de trabalhos	
(Congresso).....	4
Apresentações de trabalhos	
(Simpósio).....	1
Apresentações de trabalhos	
(Outra).....	5

Produção técnica

Trabalhos técnicos	
(parecer).....	1
Curso de curta duração ministrado	
(extensão).....	1
Curso de curta duração ministrado	
(outro).....	4
Programa de Rádio ou TV (mesa	
redonda).....	1
Site.....	47
..	

Orientações

Orientação concluída (iniciação	
científica).....	1

Eventos

Participações em eventos	
(congresso).....	4
Participações em eventos	
(simpósio).....	1
Participações em eventos	
(encontro).....	2
Participações em eventos	
(outra).....	11
Organização de evento	
(outro).....	1
Participação em banca de trabalhos de conclusão	
(graduação).....	1