UNIVERSIDADE FEDERAL DE GOIÁS
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL
E SAÚDE PÚBLICA

CLEBER CAMILO DE MELO FILHO

Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal

Goiânia
2018

sistema de bibliotecas ufg

PRPG
PRÓ-REITORIA DE PÓS-GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[ ] Dissertação** **[x] Tese**

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Cleber Camilo de Melo Filho

Título do trabalho: Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Cleber Camilo de Melo Filho
Assinatura do(a) autor(a)[2]

Ciente e de acordo:

Assinatura do(a) orientador(a)[2] Data: 27 / 05 / 2019

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

[2] A assinatura deve ser escaneada.

CLEBER CAMILO DE MELO FILHO

Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Medicina Tropical e Saúde Pública da Universidade Federal de Goiás para obtenção do Título de Doutor em Medicina Tropical e Saúde Pública.

Orientadora: Profa. Dra. Carolina Horta Andrade

Coorientador: Dr. Rodolpho de Campos Braga

Goiânia
2018

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Melo Filho, Cleber Camilo de
Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal [manuscrito] / Cleber Camilo de Melo Filho. - 2018.
xix, 164 f.: il.

Orientador: Profa. Dra. Carolina Horta Andrade; co-orientador Dr. Rodolpho de Campos Braga.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Patologia Tropical e Saúde Pública (IPTSP), Programa de Pós Graduação em Medicina Tropical e Saúde Pública, Goiânia, 2018.
Bibliografia. Anexos.
Inclui siglas, abreviaturas, símbolos, tabelas, lista de figuras, lista de tabelas.

1. Triagem virtual. 2. Descoberta de fármacos. 3. Trypanosoma cruzi . 4. Leishmania. 5. Tratamento. I. Andrade, Carolina Horta, orient. II. Título.

CDU 576.8

Programa de Pós-Graduação em Medicina Tropical e Saúde Pública
da Universidade Federal de Goiás

BANCA EXAMINADORA DA TESE DE DOUTORADO

Aluno: Cleber Camilo de Melo Filho

Orientadora: Profa. Dra. Carolina Horta Andrade

Coorientador: Dr. Rodolpho de Campos Braga

Membros:

1. **Dra. Carolina Horta Andrade (FF/UFG)**
2. **Dra. Ana Maria de Castro (IPTSP/UFG)**
3. **Dr. Milton Adriano Pelli de Oliveira (IPTSP/UFG)**
4. **Dra. Elisângela de Paula Silveira Lacerda (ICB/UFG)**
5. **Dr. Marcus Tullius Scotti (DQ/UFPB)**

Data: 02/04/2018

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE PATOLOGIA TROPICAL E SAÚDE PÚBLICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL E SAÚDE PÚBLICA
Rua 235, s/n – Setor Universitário - Goiânia/GO – CEP: 74.605-050
Fones: (62) 3209.6362 - 3209.6102 – Fax: (62) 3209.6363 - e-mail : ppgmtsp.ufg@gmail.com

ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE CLEBER CAMILO DE MELO FILHO - Aos dois dias do mês de abril do ano de 2018 (02/04/2018), às 8:00 horas, reuniram-se os componentes da Banca Examinadora: Profs. Drs. CAROLINA HORTA ANDRADE, ANA MARIA DE CASTRO, MILTON ADRIANO PELLI DE OLIVEIRA, MARCUS TULLIUS SCOTTI e ELISANGELA DE PAULA SILVEIRA LACERDA, para, sob a presidência da primeira, e em sessão pública realizada no CENTRO DE EVENTOS DA UFG/CAMPUS SAMAMBAIA, procederem à avaliação da defesa de tese intitulada: **"IDENTIFICAÇÃO DE NOVOS COMPOSTOS BIOATIVOS CONTRA TRIPANOSSOMATÍDEOS A PARTIR DE ESTRATÉGIAS INTEGRADAS EM QUÍMICA MEDICINAL"** em nível de **DOUTORADO**, área de concentração em **PARASITOLOGIA**, de autoria de **CLEBER CAMILO DE MELO FILHO** discente do PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL E SAÚDE PÚBLICA, da Universidade Federal de Goiás. A sessão foi aberta pela Orientadora Profa. Dra. CAROLINA HORTA ANDRADE, que fez a apresentação formal dos membros da Banca e orientou a Candidata sobre como utilizar o tempo durante a apresentação de seu trabalho. A palavra a seguir, foi concedida ao autor da tese que, em 30 minutos, procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca argüiu o Candidato, tendo-se adotado o sistema de diálogo seqüencial. Terminada a fase de argüição, procedeu-se à avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1034/2014 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Medicina Tropical e Saúde Pública a Banca, em sessão secreta, expressou seu Julgamento, considerando o candidato **Aprovado** ou **Reprovado:**

Banca Examinadora	**Aprovado / Reprovado**
Dra. Carolina Horta Andrade	Aprovado
Dr. Milton Adriano Pelli de Oliveira	APROVADO
Dra. Ana Maria de Castro	Aprovado
Dr. Marcus Tullius Scotti	Aprovado
Dra. Elisangela De Paula Silveira Lacerda	APROVADO

Em face do resultado obtido, a Banca Examinadora considerou o candidato habilitado, (**Habilitado ou não Habilitado**), cumprindo todos os requisitos para fins de obtenção do título de **DOUTOR EM MEDICINA TROPICAL E SAÚDE PÚBLICA**, na área de concentração em **PARASITOLOGIA**, pela Universidade Federal de Goiás. Cumpridas as formalidades de pauta, às 11 h 55min, a presidência da mesa encerrou esta sessão de defesa de tese e para constar eu, KARINY VIEIRA SOARES E SILVA, secretária do Programa de Pós-Graduação em Medicina Tropical e Saúde Pública lavrei a presente Ata que depois de lida e aprovada, será assinada pelos membros da Banca Examinadora e por mim em duas vias de igual teor.
A Banca Examinadora aprovou a seguinte alteração no título da Tese:

_____

_____

_____

Dra. Carolina Horta Andrade (FF/UFG) _____
Dra. Ana Maria de Castro (IPTSP/UFG) Ana Maria de Castro.
Dr. Milton Adriano Pelli de Oliveira (ISPTSP/UFG) _____
Dr. Marcus Tullius Scotti (UFPB) _____
Dra. Elisangela de Paula Silveira Lacerda (ICB/UFG) _____
Secretário da Pós-Graduação: _____

Dedico este trabalho aos meus pais, Cleber e Judite, que sempre me incentivaram a lutar pelos meus sonhos.

AGRADECIMENTOS

Primeiramente a Deus, pela oportunidade de estar aqui hoje, realizando mais esta etapa.

Aos meus pais, Cleber e Judite, pelo carinho e por me incentivarem, desde a infância, a buscar o conhecimento e a lutar pelos meus sonhos.

À minha esposa Wagna, pelo amor, carinho, companheirismo e compreensão.

À professora Dra. Carolina Horta Andrade, pelo exemplo de profissionalismo e competência, pela orientação, paciência e oportunidade de realizar este trabalho.

Ao Dr. Rodolpho de Campos Braga, pela contribuição, ensinamentos e disposição em ajudar.

Ao Dr. Eugene Muratov, pelas importantes contribuições no desenvolvimento de modelos de QSAR.

A todos os colegas do LabMol pelos momentos de descontração na hora do famoso cafezinho e pela constante troca de conhecimentos.

A todos os professores que participaram e contribuíram nessa jornada, iniciada há 22 anos, no jardim de infância.

À OpenEye Scientific Software e ChemAxon pela licença acadêmica de programas utilizados neste trabalho.

Ao CNPq, CAPES e FAPEG pelo apoio financeiro.

Ao Portal de Periódicos da CAPES

SUMÁRIO

TABELAS, FIGURAS E ANEXOS xi
SÍMBOLOS, SIGLAS E ABREVIATURAS xiv
RESUMO xviii
ABSTRACT xix
1. INTRODUÇÃO / REVISÃO DA LITERATURA 1
1.1 Tripanossomatídeos 1
1.1.1. Doença de Chagas 1
1.1.1.1. Distribuição e epidemiologia 2
1.1.1.2. Morfologia do parasito e ciclo biológico 4
1.1.1.3. Quimioterapia 7
1.1.2. Leishmanioses 8
1.1.2.1. Distribuição e epidemiologia 9
1.1.2.2. Formas clínicas das leishmanioses 11
1.1.2.3. Morfologia do parasito e ciclo biológico 13
1.1.2.4. Quimioterapia 18
1.2. Descoberta e desenvolvimento de novos fármacos 20
1.2.1. A indústria farmacêutica e as doenças tropicais negligenciadas 23
1.3. Planejamento de fármacos auxiliado por computador 24
1.3.1. Relações quantitativas entre estrutura e atividade (QSAR) 25
1.3.1.1. Princípios 26
1.3.1.2. Descritores moleculares 28
1.3.1.3. Métodos de aprendizado de máquina 29
1.3.1.4. Boas práticas de desenvolvimento e validação de modelos de QSAR 31
1.3.1.4.1. Preparo do conjunto de dados 32
1.4. Bases moleculares para a descoberta de novos fármacos contra as leishmanioses 33
1.4.1. Via glicolítica 34
1.4.1.1. Piruvato quinase (PK) 36
1.4.2. Biossíntese de esteróis de membrana 39
2. JUSTIFICATIVA 42
3. OBJETIVOS 44
3.1. Objetivos específicos 44

3.1.1. Identificação de novos compostos bioativos contra *Trypanosoma cruzi* 44
3.1.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp. 44
3.1.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp. 45
4. MÉTODOS 46
4.1. Identificação de novos compostos bioativos contra *T. cruzi* 46
4.1.1. Estratégias Computacionais 46
4.1.1.1. Conjuntos de dados 46
4.1.1.2. Preparo e padronização dos conjuntos de dados 46
4.1.1.3. Balanceamento dos conjuntos de dados 47
4.1.1.4. Geração e validação dos modelos de QSAR 48
4.1.1.5. Fingerprints moleculares 48
4.1.1.6. Método de aprendizado de máquina 48
4.1.1.7. Validação cruzada externa 5-fold 48
4.1.1.8. Definição do domínio de aplicabilidade (DA) 49
4.1.1.9. Avaliação da performance dos modelos de QSAR 49
4.1.1.10. Triagem virtual 50
4.1.2. Avaliação experimental 51
4.1.2.1. Cultura de células 51
4.1.2.2. Preparo dos compostos 52
4.1.2.3. Ensaios de atividade antiparasitária 52
4.1.3. Avaliação multi-paramétrica *in silico* dos compostos testados 53
4.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp. 53
4.2.1. Conjunto de dados 53
4.2.1.1. Preparo e padronização do conjunto de dados 54
4.2.1.2. Balanceamento do conjunto de dados 54
4.2.1.3. Preparo da estrutura tridimensional dos compostos 54

4.2.2. Modelo baseado em *docking* (SBDD) 55

4.2.2.1. Preparo da proteína 55

4.2.2.2. Docking 55

4.2.3. Modelos baseados na forma e volume moleculares (modelos *shape-based*) 56

4.2.4. Avaliação dos modelos SBDD e *shape-based* 57

4.2.5. Modelos de QSAR 59

4.2.5.1. Fingerprints moleculares 59

4.2.5.2. Método de aprendizado de máquina 60

4.2.5.3. Validação cruzada externa 5-fold 60

4.2.5.4. Definição do DA 60

4.2.5.5. Modelos consenso 60

4.2.5.6. Avaliação da robustez e performance dos modelos de QSAR 61

4.2.6. Triagem virtual 62

4.2.7. Avaliação experimental 62

4.2.7.1. Culturas de L. infantum, macrófagos peritoneais e células NCTC ... 63

4.2.7.2. Avaliação da atividade contra L. infantum e citotoxicidade contra células de mamíferos 63

4.2.7.2.1. Avaliação da atividade em promastigotas 63

4.2.7.2.2. Avaliação da atividade em amastigotas intracelulares 64

4.2.7.2.3. Avaliação da citotoxicidade 64

4.2.7.3. Análise estatística 64

4.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp. 65

4.3.1. Conjuntos de dados 65

4.3.1.1. Preparo e padronização dos conjuntos de dados 66

4.3.1.2. Balanceamento dos conjuntos de dados 66

4.3.1.3. Geração de decoys para o conjunto de dados de inibidores de CYP51 67

4.3.1.4. Preparo da estrutura tridimensional dos inibidores de CYP51 e decoys 67

4.3.2. Modelos baseados na forma e volume moleculares (*shape-based*) 68
4.3.3. Modelos de QSAR para predição da atividade contra amastigotas de *L. infantum* 69
4.3.3.1. Descritores moleculares 69
4.3.3.2. Método de aprendizado de máquina 69
4.3.3.3. Validação cruzada externa 5-fold 70
4.3.3.4. Definição do DA 70
4.3.3.5. Modelos consenso 70
4.3.3.6. Avaliação da robustez e performance dos modelos de QSAR 70
4.3.4. Triagem virtual 71
4.3.5. *Docking* dos *hits* virtuais nas enzimas PK e CYP51 72
4.3.5.1. Preparo das proteínas 72
4.3.5.2. Preparo dos ligantes 72
4.3.5.3. Docking 73
5. RESULTADOS 74
5.1. Identificação de novos compostos bioativos contra *Trypanosoma cruzi* utilizando triagem virtual baseada em QSAR 74
5.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp. 121
5.2.1. Validação dos modelos *shape-based* e do modelo baseado em *docking* 121
5.2.2. Modelos de QSAR 123
5.2.3. Triagem virtual 126
5.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp. 129
5.3.1. Validação dos modelos *shape-based* 130
5.3.2. Modelos de QSAR para predição da atividade contra amastigotas de *L. infantum* 132
5.3.3. Triagem virtual 135
6. DISCUSSÃO 139
7. CONCLUSÕES 143
REFERÊNCIAS 145

ANEXOS **163**
Anexo 1 - Artigos publicados através de colaborações durante o doutorado **163**
Anexo 2 - Artigo submetido e comprovante de submissão **164**

TABELAS, FIGURAS E ANEXOS

Figura 1 - Distribuição global da doença de Chagas, baseada em estimativas oficiais, no período de 2006-2015 (Adaptado de WHO, 2017). 3

Figura 2 - Forma epimastigota de *Trypanosoma cruzi* (Adaptado de BOGITSH; CARTER; OELTMANN, 2013)..... 4

Figura 3 - Forma tripomastigota de *Trypanosoma cruzi* (Adaptado de: BOGITSH; CARTER; OELTMANN, 2013; MACHADO; TANOWITZ, 2012). 5

Figura 4 - Forma amastigota de *Trypanosoma cruzi* (Adaptado de: MACHADO; TANOWITZ, 2012; TEIXEIRA et al., 2012). 5

Figura 5 - Ciclo biológico do *Trypanosoma cruzi* (Adaptado de CDC, 2017b). 6

Figura 6 - Estrutura do benznidazol, utilizado no tratamento da doença de Chagas..... 8

Figura 7 - Distribuição dos países com maior número de casos de leishmanioses em 2016 (Adaptado de WHO, 2017)..... 10

Figura 8 - Distribuição das leishmanioses por região no Brasil (Adaptado de MS, 2014a,b). 11

Figura 9 - Forma promastigota de *Leishmania* spp (Adaptado de BOGITSH; CARTER; OELTMANN, 2013; CDC, 2013). 14

Figura 10 - Forma amastigota de *Leishmania* spp (Adaptado de BOGITSH; CARTER; OELTMANN, 2013; MURRAY et al., 2005). 15

Figura 11 - Ciclo biológico de *Leishmania* spp (Adaptado de BOELAERT; SUNDAR, 2014; CDC, 2015). 16

Figura 12 - Etapas envolvidas na descoberta e desenvolvimento de novos fármacos (Adaptado de ELEBRING; GILL; PLOWRIGHT, 2012; NWAKA; RIDLEY, 2003). 22

Figura 13 - Representação do processo de construção de um modelo de QSAR (Adaptado de TROPSHA, 2010). 27

Figura 14 - Esquema geral do método *Random Forest* (Adaptado de NEVES, 2015). 31

Figura 15 - Representação das etapas para o preparo do conjunto de dados (FOURCHES; MURATOV; TROPSHA, 2015). 32

Figura 16 - Esquema geral da via glicolítica em tripanossomatídeos (Adaptado de VERLINDE et al., 2001). 35

Figura 17 - Estrutura da PK de *L. mexicana* (*Lm*PK) em complexo com ATP, oxalato e F-2,6-BP (Adaptado de MORGAN et al., 2010). 37

Figura 18 - Biossíntese de esteróis a partir do acetil-CoA até a formação de colesterol (vertebrados), ergosterol (fungos e protozoários) e fitoesterol (plantas) (Adaptado de CHOI; PODUST; ROUSH, 2014; HARGROVE et al., 2012). 40

Figura 19 - Representação do desempenho de diferentes classificadores (pontos A, B, C e D) e da curva ROC (Adaptado de BRAGA; ANDRADE, 2013). 57

Figura 20 - Melhor modelo *shape-based* (modelo V) selecionado após a validação com inibidores e não inibidores de PK.. 122

Figura 21 - Fragmentos mais frequentes encontrados no conjunto de inibidores da PK (compostos ativos), sua respectiva frequência e exemplos de MPP contendo esses fragmentos. 125

Figura 22 - Fragmentos mais frequentes encontrados no conjunto de não inibidores da PK (compostos inativos), sua respectiva frequência e exemplos de MPP contendo esses fragmentos.. 125

Figura 23 - Resultados da triagem virtual de potenciais inibidores da enzima PK de *Leishmania* spp. 126

Figura 24 - Melhor modelo *shape-based* selecionado após a validação com inibidores de CYP51 e *decoys*. 131

Figura 25 - Fragmentos mais frequentes encontrados no conjunto de compostos ativos contra *L. infantum*, sua respectiva frequência e exemplos de MPP contendo esses fragmentos. 133

Figura 26 - Fragmentos mais frequentes encontrados no conjunto de compostos inativos contra *L. infantum*, sua respectiva frequência e exemplos de MPP contendo esses fragmentos.. 134

Figura 27 - Resultados da triagem virtual *dual-target* de potenciais inibidores das enzimas PK e CYP51 e com potencial de serem ativos contra as formas amastigotas de *Leishmania infantum*. 135

Tabela 1 - Espécies de *Leishmania* patogênicas em humanos. 12

Tabela 2 - Fármacos utilizados na quimioterapia contra as leishmanioses e suas limitações. 18

Tabela 3 - Representação genérica de uma matriz de dados para um estudo de QSAR. 27

Tabela 4 - Validação dos modelos *shape-based* (I-V) e SBDD (*docking*) a partir do conjunto de dados de inibidores e não inibidores da enzima PK (PubChem AID: 1721), balanceados na proporção de 1:36. 121

Tabela 5 - Parâmetros estatísticos dos modelos de QSAR após validação cruzada externa *5-fold*. 123

Tabela 6 - Estruturas dos 14 potenciais inibidores da enzima PK de *Leishmania* selecionados após triagem virtual através de modelos *shape-based*, baseado em *docking* e QSAR. 127

Tabela 7 - Validação dos modelos *shape-based* a partir do conjunto de dados de inibidores da enzima CYP51 e *decoys*, balanceados na proporção de 1:36. 130

Tabela 8 - Parâmetros estatísticos dos modelos de QSAR após validação cruzada externa *5-fold*. 132

Tabela 9 - Estruturas dos 15 potenciais inibidores das enzimas PK e CYP51 e preditos como ativos contra amastigotas de *L. infantum* selecionados após triagem virtual através de modelos *shape-based* e QSAR. 136

SÍMBOLOS, SIGLAS E ABREVIATURAS

1D	unidimensional
2D	bidimensional
2-PGA	2-fosfoglicerato
3D	tridimensional
3-PGA	3-fosfoglicerato
4D	quatro dimensões
ADMET	absorção, distribuição, metabolismo, excreção e toxicidade
ADP	Adenosina difostato
ATP	Adenosina trifosfato
AUC	Área sob a curva ROC
BZN	benznidazol
CADD	Planejamento de fármacos auxiliado por computador (do inglês *Computer Assisted Drug Design*)
CART	*Classification and regression trees*
CCR	Taxa de classificação correta (do inglês *Correct Classification Rate*)
CC_{50}	Concentração que reduz o número de células de mamíferos pela metade
CYP51	Esterol 14 α-desmetilase
DA	Domínio de aplicabilidade
DHAP	Dihidroxiacetona fosfato
DNDi	*Drugs for Neglected Diseases initiative*
DUD	Diretório de *decoys* úteis, do inglês *directory of useful decoys*
DTN	Doenças tropicais negligenciadas
ECFP4	*standard-connectivity fingerprints*
EC_{50}	Concentração correspondente à redução de 50% da infecção por *T. cruzi*

EF	Fator de enriquecimento (do inglês *enrichment factor*)
FCFP4	*functional-class fingerprints*
FDA	*Food and Drug Administration*
G-3-P	gliceraldeído 3-fosfato
Gly-3-P	glicerol 3-fosfato
HCS	*High Content Screening*
HTS	Triagem biológica automatizada em grande escala (do inglês *High Throughput Screening*)
HIA	Absorção intestinal humana, (do inglês *human intestinal absorption*)
IOWH	*Institute for One World Health*
IUPAC	União Internacional de Química Pura e Aplicada, (do inglês *International Union of Pure and Applied Chemistry*)
k	Coeficiente kappa
kDNA	DNA do cinetoplasto
***k*NN**	*k-nearest neighbors*
LBDD	Planejamento baseado no ligante (do inglês *Ligand-Based Drug Design*)
LBVS	Triagem virtual baseada no ligante (do inglês *ligand-based virtual screening*)
LC	Leishmaniose Cutânea
LM	Leishmaniose Mucocutânea
LogP	coeficiente de partição octanol/água
LT	Leishmaniose Tegumentar
LV	Leishmaniose Visceral
MACCS	*Molecular Access System*
MLF	miltefosina
MPP	Mapas de probabilidade predita

MS Ministério da Saúde

NCE Nova Entidade Química (do inglês, *New Chemical Entity*)

NFX nifurtimox

NPV Valor preditivo negativo (do inglês *Negative Predictive Value*)

OECD Organização para a Cooperação e Desenvolvimento Econômico (do inglês *Organization for Economic Co-operation and Development*)

OMS Organização Mundial da Saúde

OOB *Out-of-bag*

P&D Pesquisa e Desenvolvimento

PCA Análise de componentes principais (do inglês *principal component analysis*)

PDB *Protein Data Bank*

PEP Fosfoenolpiruvato

PK Piruvato quinase

PKDL Leishmaniose Dérmica Pós-Calazar (do inglês *Post Kala-azar Dermal Leishmaniases*)

PPPs Parcerias público-privadas

PPV Valor preditivo positivo (do inglês *Positive Predictive Value*)

QSAR Relações quantitativas entre estrutura e atividade (do inglês *Quantitative Structure-Activity Relationships*)

RF Florestas aleatórias (do inglês *random forest*)

RMSD Desvio médio quadrático (do inglês *Root Mean Square Deviation*)

RNAi Interferência por RNA

ROC *Receiver operating characteristic*

ROCS *Rapid Overlay of Chemical Structures*

SAR Relação entre a estrutura e atividade (do inglês *Structure-Activity Relationships*)

SBDD Planejamento de fármacos baseado na estrutura (do inglês *Structure-*

Based Drug Design)

SBVS Triagem virtual baseada na estrutura (do inglês *structure-based virtual screening*)

SE Sensibilidade

SI Índice de seletividade

SFB Soro fetal bovino

SP Especificidade

SVM Máquina de suporte de vetores (do inglês *support vector machine*)

VS Triagem virtual (do inglês *Virtual Screening*)

1,3-PGA 1,3-bisfosfoglicerato

F-1,6-BP Frutose-1,6-bifosfato

F-2,6-PB Frutose-2,6-bifosfato

RESUMO

Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal

Tripanossomatídeos são parasitos causadores de diferentes doenças que afetam principalmente populações de baixa renda de países subdesenvolvidos e em desenvolvimento, sendo classificadas como doenças tropicais negligenciadas (DTNs). Os poucos fármacos disponíveis para o tratamento dessas parasitoses apresentam diversos problemas relacionados à baixa eficácia, resistência e toxicidade, tornando urgente a busca por novas opções terapêuticas. Nesse contexto, o objetivo deste trabalho foi identificar novos candidatos a fármacos contra tripanossomatídeos através da aplicação de estratégias integradas em Química Medicinal. Para atingir esse objetivo, foram desenvolvidos três estudos combinando diferentes abordagens do planejamento de fármacos auxiliado por computador (CADD) e validação experimental. No primeiro estudo, a partir de uma triagem virtual (VS) baseada em modelos de relações quantitativas entre estrutura e atividade (QSAR), foram identificados e selecionados 39 *hits* virtuais para avaliação experimental. Dentre estes, 13 compostos foram ativos e seletivos *in vitro* contra *T. cruzi* intracelular. Foi realizada uma análise multiparamétrica *in silico* para avaliar e comparar algumas propriedades farmacocinéticas (ADMET), físico-químicas, com a potência e seletividade, que nos permitiu priorizar os compostos mais promissores para futuros ensaios *in vivo*. No segundo estudo, foi realizada uma VS para identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp., que levou a identificação e seleção de 14 novos potenciais inibidores dessa enzima. Estes compostos foram enviados para avaliação experimental contra *L. infantum* amastigota. No terceiro estudo, foi realizada uma VS com dois alvos (*dual-target*), que permitiu a identificação e seleção de 15 potenciais inibidores de ambas as enzimas PK e 14 α–esterol desmetilase (CYP51) de *Leishmania* spp., também preditos, através de modelos de QSAR, como ativos contra forma amastigota de *L. infantum*. Portanto, este trabalho demonstrou o potencial da integração de métodos computacionais e experimentais para a identificação de compostos ativos contra tripanossomatídeos.

Palavras-chave: Triagem virtual; descoberta de fármacos; *Trypanosoma cruzi*; *Leishmania*; tratamento.

ABSTRACT

Identification of novel bioactive compounds against trypanosomatids through integrated strategies in Medicinal Chemistry

Trypanosomatids are parasites that cause different diseases affecting mainly low-income populations from countries in development or underdeveloped countries. These diseases are classified as neglected tropical diseases (NTDs). The few drugs available for the treatment of these parasitic diseases present several problems regarding low efficacy, resistance, and toxicity, making the need of new therapeutic options an urgent task. In this sense, the aim of this work was the identification of novel drug candidates against trypanosomatids through the application of integrated strategies in Medicinal Chemistry. To achieve this goal, three studies were performed combining different computer assisted drug design (CADD) approaches and experimental validation. In the first study, 39 virtual hits were identified by quantitative structure activity relationships (QSAR)-based virtual screening (VS) and selected for experimental evaluation. Among them, 13 compounds were active and selective *in vitro* against intracellular *T. cruzi*. Additionally, an *in silico* multi-parameter analysis for evaluation and comparison of some pharmacokinetic (ADMET) and physicochemical properties, with potency and selectivity, allowed the prioritization of the most promising compounds for further *in vivo* assays. In the second study, VS was performed for identification of novel potential inhibitors of pyruvate kinase (PK) from *Leishmania* spp., allowing the discovery of 14 novel potential inhibitors of this enzyme. These compounds were submitted to experimental evaluation against *L. infantum* amastigotes. In the third study, a dual-target VS was performed, allowing the identification of 15 potential inhibitors of both PK and sterol 14 α-demethylase (CYP51) of *Leishmania* spp., also predicted as active compounds against *L. infantum* amastigotes by QSAR models. Therefore, this work has demonstrated the potential of the integration of computational and experimental methods for the identification of active compounds against trypanosomatids.

Keywords: Virtual screening; drug discovery; *Trypanosoma cruzi*; *Leishmania*; treatment.

1. INTRODUÇÃO / REVISÃO DA LITERATURA

1.1 Tripanossomatídeos

Trypanosomatidae é uma família composta por protozoários pertencentes à ordem *Kinetoplastida*, que se caracterizam pela presença de flagelo e uma estrutura intracelular única chamada cinetoplasto, que consiste em um disco formado por círculos de DNA interligados (kDNA), localizados dentro de uma grande mitocôndria (GUNN; PITT, 2012). Os tripanossomatídeos são parasitos causadores de diferentes doenças que afetam principalmente populações de baixa renda de países subdesenvolvidos e em desenvolvimento. Dentre essas doenças, se destacam a doença de Chagas, as leishmanioses e a doença do sono (ROBERTS; JAVONY, 2009). As duas primeiras doenças, que são endêmicas no Brasil e outros países da América Latina, serão apresentadas a seguir.

1.1.1. Doença de Chagas

A doença de Chagas, ou tripanossomíase americana, é uma doença tropical negligenciada (DTN) causada pelo protozoário flagelado *Trypanosoma cruzi*, pertencente à ordem *Kinetoplastida*, família *Trypanosomatidae*, que apresenta um ciclo de vida heteroxeno que envolve um inseto vetor hematófago (triatomíneo) e um hospedeiro vertebrado (mamífero) (FRANCO-PAREDES, 2014). O protozoário na forma tripomastigota metacíclica, presente nas fezes de triatomíneos, adentra o hospedeiro vertebrado através da ferida deixada no local da picada durante o repasto sanguíneo (KIRCHHOFF, 2011). Outras formas de transmissão incluem a ingestão de alimento contaminado, a transmissão congênita, transfusão sanguínea e transplante de órgãos (BERN, 2015).

A infecção por *T. cruzi* se apresenta clinicamente em duas fases: aguda e crônica. A fase aguda acontece durante as primeiras semanas ou meses após a infecção sendo assintomática em aproximadamente 90-95% dos casos (FRANCO-PAREDES, 2014). Em pacientes sintomáticos, pode haver febre, fadiga, dores no corpo, dores de

cabeça e erupções cutâneas. Além disso, em alguns casos podem ser observados, no local da picada e da contaminação pelas fezes do triatomíneo, sinais como o chagoma (inchaço cutâneo) e o edema de pálpebra, classicamente conhecido como sinal de Romaña. Em casos raros, crianças podem apresentar miocardite ou meningoencefalite (BERN, 2015; CDC, 2017a). Na fase crônica, apesar de grande parte dos pacientes permanecerem na forma indeterminada, cerca de 10-30% desenvolve a forma sintomática, geralmente anos ou décadas após a infecção. A cardiomiopatia é a complicação mais frequente da fase crônica sintomática, resultando em cardiomegalia, arritmias, contrações ventriculares prematuras e bloqueio atrioventricular, podendo levar à morte súbita dos pacientes. A segunda complicação mais frequente está relacionada a disfunções do trato gastrointestinal levando aos quadros de megaesôfago e megacólon, os quais estão associados à destruição de neurônios nesses órgãos (BERN, 2015; KIRCHHOFF, 2011; WHO, 2018).

1.1.1.1. Distribuição e epidemiologia

Estima-se que aproximadamente 6-7 milhões de pessoas no mundo todo estejam infectadas por *T. cruzi*, principalmente nas áreas endêmicas localizadas em 21 países da América Latina (WHO, 2017). Estimativas recentes indicam a existência de aproximadamente 300.000 casos crônicos de doença de Chagas nos Estados Unidos, colocando a região sul do país na lista de áreas endêmicas. Essa mudança na distribuição da doença se deve principalmente ao crescente número de imigrantes originários de áreas endêmicas da América Latina. Os movimentos migratórios também podem explicar o aparecimento de casos na Europa (KIRCHHOFF, 2011). Apesar do papel predominante da imigração, ao longo dos anos, um pequeno número de casos autóctones da doença foi reportado nos Estados Unidos, visto que 11 espécies de triatomíneos são encontradas no país. Além disso, o ciclo de transmissão silvestre na região já é conhecido há pelo menos um século (KLOTZ et al., 2014; PARISE et al., 2016). A figura 1 mostra a distribuição global da doença de Chagas, baseada em estimativas oficiais, no período de 2006-2015.

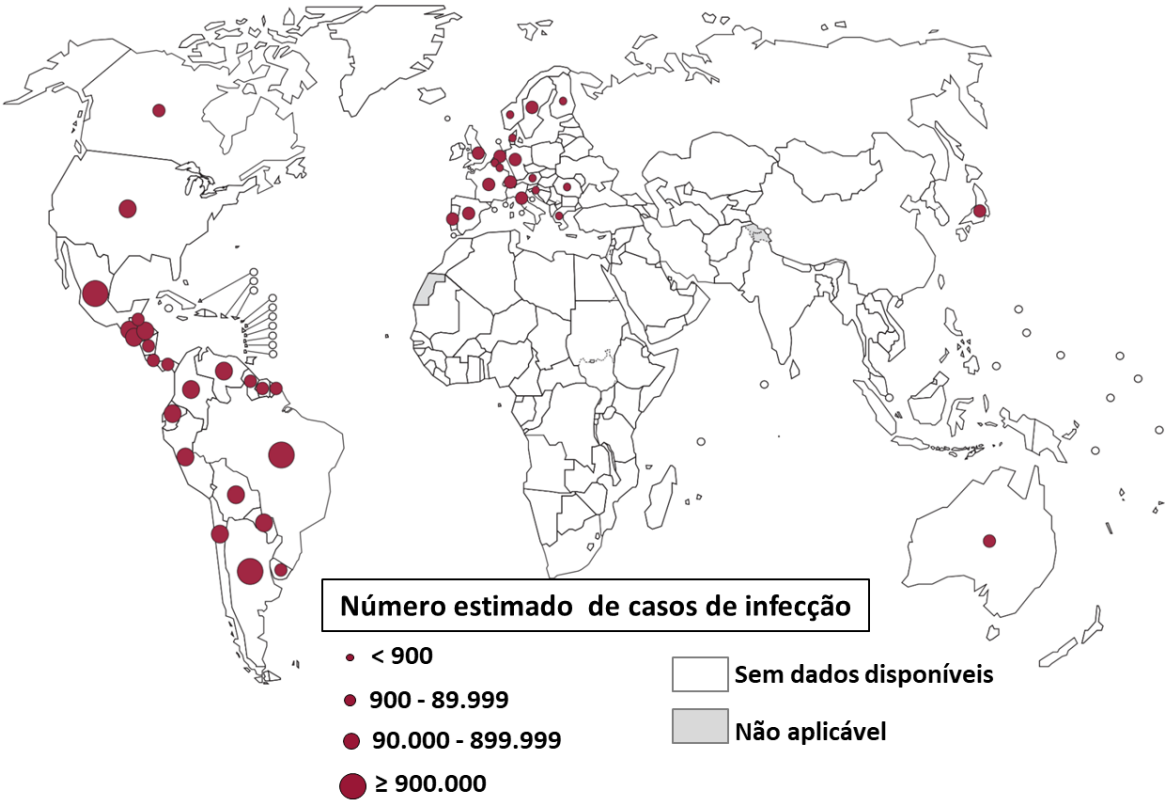


Figura 1 - Distribuição global da doença de Chagas, baseada em estimativas oficiais, no período de 2006-2015 (Adaptado de WHO, 2017).

Segundo dados do Ministério da Saúde (MS), entre o período de 2007 a 2016, foram registrados casos confirmados de doença de Chagas aguda na maioria dos estados brasileiros (1.989 casos), com uma média anual de 200 casos. A maior distribuição de casos, aproximadamente 95%, se concentrou na região Norte, sendo o estado do Pará responsável por 85% dos casos nessa região. Em relação às prováveis formas de transmissão, 69% dos casos foram por transmissão oral, 9% vetorial e 21% não foram identificadas (MS, 2017a). Além disso, segundo o MS, estima-se que 1,9 a 4,6 milhões de pessoas (1,0 a 2,4% da população) estão infectadas por *T. cruzi*, o que reflete na alta taxa de mortalidade pela doença no país (MS, 2017a).

No período de 2007 a 2016, foram registrados 27 casos de doença de Chagas aguda na região Centro-Oeste (1,3% dos casos no país). Destes 27 casos, 24 foram registrados no estado de Goiás, correspondendo a 89% dos casos na região (MS, 2017b).

1.1.1.2. Morfologia do parasito e ciclo biológico

O parasito passa por diferentes formas durante seu ciclo biológico: epimastigota, tripomastigota e amastigota (ORTEGA-BARRÍA, 2008).

A forma epimastigota é a forma replicativa encontrada no inseto vetor e não infectante para o hospedeiro vertebrado (Figura 2). Ela apresenta aspecto fusiforme, com o cinetoplasto localizado na sua porção anterior, próximo do núcleo. O flagelo forma uma pequena membrana ondulante e sua porção distal livre se projeta anteriormente (BOGITSH; CARTER; OELTMANN, 2013; CASTRO, 2011).

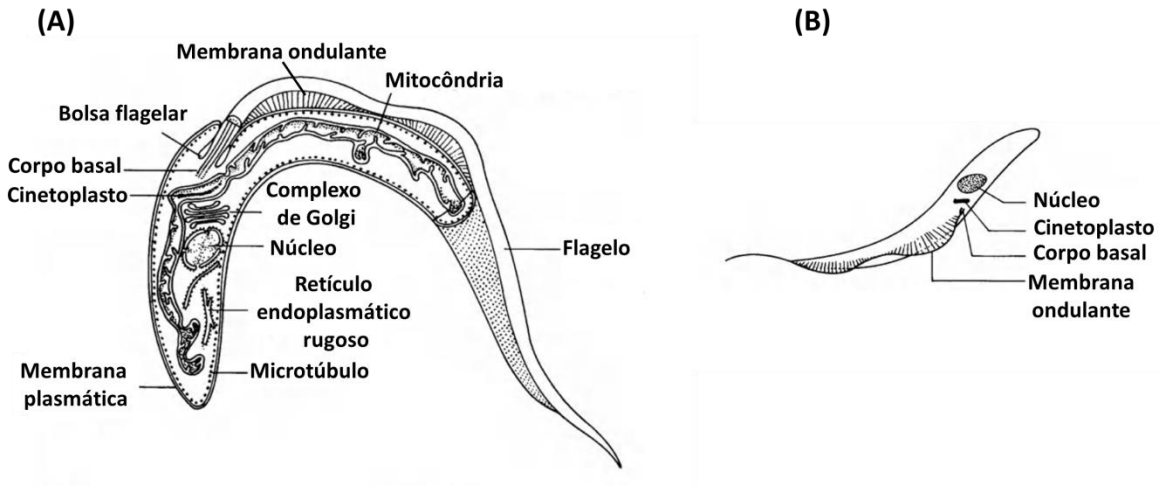


Figura 2 - Forma epimastigota de *Trypanosoma cruzi*. **(A)** Esquema geral da forma epimastigota demonstrando suas principais características. Destaca-se o aspecto fusiforme, o cinetoplasto localizado na porção anterior (próximo ao núcleo), a membrana ondulante e o flagelo. **(B)** Esquema simplificado da forma epimastigota, da forma como é normalmente vista em microscópio óptico. (Adaptado de BOGITSH; CARTER; OELTMANN, 2013).

A forma tripomastigota é encontrada no sangue circulante do hospedeiro vertebrado e apresenta formato alongado (Figura 3). O flagelo livre é moderadamente longo, enquanto a membrana ondulante é curta. O cinetoplasto está localizado na porção terminal ou subterminal (ROBERTS; JAVONY, 2009). A forma amastigota é a forma replicativa intracelular encontrada nos tecidos do hospedeiro vertebrado (Figura 4). Ela apresenta formato esférico ou oval e um flagelo livre muito curto, localizado dentro de uma bolsa (ORTEGA-BARRÍA, 2008).

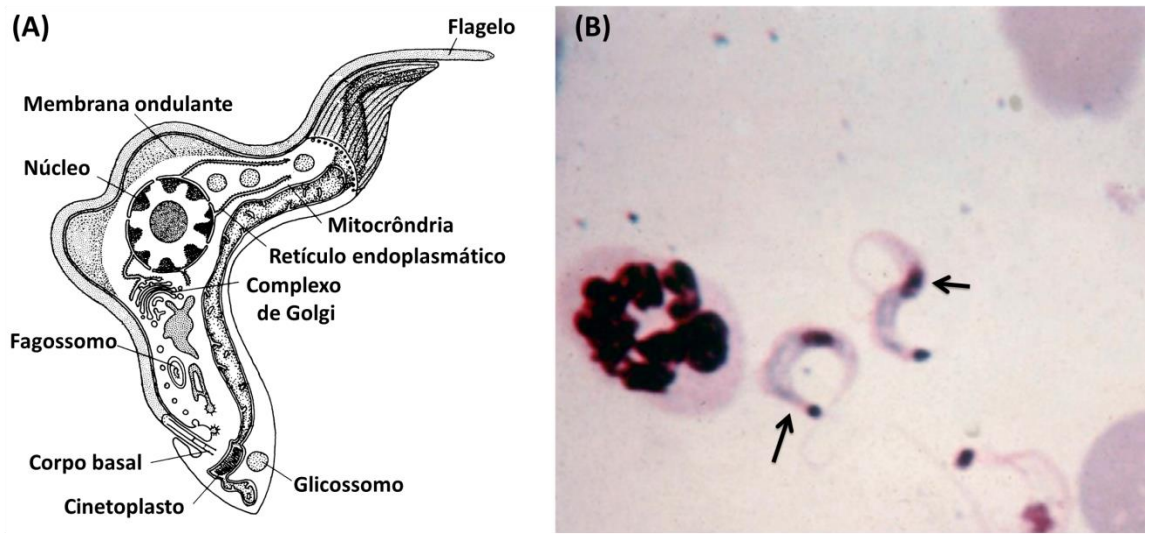


Figura 3 - Forma tripomastigota de *Trypanosoma cruzi*. **(A)** Esquema geral da forma tripomastigota demonstrando suas principais características. Destaca-se o formato alongado, o cinetoplasto localizado na porção terminal, a membrana ondulante e o flagelo. **(B)** Tripomastigotas (setas) observados ao microscópio óptico a partir de esfregaço de sangue periférico. (Adaptado de: BOGITSH; CARTER; OELTMANN, 2013; MACHADO; TANOWITZ, 2012).

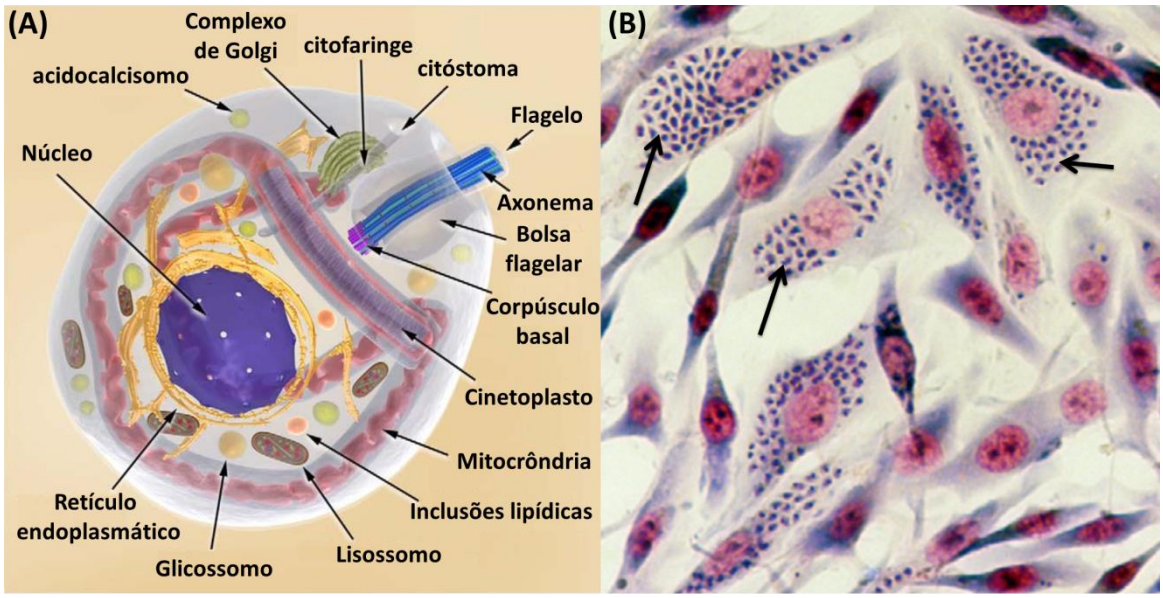


Figura 4 - Forma amastigota de *Trypanosoma cruzi*. **(A)** Esquema geral da forma amastigota demonstrando suas principais características. Destaca-se o formato esférico ou oval, o cinetoplasto e a presença de um flagelo curto, localizado na bolsa flagelar. **(B)** Amastigotas intracelulares (setas) em uma cultura de mioblastos. (Adaptado de: MACHADO; TANOWITZ, 2012; TEIXEIRA et al., 2012).

O ciclo biológico de *T. cruzi* é complexo e envolve insetos vetores (triatomíneos) e hospedeiros mamíferos (Figura 5).

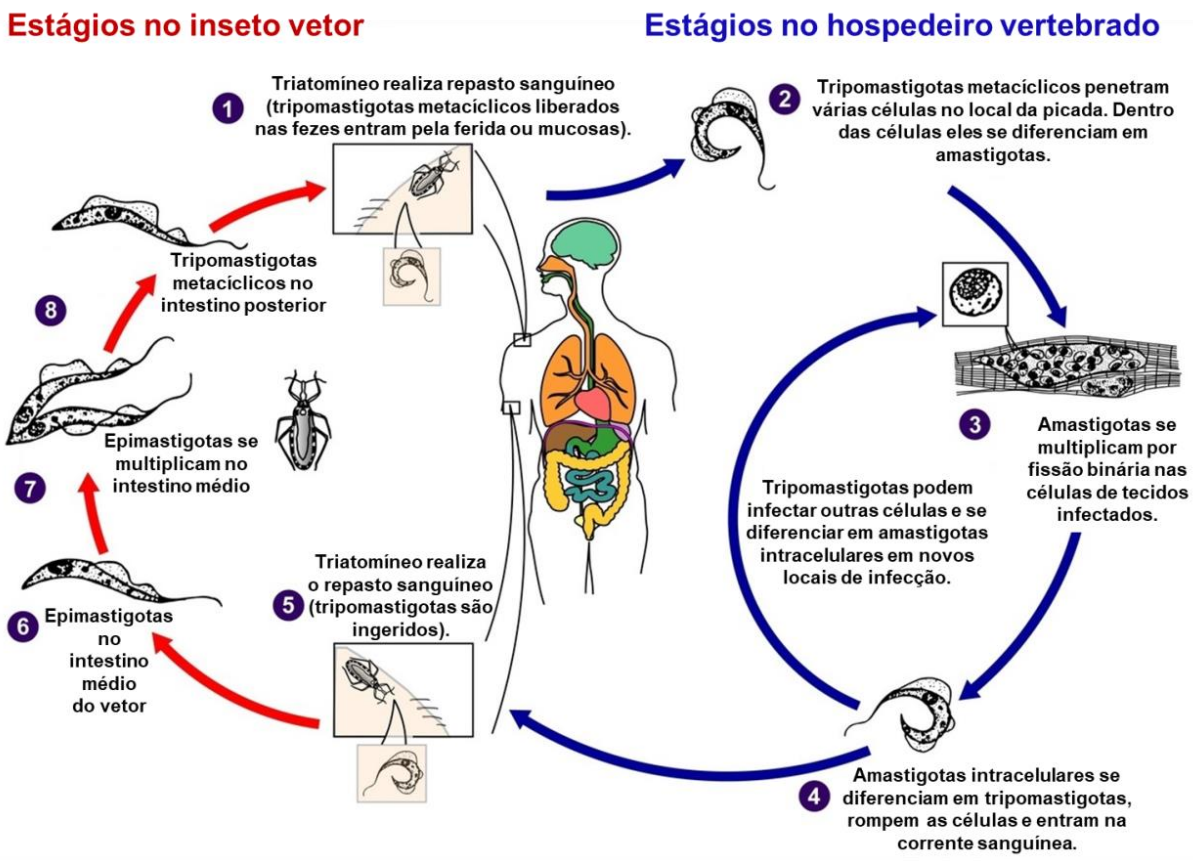


Figura 5 - Ciclo biológico do *Trypanosoma cruzi*. **(1)** O inseto triatomíneo realiza repasto sanguíneo e libera tripomastigotas nas fezes, os quais adentram o hospedeiro vertebrado através de abertura no local da picada ou mucosas intactas. **(2)** dentro do hospedeiro vertebrado, os tripomastigotas invadem células e se diferenciam em amastigotas intracelulares. **(3)** Os amastigotas se multiplicam dentro das células. **(4)** Os amastigotas intracelulares se diferenciam em tripomastigotas, rompem as células e são liberados na corrente sanguínea, podendo adentrar novas células, se diferenciar novamente em amastigotas, e se replicarem. **(5)** O triatomíneo, ao realizar repasto sanguíneo, ingere tripomastigotas circulantes. **(6)** Os tripomastigotas ingeridos se diferenciam em epimastigotas no intestino médio do vetor. **(7)** Os epimastigotas se multiplicam no intestino médio do vetor. **(8)** Os epimastigotas se diferenciam em tripomastigotas metacíclicos no intestino posterior (Adaptado de CDC, 2017b).

O vetor triatomíneo se infecta ao ingerir sangue de mamíferos que apresentam tripomastigotas circulantes, que são formas infectantes e não replicativas. Ao chegarem ao intestino médio do vetor, os parasitos se diferenciam em epimastigotas, os quais se multiplicam no ambiente extracelular. Após migrarem para o intestino posterior, os epimastigotas se diferenciam em tripomastigotas metacíclicos (formas não replicativas e infectantes), os quais são liberados junto às fezes do vetor durante o próximo repasto sanguíneo (KIRCHHOFF, 2010). A transmissão para um próximo hospedeiro vertebrado (mamífero) ocorre quando a pele (com rupturas), as membranas mucosas ou a conjuntiva entram em contato com as fezes contendo as formas metacíclicas infectantes. Dentro do hospedeiro vertebrado, os parasitos adentram uma variedade de tipos de células e, após se diferenciarem em amastigotas, se multiplicam no ambiente intracelular (KIRCHHOFF, 2011; ROBERTS; JAVONY, 2009). Após se multiplicarem dentro das células hospedeiras, os amastigotas se diferenciam em tripomastigotas e, em seguida, rompem as células. Os tripomastigotas liberados invadem tecidos adjacentes e se espalham hematogenicamente para diversos locais, nos quais eles iniciam novos ciclos assíncronos de multiplicação intracelular, mantendo assim a parasitemia em níveis suficientes para infectar novos insetos vetores (FRANCO-PAREDES, 2014).

1.1.1.3. Quimioterapia

Apenas um fármaco está disponível para o tratamento da doença de Chagas na maioria dos países: o derivado nitro-heterocíclico benznidazol (BZN), descoberto no final dos anos 60 (PAUCAR; MORENO-VIGURI; PÉREZ-SILANES, 2016). Existe uma necessidade urgente de novas alternativas terapêuticas para a doença de Chagas, uma vez que o BZN apresenta eficácia limitada no tratamento da fase crônica da doença, longo período de tratamento (60 dias) e efeitos adversos importantes como erupções cutâneas, náuseas, nefrotoxicidade e hepatotoxicidade (CLAYTON, 2010).

O BZN é um derivado 2-nitroimidazólico usado como primeira escolha no tratamento da doença de Chagas (Figura 6). Esse composto age como um pró-fármaco, o qual é ativado por nitroredutases, especialmente a nitroredutase do tipo I (PAUCAR; MORENO-VIGURI; PÉREZ-SILANES, 2016).

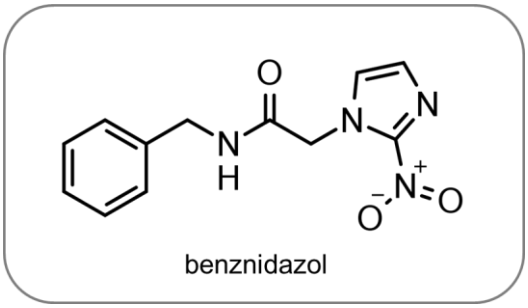


Figura 6 - Estrutura do benznidazol, utilizado no tratamento da doença de Chagas.

A cura clínica por BZN pode ser alcançada em 60-90% dos casos agudos da doença e em 90% das crianças infectadas por transmissão congênita, se tratadas em seu primeiro ano de vida (GASPAR et al., 2015). Entretanto, a eficácia do BZN na fase crônica é limitada, com taxas de cura de apenas 15-35%. A baixa solubilidade do fármaco associada ao tratamento com altas doses, por longos períodos de tempo, pode levar a efeitos adversos como hipersensibilidade cutânea, supressão da medula óssea, hepatite tóxica, intolerância digestiva e neuropatia periférica (BERMUDEZ et al., 2016). Novas formulações pediátricas de BZN foram desenvolvidas nos últimos anos (GASPAR et al., 2015). Em 2017, o FDA (do inglês *Food and Drug Administration*) aprovou o BZN para uso pediátrico em pacientes de 2 a 12 anos, sendo o primeiro tratamento para doença de Chagas a ser aprovado nos Estados Unidos. Entretanto, o processo foi realizado mediante o programa de "aceleração" de aprovações, o qual é usado em circunstâncias especiais como, por exemplo, condições graves sem opções terapêuticas (FDA, 2017). Portanto, ainda há uma grande necessidade de novos fármacos mais seguros e que sejam eficientes na fase crônica da doença de Chagas.

1.1.2. Leishmanioses

As leishmanioses compreendem um conjunto de doenças causadas por protozoários intracelulares do gênero *Leishmania*, pertencentes à ordem *Kinetoplastida*, família *Trypanosomatidae*. *Leishmania* é um protozoário heteroxeno, encontrado nas formas promastigota e amastigota e que, essencialmente, requer dois diferentes hospedeiros para completar o seu ciclo biológico: (1) um inseto vetor (hospedeiro invertebrado), representado pelas fêmeas da ordem Diptera, família *Psychodidae*, subfamília *Phlebotominae*, gêneros *Lutzomya*, *Psychodopygus* e *Phlebotomus* e (2)

como hospedeiro vertebrado, humanos, cães e uma ampla variedade de mamíferos silvestres (AKHOUNDI et al., 2016; HERWALDT, 1999).

Essas doenças afetam, principalmente, indivíduos de baixa condição socioeconômica e constituem um grande problema de saúde pública, uma vez que interferem na produtividade e qualidade de vida das pessoas acometidas. Além disso, em casos mais graves, como a leishmaniose visceral, pode levar a morte (ALVAR; YACTAYO; BERN, 2006). Os fatores socioeconômicos e a dificuldade de acesso ao tratamento e/ou descontinuidade do mesmo, somados as limitadas opções terapêuticas, em casos de resistência aos fármacos disponíveis, permitem classificar as leishmanioses como doenças tropicais negligenciadas (DTN) (OKWOR; UZONNA, 2016). Além disso, diversos fatores de risco são determinantes para a expansão dessas doenças, como: desnutrição, mudanças climáticas, movimentos migratórios, conflitos e condições de imunossupressão, como co-infecção com HIV (HOUWELING et al., 2016). Neste contexto, diversos problemas podem dificultar o controle das leishmanioses: o abandono do tratamento devido a limitações geográficas e financeiras, os custos elevados para o deslocamento aos locais de tratamento, a falta de opções terapêuticas por via oral, a toxicidade dos fármacos disponíveis e a difícil adesão dos pacientes ao esquema terapêutico (OKWOR; UZONNA, 2016).

1.1.2.1. Distribuição e epidemiologia

Segundo dados da Organização Mundial da Saúde (OMS), as leishmanioses são prevalentes em 98 países dos cinco continentes. Estimativas sugerem que 1,3 milhões de casos ocorram anualmente, dentre eles, aproximadamente 300.000 estão relacionados à forma visceral da doença e 1 milhão às formas cutânea e subcutânea. Estima-se que 20.000 a 50.000 mortes por leishmaniose visceral ocorram anualmente (WHO, 2015). Além disso, cerca de 350 milhões de pessoas vivem em áreas de risco para transmissão de leishmanioses (WHO, 2010).

Cerca de 90% dos casos de leishmaniose visceral, no mundo, se concentram em apenas sete países (Bangladesh, Brasil, Etiópia, Índia, Nepal, Sudão do Sul e Sudão). A maior parte dos casos de leishmaniose cutânea ocorre em dez países (Afeganistão, Argélia, Brasil, Colômbia, Irã, Paquistão, Peru, Arábia Saudita, Síria e Tunísia), enquanto a leishmaniose mucocutânea se concentra principalmente no Brasil, Peru e Bolívia (WHO, 2015). Entretanto, a notificação de casos é mandatória em apenas 34%

das áreas endêmicas (WHO, 2015). Mesmo nessas áreas, há problemas de subnotificação de casos, sugerindo que o número seja ainda maior que o reportado (SINGH et al., 2010). Nota-se que o Brasil é um país endêmico para todas as formas clínicas da doença.

A distribuição de casos de leishmanioses tem se expandido rapidamente nas últimas décadas. Dados na literatura indicam o recente aparecimento de casos em regiões não endêmicas, representados por indivíduos provenientes de regiões endêmicas. Tal fato pode ser associado ao crescimento de atividades de turismo, operações militares e ao aumento do fluxo migratório (PAVLI; MALTEZOU, 2010). A figura 7 mostra os 25 países com maior número de casos reportados à OMS em 2016.

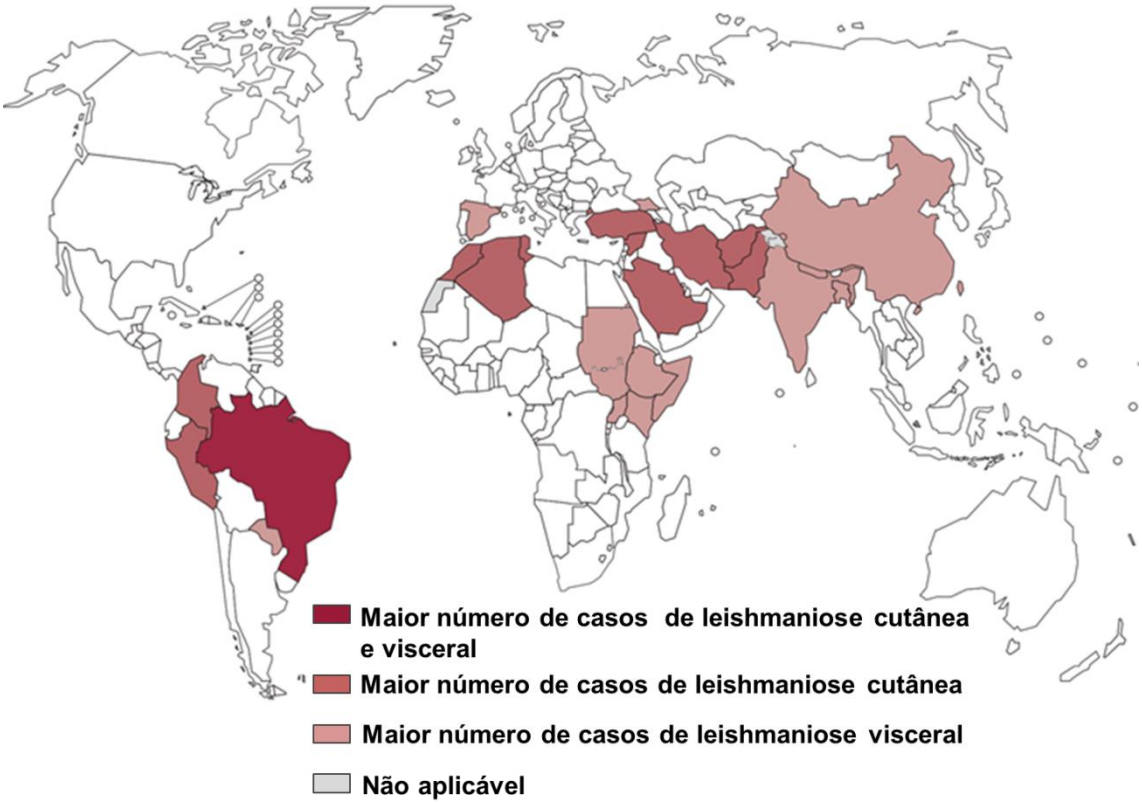


Figura 7- Distribuição dos países com maior número de casos de leishmanioses em 2016 (Adaptado de WHO, 2017).

Segundo dados do Ministério da Saúde, em 2014, no Brasil, foram reportados 20.296 casos de leishmaniose tegumentar (cutânea/mucocutânea). A região Norte apresentou o maior número de casos, com 10.387 (51,2 %), seguida da região Nordeste com 4.969 casos (24,5%), Centro-Oeste com 3.038 casos (15%), Sudeste com 1.460 casos (7,2%) e Sul com 373 casos (1,8%) (Figura 8A) (MS, 2014a). As regiões Norte e

Nordeste concentraram aproximadamente 75% dos casos de leishmaniose tegumentar reportados no país em 2014 (MS, 2014a, 2014b).

Em relação à leishmaniose visceral, em 2014, foram reportados 3.453 casos. A região Nordeste ocupou a primeira posição em número de casos, com 2.022 (58,6%), seguida pela região Sudeste com 456 casos (13,2%), Norte com 404 casos (11,7%), Centro-Oeste com 184 casos (5,3%) e Sul, com apenas quatro casos (0,1%) (Figura 8B) (MS, 2014c).

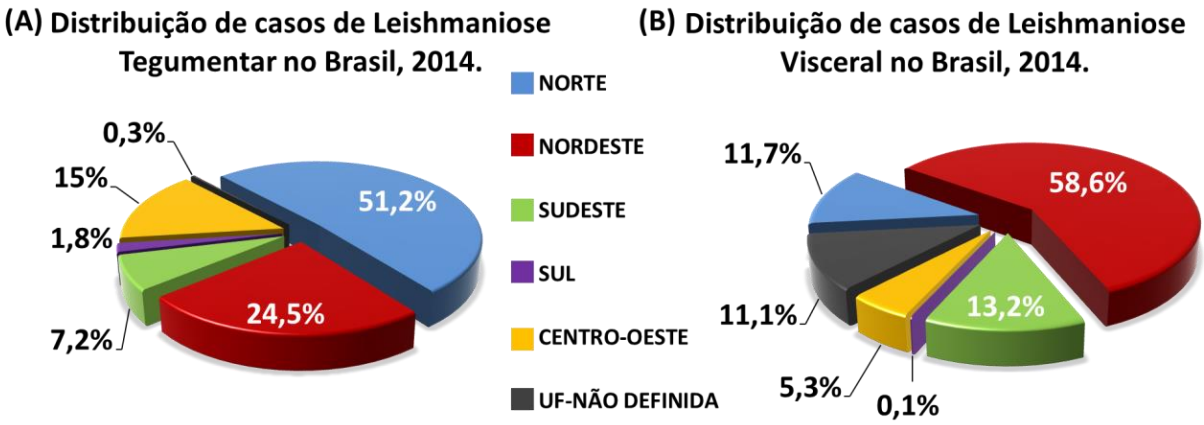


Figura 8 - Distribuição das leishmanioses por região no Brasil. **(A)** Distribuição de casos reportados de Leishmaniose Tegumentar, no Brasil, em 2014; **(B)** Distribuição de casos de reportados de Leishmaniose Visceral, no Brasil, em 2014. (Adaptado de MS, 2014a,b).

1.1.2.2. Formas clínicas das leishmanioses

As leishmanioses são caracterizadas pela diversidade de manifestações clínicas, cuja ocorrência depende tanto da espécie do parasito envolvida, quanto da resposta imune do hospedeiro (KAYE; SCOTT, 2011). Esses fatores determinam se o parasito irá se estabelecer na pele, levando à forma cutânea da doença, ou no baço, medula óssea e fígado, levando à forma visceral da doença (BOELAERT; SUNDAR, 2014).

Pelo menos 20 espécies de *Leishmania* são patogênicas em seres humanos. As três principais formas clínicas encontradas são: visceral (LV), cutânea (LC) e mucocutânea (LM), sendo as duas últimas por vezes agrupadas e denominadas leishmaniose tegumentar (LT) (BOELAERT; SUNDAR, 2014). A tabela 1 demonstra as principais espécies encontradas em seres humanos, sua localização geográfica e principal tropismo de cada uma.

Tabela 1 - Espécies de *Leishmania* patogênicas em humanos.

		Subgênero			
		***L.* (*Leishmania*)**	***L.* (*Leishmania*)**	***L.* (*Viannia*)**	***L.* (*Viannia*)**
Localização Geográfica	**Velho Mundo***	*L. donovani* *L. infantum*	*L. major* *L.tropica* *L. killicki*[a] *L. aethiopica* *L. infantum*		
	Novo Mundo**	*L. infantum*	*L. infantum* *L. mexicana* *L. pifanoi*[a] *L. venezuelensis* *L. garnhami*[a] *L. amazonensis*	*L. braziliensis* *L. guyanensis* *L. panamensis* *L. shawi* *L. naïffi* *L. lainsoni* *L. lindenbergi* *L. peruviana* *L. colombiensis*[b]	*L. braziliensis* *L. panamensis* *L. guyanensis*
Principal tropismo		Viscerotrópica	Dermotrópica	Dermotrópica	Mucotrópica

* Europa, África e Ásia; ** Américas; [a] Condição de espécie sob discussão; [b] Posição taxonômica sob discussão. (Adaptado de WHO, 2010).

A LV, também conhecida por calazar, é uma doença crônica e letal causada por parasitos do complexo *L. donovani-L. infantum* (Tabela 1). A maioria das infecções é assintomática. Entretanto, nutrição inadequada e supressão da resposta imune, especialmente em infecções por HIV, levam ao desenvolvimento de LV (WHO, 2010). Se não tratada, a LV pode ser fatal, sendo a morte associada à hemorragia, anemia grave ou infecção secundária (PAVLI; MALTEZOU, 2010). Indivíduos que se recuperam da LV podem desenvolver manifestações cutâneas crônicas, conhecidas como leishmaniose dérmica pós-calazar (PKDL, do inglês Post Kala-azar Dermal Leishmaniasis), caracterizadas por erupções maculo-papulares e nodulares que se iniciam geralmente próximas à boca e depois se espalham (MURRAY et al., 2005; WHO, 2016).

A LC é caracterizada por lesões que são iniciadas por pápula no local em que ocorre a picada do flebotomíneo, que gradualmente se transforma em nódulo, o qual permanece avermelhado e indolor. Lentamente, depois de 1 a 3 meses, a parte central

sofre necrose e a crosta formada se solta deixando uma úlcera de bordas elevadas (BOELAERT; SUNDAR, 2014; DAVID; CRAFT, 2012). Quando as úlceras são curadas, cicatrizes permanentes são deixadas, podendo levar a consequências sociais marcantes dependendo do local e extensão das lesões (WHO, 2016).

Apesar de esforços terem sido feitos para se classificar as diversas manifestações clínicas de LC, a sobreposição nas apresentações da doença, entre as espécies, é muito grande (DAVID; CRAFT, 2012). Dentre as variações clínicas da LC, destacam-se: a leishmaniose cutânea difusa, a leishmaniose cutânea disseminada e a leishmaniose recidiva cútis (HERWALDT, 1999; MURRAY et al., 2005; WHO, 2010).

Em uma população de indivíduos que apresentam LC, cerca de 1-10% dos pacientes desenvolve leishmaniose mucocutânea (LM). A progressão da doença em mucosas é dependente de uma combinação entre a imunidade celular do hospedeiro e a virulência do parasito (DAVID; CRAFT, 2012).

A LM é comumente encontrada no Novo Mundo, e está relacionada, principalmente, aos parasitos do subgênero *Viannia*: *L. (V) braziliensis*, *L. (V.) amazonensis*, *L. (V) panamensis* e *L. (V) guyanensis* (DAVID; CRAFT, 2012; WHO, 2010). No Velho Mundo, as lesões mucosas ocorrem principalmente em pacientes com PKDL e co-infecção por HIV. Esses indivíduos podem desenvolver lesões na mucosa bucal, nasal e genital (WHO, 2010). A LM pode ocorrer simultaneamente com a LC. Entretanto, ela se desenvolve principalmente meses ou anos após a cura de uma lesão cutânea. No seu estágio mais avançado, a doença leva à necrose de tecidos e mutilações que podem desfigurar o indivíduo, levando a consequências sociais e psicológicas marcantes (BOELAERT; SUNDAR, 2014).

1.1.2.3. Morfologia do parasito e ciclo biológico

Durante o ciclo biológico, os parasitos do gênero *Leishmania* assumem duas formas distintas: promastigota e amastigota (GUNN; PITT, 2012).

A forma promastigota (Figura 9), encontrada no hospedeiro invertebrado, apresenta formato alongado e um longo flagelo livre posicionado na porção anterior do parasito, a partir do axonema, o qual exerce tanto a função de locomoção quanto de aderência à parede do intestino do inseto vetor (BOGITSH; CARTER; OELTMANN, 2013). O núcleo da célula, arredondado ou oval, é único e está localizado no centro ou próximo a ele. O cinetoplasto está localizado na porção anterior do organismo. Além

disso, partindo do cinetoplasto, apresenta duas projeções da mitocôndria: uma na porção posterior da célula, por vezes responsável por alongar o comprimento da célula; e outra menor, na porção anterior (BOGITSH; CARTER; OELTMANN, 2013; MANTOOTH; ZEIBIG, 2013).

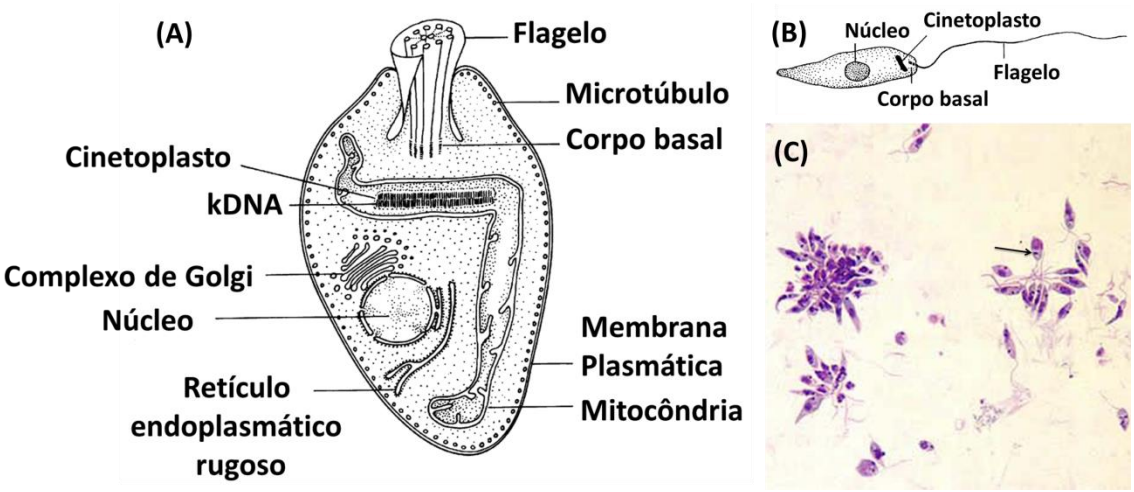


Figura 9 - Forma promastigota de *Leishmania* spp. **(A)** Esquema geral da forma promastigota demonstrando as principais organelas e suas características. Destaque para o cinetoplasto, situado dentro da mitocôndria, na porção anterior do parasito, contendo DNA (kDNA). A mitocôndria normalmente se encontra projetada nas porções anterior e posterior. **(B)** Esquema simplificado da forma promastigota, da forma como é normalmente vista em microscópio óptico. Destaque para o flagelo, na porção anterior do parasito, e para o cinetoplasto, posicionado entre a porção anterior e o núcleo. **(C)** Parasitos na forma promastigota observados em microscópio óptico. Observa-se o formato alongado dos parasitos e a presença do cinetoplasto (seta), posicionado na porção anterior, entre o núcleo e o flagelo. (Adaptado de BOGITSH; CARTER; OELTMANN, 2013; CDC, 2013).

A forma amastigota (Figura 10), que se desenvolve dentro de células do sistema mononuclear fagocítico do hospedeiro vertebrado, apresenta formato arredondado ou ovóide. Essa forma é caracterizada por um núcleo simples, cinetoplasto em forma de haste e um flagelo muito curto, que não se projeta para a superfície da célula, ficando dentro da bolsa flagelar. Pelo fato de ficar restrito à bolsa flagelar, o flagelo só pode ser observado por microscopia eletrônica (BOGITSH; CARTER; OELTMANN, 2013; VANNIER-SANTOS; MARTINY; DE SOUZA, 2002). Apesar de apresentar as mesmas organelas que a forma promastigota, algumas apresentam redução no tamanho

(e.g. mitocôndria) e provavelmente na função (BOGITSH; CARTER; OELTMANN, 2013).

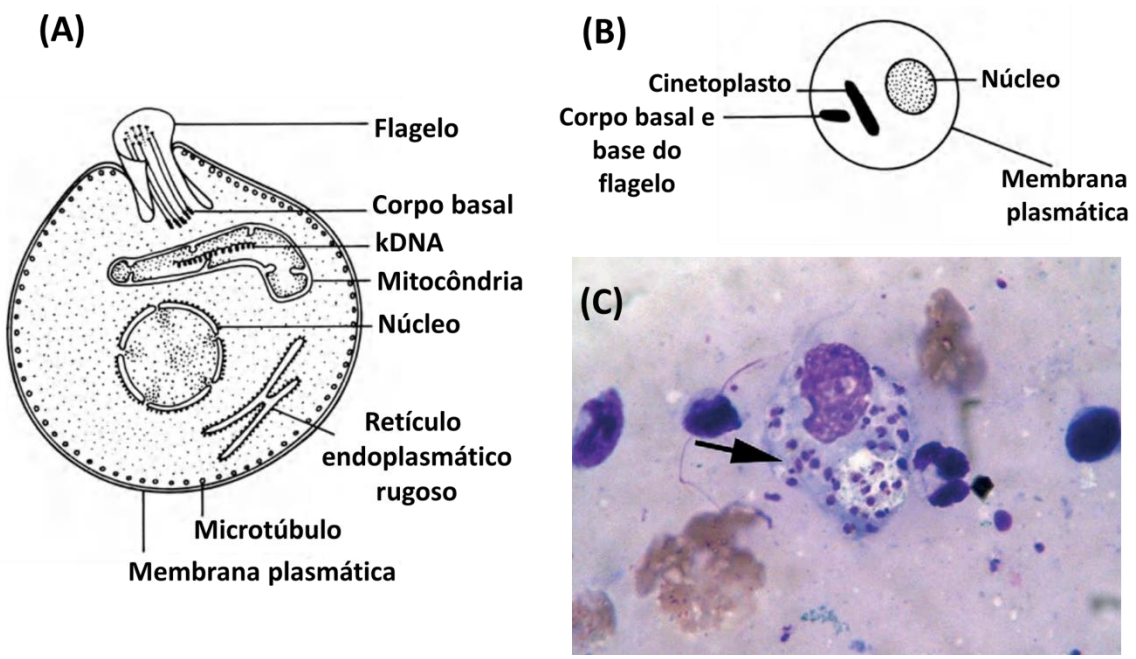


Figura 10 - Forma amastigota de *Leishmania* spp. **(A)** Esquema geral da forma amastigota e suas principais características. Observa-se que a mitocôndria apresenta tamanho reduzido em comparação à forma promastigota. O cinetoplasto, situado dentro da mitocôndria e contendo o kDNA, se apresenta em forma de haste. O flagelo fica recolhido na bolsa flagelar e não se projeta para fora da célula. **(B)** Esquema simplificado da forma amastigota, em sua aparência característica quando observada em microscópio óptico. **(C)** Esfregaço obtido a partir de raspado de lesão ocasionada por infecção por *L. panamensis*. A seta indica a presença de amastigotas localizadas no interior de um macrófago. (Adaptado de BOGITSH; CARTER; OELTMANN, 2013; MURRAY et al., 2005).

Os parasitos do gênero *Leishmania* apresentam um ciclo de vida dimórfico (alternam entre as formas promastigota e amastigota) e são heteroxenos, ou seja, requerem dois hospedeiros para que seu ciclo biológico seja completo: um hospedeiro invertebrado (flebotomíneo) e um hospedeiro vertebrado (mamífero) (AKHOUNDI et al., 2016). O ciclo biológico de *Leishmania* spp. está representado na figura 11.

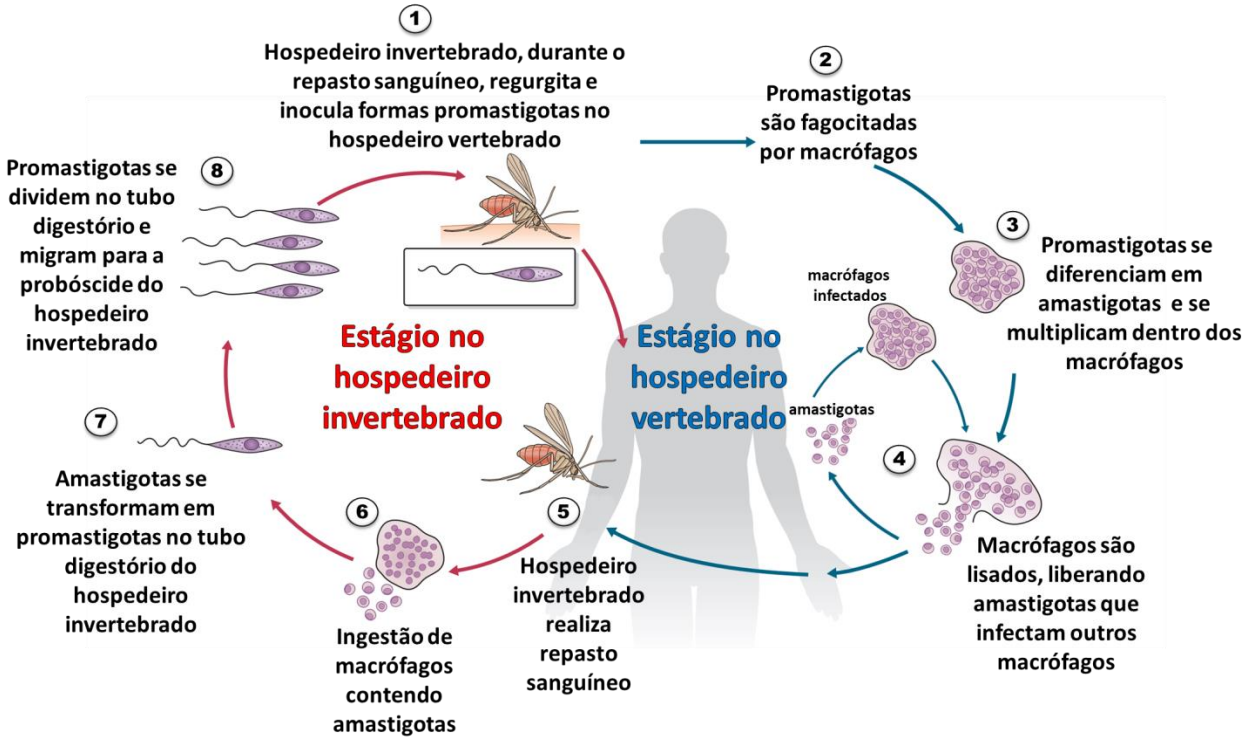


Figura 11 - Ciclo biológico de *Leishmania* spp. Setas azuis indicam os estágios do ciclo biológico que ocorrem no hospedeiro vertebrado. Setas vermelhas indicam os estágios do ciclo biológico que ocorrem no hospedeiro invertebrado. **(1)** Durante o repasto sanguíneo, fêmeas de flebotomíneos inoculam promastigotas metacíclicas no hospedeiro vertebrado. **(2)** Após serem inoculadas, as promastigotas metacíclicas são fagocitadas por macrófagos e outros tipos de células mononucleares fagocitárias. **(3)** Dentro do vacúolo parasitóforo, as promastigotas metacíclicas se diferenciam em amastigotas, as quais se multiplicam por divisão binária. **(4)** Os macrófagos são lisados e as amastigotas são liberadas para infectar outras células mononucleares fagocitárias. **(5 e 6)** Ao realizar o repasto sanguíneo, o hospedeiro invertebrado (fêmeas de flebotomíneos) ingerem macrófagos contendo amastigotas, os quais se rompem dentro do trato digestivo do inseto liberando as amastigotas. **(7)** As formas amastigotas se diferenciam em promastigotas procíclicas no tubo digestivo do inseto vetor. **(8)** As promastigotas procíclicas passam por uma série de alterações morfológicas e bioquímicas até se tornarem promastigotas metacíclicas, as quais migram para a probóscide do hospedeiro invertebrado. (Adaptado de BOELAERT; SUNDAR, 2014; CDC, 2015).

Ao picar um hospedeiro vertebrado infectado, as fêmeas de flebotomíneos ingerem macrófagos contendo amastigotas. Após adentrarem o trato digestivo, os

macrófagos são rompidos, liberando as formas amastigotas (intestino posterior para protozoários do subgênero *Viannia*; intestino médio para protozoários do subgênero *Leishmania*) (GUNN; PITT, 2012; MICHALIK, 2005). No intestino médio do vetor, as amastigotas se diferenciam em promastigotas procíclicas (forma reprodutiva e não infectante) em resposta à redução de temperatura e aumento do pH. Essas formas apresentam um flagelo relativamente pequeno, não são muito móveis e se multiplicam por divisão binária. Além disso, as formas procíclicas podem se aderir à parede do intestino médio através do flagelo (GUNN; PITT, 2012).

As promastigotas passam por uma série de alterações morfológicas e bioquímicas, à medida que se movem do intestino médio para a região da válvula estomodeal. As promastigotas procíclicas, através de um processo denominado metaciclogênese, se diferenciam em promastigotas metacíclicas (forma não reprodutiva e infectante). Encontradas na válvula estomodeal do inseto vetor, as formas metacíclicas não possuem a capacidade de se dividir e, após sofrer alterações bioquímicas em sua superfície, perdem a capacidade de adesão ao epitélio do tubo digestivo. Consequentemente, as promastigotas metacíclicas se destacam e migram para a faringe e probóscide do flebotomíneo (GUNN; PITT, 2012; KAYE; SCOTT, 2011).

Durante o repasto sanguíneo, as promastigotas metacíclicas são regurgitadas pelo flebotomíneo, juntamente com proteofosfoglicanos imunomodulatórios derivados do parasito e vários componentes da saliva. Depois de regurgitadas, as promastigotas metacíclicas são introduzidas na pele do hospedeiro vertebrado e fagocitadas por um dos vários tipos de células do sistema imune encontradas no local, tais como macrófagos residentes, células dendríticas da derme (células de Langerhans) e células dendríticas derivadas de monócitos (KAYE; SCOTT, 2011).

A interação inicial entre o parasito e o fagócito ocorre através da extremidade do flagelo. Essa interação é mediada pelos fosfoglicanos e a enzima gp63, localizadas na superfície da promastigota metacíclica, as quais interagem com uma variedade de receptores localizados no macrófago. Após a fixação à superfície do macrófago, o parasito é fagocitado e internalizado em uma vesícula denominada fagossomo. Em seguida, lisossomos se fundem ao fagossomo e liberam enzimas hidrolíticas e peptídeos antimicrobianos no seu interior. Além disso, o conteúdo interno, após essa fusão se torna acidificado e a estrutura resultante é conhecida como vacúolo parasitóforo ou fagolisossomo. A combinação entre o aumento da temperatura, relacionado à entrada no hospedeiro vertebrado, e a diminuição do pH dentro do vacúolo parasitóforo, induzem o

parasito a se diferenciar para forma amastigota. Essa diferenciação, que ocorre de 1-4 horas após a fagocitose, é essencial para a sobrevivência do parasito (GUNN; PITT, 2012; VANNIER-SANTOS; MARTINY; DE SOUZA, 2002).

Apesar de os macrófagos serem especializados no combate a agentes infecciosos, os parasitos do gênero *Leishmania* apresentam mecanismos de defesa capazes de subverter sua capacidade microbicida, conseguindo sobreviver no ambiente potencialmente tóxico do vacúolo parasitóforo e multiplicar-se até a ruptura da célula, quando as formas amastigotas são liberadas para infectar outros macrófagos, propagando a infecção. Esses macrófagos infectados podem ser ingeridos pelas fêmeas do inseto vetor durante o repasto sanguíneo, dando continuidade ao ciclo no hospedeiro invertebrado (GUNN; PITT, 2012; VANNIER-SANTOS; MARTINY; DE SOUZA, 2002).

1.1.2.4. Quimioterapia

Uma vez que vacinas efetivas contra as leishmanioses ainda não foram desenvolvidas, a principal forma de controle dessas doenças consiste na quimioterapia. Porém, vários desafios são encontrados na quimioterapia disponível atualmente, como: poucos fármacos disponíveis, aparecimento de resistência aos fármacos existentes, toxicidade e custo incompatível com a realidade socioeconômica encontrada em países em desenvolvimento (CHAWLA; MADHUBALA, 2010; JAIN; JAIN, 2013). Além desses fatores, o longo período de tratamento e a escassez de opções terapêuticas por via oral (com exceção da miltefosina) contribuem para que os pacientes não concluam o tratamento, aumentado, consequentemente, o risco de desenvolvimento de resistência (FREITAS-JUNIOR et al., 2012). Os fármacos usados no tratamento das leishmanioses e suas principais limitações estão listados na tabela 2.

Tabela 2 - Fármacos utilizados na quimioterapia contra as leishmanioses e suas limitações.

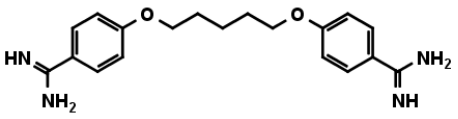
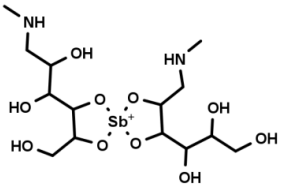
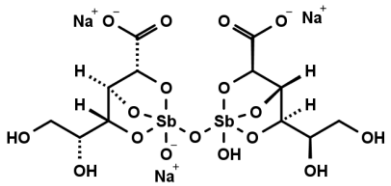
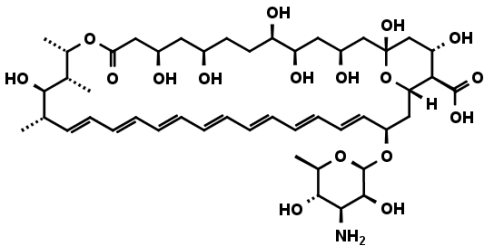
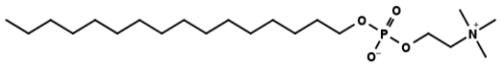
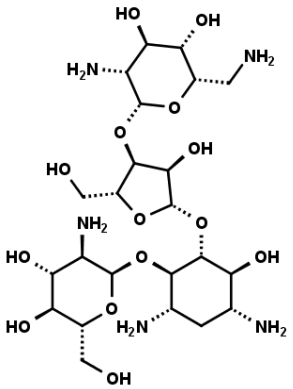
Fármaco	Estrutura química	Limitações
pentamidina		Resistência; Nefrotoxicidade; Cardiotoxicidade; Via parenteral. (KOBETS; GREKOV; LIPOLDOVA, 2012)

Tabela 2 (cont.) – Fármacos utilizados na quimioterapia contra as leishmanioses e suas limitações.

Fármaco	Estrutura química	Limitações
antimoniato de meglumina		Casos de resistência reportados; Cardiotoxicidade; Pancreatite; Administração apenas por via parenteral. (JAIN; JAIN, 2013; NUSSBAUM et al., 2010)
estibogluconato de sódio		
anfotericina B		Resistência *in vitro*; Baixa biodisponibilidade; Nefrotoxicidade; Administração apenas por via parenteral; Difícil regime de tratamento; Formulação lipossomal apresenta alto custo. (CASTILLO et al., 2010; KOBETS; GREKOV; LIPOLDOVA, 2012)
miltefosina		Resistência *in vitro*; Longa meia-vida; Nefrotoxicidade; Perturbações gastrointestinais; Teratogênico. (FREITAS-JUNIOR et al., 2012; NUSSBAUM et al., 2010)
paromomicina		Resistência *in vitro*; Nefrotoxicidade; Ototoxicidade; Hepatotoxicidade. (CASTILLO et al., 2010; FREITAS-JUNIOR et al., 2012)

1.2. Descoberta e desenvolvimento de novos fármacos

A descoberta e o desenvolvimento de novos fármacos é um processo complexo e que demanda tempo, compreendendo várias etapas e elevado investimento financeiro. Esse processo consiste na integração de diversas áreas estratégicas relacionadas à inovação, conhecimento, tecnologia, gerenciamento e altos investimentos em Pesquisa e Desenvolvimento (P&D) (DIMASI; HANSEN; GRABOWSKI, 2003; LOMBARDINO; LOWE, 2004). Todo o processo envolve, em média, 13,5 anos de investimentos em P&D (NICOLAOU, 2014). Estima-se que, na última década, os custos envolvidos em todo o processo de descoberta e desenvolvimento de um novo fármaco, desde a sua concepção até o lançamento no mercado, tenham variado de US$ 200 milhões a US$ 800 milhões, com algumas estimativas indicando custos acima de US$ 1 bilhão (MORGAN et al., 2011c). Segundo Paul e colaboradores, as grandes companhias farmacêuticas investem anualmente US$ 50 bilhões em P&D, aproximadamente. Segundo os mesmos autores, os gastos para o lançamento de uma nova entidade química (NCE, do inglês *New chemical entity*) no mercado, são de aproximadamente US$ 1,8 bilhões (PAUL et al., 2010). Em 2015, dados do *Tufts Center for the study of Drug Development* indicaram que o custo total para desenvolvimento de um novo fármaco é de aproximadamente US$ 2,6 bilhões. Entretanto, tais dados foram objeto de intensa discussão e questionamentos quanto ao método utilizado para se estimar esses gastos (AVORN, 2015).

A gênese de fármacos pode ser dividida em duas grandes etapas: a etapa de descoberta (pesquisa básica, identificação e otimização de compostos líderes) e a etapa de desenvolvimento (estudos pré-clínicos, clínicos e aprovação pela agência reguladora) (ELEBRING; GILL; PLOWRIGHT, 2012; LOMBARDINO; LOWE, 2004). A figura 12 apresenta de forma simplificada essas etapas.

A descoberta de novos fármacos envolve diversas estratégias em Química Medicinal, que são aplicadas na identificação e validação de novos alvos biológicos, identificação de substâncias bioativas (*hits*), identificação e otimização de compostos líderes e na avaliação e otimização de propriedades farmacocinéticas (IMMING, 2011). Segundo definição da União Internacional de Química Pura e Aplicada (IUPAC, do inglês *International Union of Pure and Applied Chemistry*): “A Química Medicinal é

uma disciplina baseada na Química, a qual envolve aspectos das ciências biológicas, médicas e farmacêuticas e está relacionada à invenção, descoberta, planejamento, identificação e preparo de compostos biologicamente ativos, bem como ao estudo do seu metabolismo, interpretação do seu mecanismo de ação em nível molecular e a construção de relações entre a estrutura e atividade"(WERMUTH et al., 1998).

A etapa de descoberta se inicia com a identificação de um potencial alvo biológico relacionado a determinado quadro clínico ou doença e, no caso de doenças parasitárias, relacionado à sobrevivência do parasito e estabelecimento do mesmo no organismo do hospedeiro. Em geral, o alvo biológico consiste em determinada proteína (macromolécula), por exemplo, uma enzima, receptor ou transportador. Preferencialmente, este alvo deve ser validado, ou seja, deve demonstrar um papel importante no processo fisiopatológico estudado (GASHAW et al., 2011; LOMBARDINO; LOWE, 2004; NWAKA; RIDLEY, 2003). Diversas técnicas podem ser usadas na validação de alvos biológicos, como, por exemplo, técnicas que bloqueiam a expresão de genes que codificam o alvo (gene *knock out*, por exemplo), observando a influência, no processo fisiopatológico estudado, da não expressão desses genes (HUGHES et al., 2011). Em doenças infecciosas e parasitárias, a validação do alvo consiste em se avaliar o impacto, para o agente etiológico, da não expressão de determinados genes que codificam determinado alvo, evidenciando assim a importância do mesmo para a sobrevivência do parasito ou microrganismo (CROWTHER et al., 2010; NWAKA; RIDLEY, 2003).

Após a definição do alvo biológico, estudos são realizados para se buscar compostos/ligantes (*hits*) que tenham capacidade de modular *in vitro* a atividade do alvo selecionado. Os *hits*, compostos de origem natural ou sintética, podem ser identificados a partir de triagens experimentais (biológicas ou bioquímicas), virtuais (computacionais) ou por meio de planejamento racional. Um exemplo de triagem experimental é a triagem biológica automatizada em grande escala (HTS, do inglês *High Throughput Screening*), a qual consiste na triagem de uma grande quimioteca de compostos contra um determinado alvo (e.g., enzima) ou um sistema mais complexo (e.g., ensaios em células). Nos estudos computacionais se destacam a organização e compilação de bases de dados, a aplicação de filtros moleculares e a triagem virtual (VS, do inglês *Virtual Screening*). (DUFFY et al., 2012; HUGHES et al., 2011; MIGNANI et al., 2016).

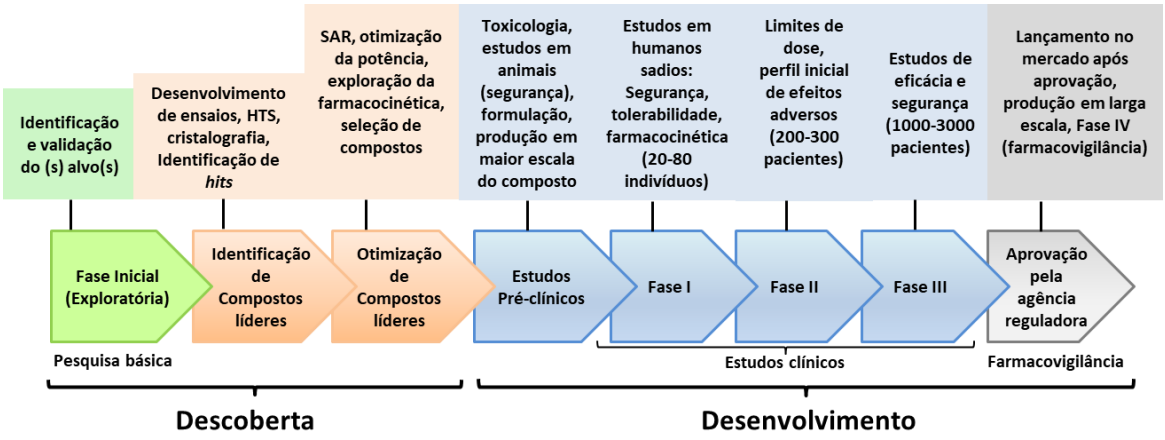


Figura 12 - Etapas envolvidas na descoberta e desenvolvimento de novos fármacos (Adaptado de ELEBRING; GILL; PLOWRIGHT, 2012; NWAKA; RIDLEY, 2003).

Uma vez identificados, os *hits* devem ser otimizados quanto às propriedades farmacodinâmicas (potência, afinidade, seletividade) e farmacocinéticas (absorção, distribuição, metabolismo, biodisponibilidade). Após otimização, os compostos selecionados são denominados compostos líderes (DUFFY et al., 2012).

Durante a otimização de compostos líderes, estudos da relação entre a estrutura e atividade (SAR, do inglês *Structure-Activity Relationships*) e da relação quantitativa entre a estrutura e atividade (QSAR, do inglês *Quantitative Structure-Activity Relationships*) são realizados com objetivo de se obter informações que possam guiar o desenvolvimento de análogos com propriedades otimizadas, como, por exemplo, propriedades ADMET (absorção, distribuição, metabolismo, excreção e toxicidade) e certas propriedades físico-químicas importantes (lipofilicidade, solubilidade, permeabilidade, etc). Ao final da otimização, são obtidas NCEs, ou seja, novos candidatos a fármacos que serão submetidos a estudos pré-clínicos e clínicos (DUFFY et al., 2012; HUGHES et al., 2011; NWAKA; RIDLEY, 2003).

A etapa de desenvolvimento compreende os estudos pré-clínicos e clínicos. Durante os estudos pré-clínicos são realizados ensaios *in vivo* em animais com o objetivo de se avaliar a segurança dos novos candidatos a fármacos. Além disso, estudos de formulação são conduzidos, acompanhados da produção em maior escala dos compostos. Os estudos clínicos compreendem os testes em seres humanos e pode ser divididos em três fases. Na fase I, os candidatos a fármacos, os quais apresentaram resultados promissores nos estudos pré-clínicos, são testados em um pequeno grupo de voluntários sadios (20-80 indivíduos). Essa fase tem como objetivo o estabelecimento

de doses seguras e obtenção de informações a respeito do perfil ADMET dos compostos. Os estudos de fase II são realizados em indivíduos portadores da doença ou quadro clínico investigado (200-300 indivíduos). O objetivo desses estudos é obter informações sobre segurança e traçar o perfil inicial de efeitos adversos. Além disso, resultados preliminares de eficácia são analisados. Os estudos de fase III consistem em testes em grande escala, utilizando um número maior de indivíduos (1000-3000 pacientes). Nesse momento, a eficácia do candidato a fármaco é estabelecida e os efeitos adversos menos comuns são observados. Assim que informações suficientes a respeito da potência e eficácia do composto são obtidas, os resultados são submetidos a um órgão regulador para que o novo fármaco seja aprovado e comercializado (LOMBARDINO; LOWE, 2004; NICOLAOU, 2014; NWAKA; RIDLEY, 2003).

Uma vez no mercado, estudos de farmacovigilância, ou estudos de fase IV, são realizados para monitoramento do novo fármaco na população geral. Pelo fato de a amostragem de indivíduos ser mais ampla nessa fase e o tempo de estudo ser maior, possíveis problemas não detectados em fases anteriores podem ser observados, o que pode levar, em alguns casos, à retirada do fármaco do mercado ou proscrição (LOMBARDINO; LOWE, 2004; NICOLAOU, 2014; NWAKA; RIDLEY, 2003).

1.2.1. A indústria farmacêutica e as doenças tropicais negligenciadas

Um dos objetivos das grandes companhias farmacêuticas está relacionado à descoberta e desenvolvimento de fármacos que atendam às necessidades de pacientes, oferecendo novas terapias e melhoras na segurança e eficácia de tratamentos já existentes. Entretanto, o desenvolvimento de novos fármacos é também uma atividade com fins comerciais, e como tal, depende de retorno nos investimentos para sobrevivência no competitivo mercado global (MIGNANI et al., 2016).

Ao longo de décadas, os investimentos na busca de alternativas terapêuticas para as doenças tropicais negligenciadas (DTNs) foram muito pequenos quando comparados às demais doenças (NWAKA; HUDSON, 2006). Estima-se que, em 2010, apenas 1% dos investimentos em P&D de fármacos tenham sido aplicados às DTNs (RØTTINGEN et al., 2013). No período de 2000-2011, dentre as 336 NCEs aprovadas, apenas quatro (1%) foram destinadas às DTNs. No mesmo período, o número de NCEs aprovadas para desordens neuropsiquiátricas e câncer foi de 49 (15%) e 81 (24%), respectivamente (PEDRIQUE et al., 2013). Esse quadro pode estar relacionado ao fato de que essas

doenças afetam, geralmente, populações mais pobres de países pouco desenvolvidos ou em desenvolvimento, as quais não constituem um mercado consumidor atrativo e, consequentemente, não trazem boas perspectivas de lucro e retorno dos investimentos em P&D para as grandes corporações farmacêuticas. Nesse contexto, há uma grande necessidade de se desenvolver novas alternativas terapêuticas, mais eficientes e seguras, para as DTNs (HOUWELING et al., 2016; NWAKA; HUDSON, 2006).

Nos últimos anos, uma pequena mudança tem sido observada no panorama de investimentos em P&D de fármacos para DTNs. Apesar disso, os investimentos ainda são limitados para grande parte das doenças tropicais. Dentre as mudanças, tem-se observado o aumento das ações de filantropia, crescimento das parcerias público-privadas (PPPs) e a criação de organizações sem fins lucrativos, focadas na busca de novos fármacos para DTNs, como por exemplo, *Drugs for Neglected Diseases initiative* (DNDi) e *Institute for One World Health* (IOWH) (MUSGROVE; HOTEZ, 2009; POLLASTRI, 2014; RENSLO; MCKERROW, 2006).

1.3. Planejamento de fármacos auxiliado por computador

O planejamento de fármacos auxiliado por computador (CADD, do inglês *Computer Assisted Drug Design*), atualmente, permeia diversos aspectos da descoberta de fármacos e apresenta diversas vantagens como auxiliar a identificação de candidatos a fármacos de forma rápida e com custos reduzidos. Dessa forma, a aplicação de métodos computacionais permite a priorização de compostos, os quais apresentam maiores chances de sucesso nas etapas seguintes do planejamento de fármacos (BAJORATH, 2012; DUFFY et al., 2012).

A expansão do uso do CADD se deve, principalmente, aos grandes avanços, obtidos nas últimas décadas, nas áreas de *software* e *hardware*. Aliado a esses avanços, também se destaca o crescimento observado nas áreas de genômica e proteômica, que permitiu o acesso a um grande número de informações estruturais de alvos biológicos (BAJORATH, 2012; KAPETANOVIC, 2008; TALELE; KHEDKAR; RIGBY, 2010). Nos últimos anos, tem sido observado um aumento significativo no número de estruturas de proteínas depositadas no PDB (do inglês, *Protein Data Bank*). Estima-se que, desde 2008, essa base de dados tenha duplicado o número de estruturas depositadas, disponibilizando, em média, 200 novas estruturas a cada semana. Em 2014, 10.368 novas entradas foram depositadas no PDB (PDB, 2014).

De acordo com o nível de conhecimento que se tem do alvo biológico estudado, diferentes abordagens podem ser usadas para o planejamento de novos candidatos a fármacos. Quando a estrutura do alvo macromolecular ou do complexo ligante-receptor está disponível, pode ser usada a estratégia de planejamento de fármacos baseado na estrutura (SBDD, do inglês *Structure-Based Drug Design*). Entretanto, quando a estrutura do alvo não é conhecida, estratégias de planejamento baseado no ligante (LBDD, do inglês *Ligand-Based Drug Design*) são utilizadas. Em muitos casos, o uso combinado das estratégias SBDD e LBDD pode ser muito útil para o planejamento de novos candidatos a fármacos, uma vez que as duas abordagens se complementam (COHEN, 1996; JORGENSEN, 2004).

As estratégias de LBDD e SBDD podem ser aplicadas em estudos de VS, os quais consistem na triagem computacional de grandes quimiotecas disponíveis comercialmente, priorizando os compostos mais promissores contra o alvo biológico estudado a partir do uso de vários filtros moleculares, disponíveis em diversas ferramentas *in silico* (BRAGA et al., 2014; SHOICHET, 2004). Um exemplo de método SBDD é a docagem molecular ou *docking*, que consiste no acoplamento de compostos nos sítios de ligação de proteínas, determinando os modos de ligação mais prováveis e estimando a afinidade de ligação (KLEBE, 2006; SHOICHET, 2004). Alguns exemplos de métodos LBDD são os modelos farmacofóricos, os modelos baseados na forma e volume moleculares (*shape-based*) e a aplicação de modelos de QSAR como filtros para VS (BRAGA et al., 2014). A ferramenta ROCS (do inglês, *Rapid Overlay of Chemical Structures*) é um exemplo de abordagem *shape-based*, em que a forma e volume molecular de inibidores conhecidos, relacionados a determinado alvo, pode ser comparada com a forma e volume molecular de milhares de moléculas ainda não testadas, presentes em bases de dados. Além disso, essa ferramenta também compara as moléculas por meio da sobreposição de características moleculares consideradas importantes para a atividade biológica (GRANT; GALLARDO; PICKUP, 1996; HAWKINS; SKILLMAN; NICHOLLS, 2007).

1.3.1. Relações quantitativas entre estrutura e atividade (QSAR)

Nas últimas décadas, com o avanço das tecnologias de informação e comunicação, observou-se um rápido desenvolvimento na capacidade de coleta, análise e armazenamento de dados químicos e biológicos. Além disso, observou-se o crescente

uso de técnicas modernas como a química combinatória e o HTS, as quais permitiram a obtenção de um grande volume de informações, depositadas em bases de dados (GASTEIGER, 2016; TROPSHA, 2010). Dessa forma, conjuntos de dados contendo milhares de compostos testados em diversos ensaios biológicos estão disponibilizados em bases dados como PubChem e ChEMBL (BENTO et al., 2014; WANG et al., 2014). Essa riqueza de informações pode ser utilizada na construção e validação de modelos computacionais, os quais podem, posteriormente, ser aplicados na VS de novos ligantes disponíveis em quimiotecas comerciais (BRAGA et al., 2014).

Nesse contexto, modelos de QSAR podem ser aplicados de diversas formas, tais como: identificação de novos compostos promissores e predição de sua atividade biológica (e.g., potência, seletividade), discriminação de compostos com propriedades indesejáveis (e.g., modelos para predição de propriedades ADMET) e otimização da atividade biológica de compostos promissores (CHERKASOV et al., 2014).

1.3.1.1. Princípios

A abordagem de QSAR pode ser descrita como a aplicação de métodos de análise de dados e estatísticos para o desenvolvimento de modelos matemáticos, que possam predizer, corretamente, a atividade biológica ou propriedades de compostos baseados em sua estrutura química (TROPSHA, 2010). Basicamente, dois tipos de informação são necessários para se construir modelos de QSAR: (i) um conjunto de compostos com atividade biológica/propriedade definida e (ii) descritores moleculares calculados a partir desses compostos (Figura 13). Entretanto, para que essas duas informações sejam correlacionadas, são aplicados métodos de classificação ou regressão. De forma geral, essa correlação pode ser definida como $A_b = k'(D_1, D_2, ..., D_n)$, em que A_b representa a atividade biológica/propriedade do composto, D são descritores moleculares calculados ou obtidos experimentalmente, e k' é um peso, definido pelo algoritmo computacional, atribuído aos descritores (TROPSHA, 2010; VERMA; KHEDKAR; COUTINHO, 2010).

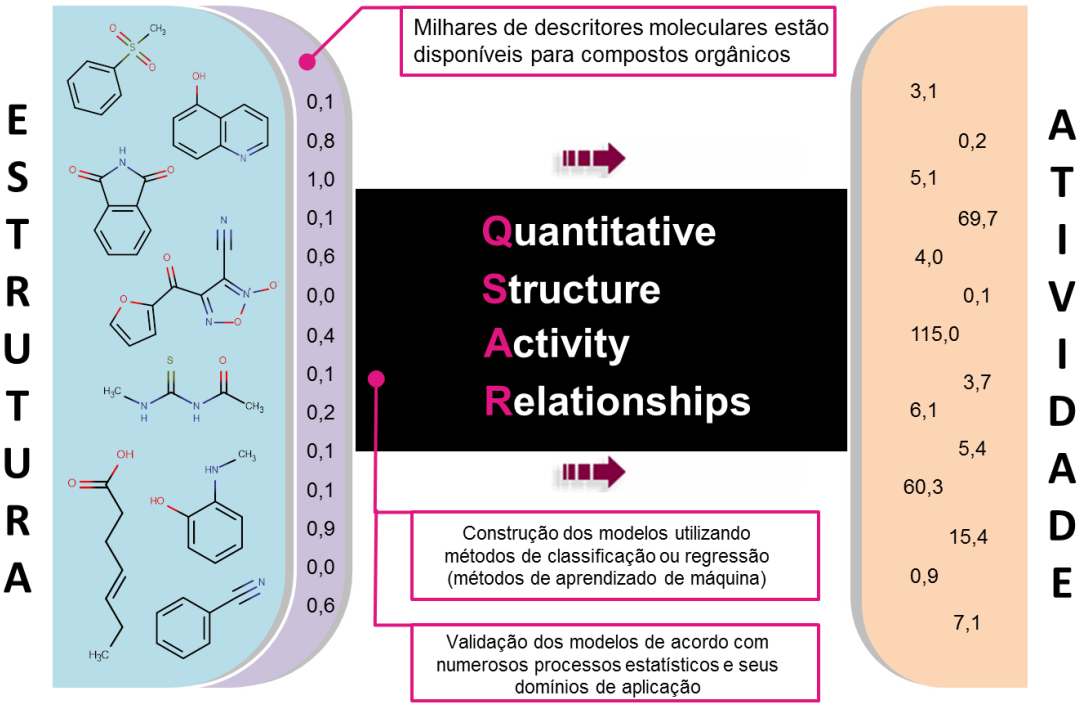


Figura 13 - Representação do processo de construção de um modelo de QSAR (Adaptado de TROPSHA, 2010).

Durante o desenvolvimento de modelos de QSAR, os descritores moleculares e os dados de atividade biológica são ordenados em uma tabela (Tabela 3), em que a variável dependente (Y), representada pela atividade biológica, é armazenada na matriz *y*, enquanto as variáveis independentes (X), representadas pelos descritores moleculares, são armazenadas na matriz *x*. A partir desses dados, métodos matemáticos são utilizados para construção de equações que correlacionem a estrutura com a atividade biológica, as quais podem ser usadas posteriormente para predição de compostos não utilizados na construção dos modelos. A atividade biológica pode ser tratada de forma categórica (binária, por exemplo, ativo ou inativo, tóxico ou não tóxico) ou contínua (por exemplo, potência do composto) (CRONIN, 2010).

Tabela 3 – Representação genérica de uma matriz de dados para um estudo de QSAR.

Compostos	Atividade	Descritor 1	Descritor 2	Descritor 3	...	Descritor n
Composto 1	y_1	x_{11}	x_{12}	x_{13}	...	x_{1n}
Composto 2	y_2	x_{21}	x_{22}	x_{23}	...	x_{2n}
Composto 3	y_3	x_{31}	x_{32}	x_{33}	...	x_{3n}
...	...	...	...	...	...	...
Composto n	y_n	x_{n1}	x_{n2}	x_{n3}	...	x_{nn}

Nos últimos anos, os estudos de QSAR têm apresentado uma grande evolução e diversificação. As análises de QSAR deixaram de ser restritas a pequenas séries de compostos congêneres, utilizando métodos simples de regressão. Atualmente, estudos de QSAR são aplicados a grandes conjuntos de dados com grande diversidade estrutural. Tal evolução se tornou possível graças à implementação, em QSAR, de diversos métodos matemáticos, estatísticos e de aprendizado de máquina (CHERKASOV et al., 2014).

1.3.1.2. Descritores moleculares

O descritor molecular é o resultado final de um procedimento matemático e lógico que transforma a informação química, codificando em números uma representação simbólica do composto. Além disso, o descritor pode ser o resultado de algum experimento padronizado, por exemplo, o coeficiente de partição octanol/água (logP), que consiste em uma medida quantitativa da lipofilicidade de compostos (TODESCHINI; CONSONNI, 2000). Os descritores moleculares são baseados em diversas teorias diferentes, como a química quântica, química orgânica, teoria dos grafos, teoria da informação, dentre outras. Eles são utilizados para modelar várias propriedades diferentes de compostos em diversas áreas científicas como toxicologia, química medicinal, analítica, ambiental, dentre outras (TODESCHINI; CONSONNI, 2009).

Atualmente, há uma grande diversidade de descritores moleculares disponíveis, os quais podem ser classificados de diversas formas, dependendo do critério considerado. De acordo com a dimensionalidade, os descritores podem ser classificados em unidimensionais (1D), os quais se baseiam em propriedades físico-químicas e da fórmula molecular (por exemplo, refratividade molar, logP, massa molecular); bidimensionais (2D), que descrevem propriedades que podem ser calculadas a partir da representação 2D, ou seja, considerando ligações e átomos (por exemplo, número de átomos de nitrogênio e oxigênio, número de ligações, índices de conectividade); tridimensionais (3D), que dependem da conformação das moléculas, ou seja, da disposição dos átomos no espaço tridimensional (por exemplo, área de superfície acessível ao solvente, volume molecular); descritores 4D em diante, que incluem, além da orientação 3D, a dinâmica das moléculas, analisando mudanças conformacionais tempo-dependentes (CHERKASOV et al., 2014).

Conforme a natureza dos descritores moleculares, eles podem ser classificados como: (i) topológicos (ex., índice de conectividade molecular de Randic´, índice de Wiener); (ii) geométricos, que são derivados de coordenadas 3D (ex., área de superfície polar, volume molecular); (iii) constitucionais, que são derivados da composição química dos compostos (ex., número de átomos e ligações, peso molecular); (iv) impressões digitais moleculares (*fingerprints*), que calculam presença/ausência ou frequência de determinados fragmentos na molécula (ex., Avalon, Morgan, etc); (v) eletrostáticos (ex.,cargas parciais, índices de polaridade); (vi) quanto-mecânicos, que são derivados das funções de onda dos elétrons (ex., energia dos orbitais moleculares) (CONSONNI; TODESCHINI, 2010; ROGERS; HAHN, 2010; TODESCHINI; CONSONNI, 2009).

1.3.1.3. Métodos de aprendizado de máquina

Nos últimos anos, houve um aumento significativo na quantidade de informações químicas e biológicas, o que impulsionou o desenvolvimento de novas ferramentas para a análise de dados, como por exemplo, os métodos de aprendizado de máquina. Seguindo essa necessidade de lidar com grandes volumes de dados, esses métodos passaram a ser aplicados em uma grande variedade de estudos de CADD (LAVECCHIA, 2014).

O aprendizado de máquina é uma área da inteligência artificial que se baseia na construção de sistemas capazes de aprender com os dados. Entretanto, não deve ser confundido com o simples ato de armazenar e liberar as informações previamente observadas. De fato, o aprendizado de máquina consiste na transferência (generalização) de propriedades obtidas de dados, previamente observados, para novos dados ainda não observados (WELLING, 2011). Esse aprendizado é construído em três estágios: (i) representação dos dados, os quais devem ser representados em uma linguagem formal com a qual o computador seja capaz de lidar; (ii) otimização da hipótese e (iii) generalização (DOMINGOS, 2012).

Na geração de um modelo de QSAR, por exemplo, os dados são armazenados em uma matriz de dados, conforme representado na tabela 3, discutida anteriormente. Em seguida, uma hipótese é gerada e otimizada para se tentar estabelecer uma relação entre os descritores e a atividade biológica. Supondo que determinada atividade biológica y seja descrita linearmente em função de um único descritor molecular x, um exemplo de hipótese pode ser representado pela equação 1:

$$y(x) = a + bx \qquad \text{(Eq. 1)}$$

Com o objetivo de encontrar valores para os parâmetros a e b, representados na equação 1, algoritmos computacionais (algoritmos de otimização, ex.: *gradient descent*, *conjugate gradient*) são utilizados para minimizar esse valores, de forma que se ajustem e sejam capazes de descrever os dados de entrada (conjunto treinamento). Em seguida, a capacidade de generalização da hipótese é testada, ou seja, sua habilidade de predizer corretamente um exemplo não usado na construção dos modelos (conjunto teste) (DOMINGOS, 2012; WELLING, 2011).

Em geral, os métodos de aprendizado são classificados em duas categorias: supervisionados e não supervisionados (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007). Os métodos supervisionados, categoria usada na geração de modelos de QSAR, se caracterizam por apresentarem uma variável y definida (nesse caso, a atividade biológica), ou seja, os descritores x estão acompanhados de uma categorização correspondente, obtida previamente em um experimento, a atividade biológica (y) real do composto. Portanto, o algoritmo tenta gerar uma função matemática que generalize a variável y. São exemplos de métodos supervisionados: máquina de suporte de vetores (SVM, do inglês *support vector machine*) e florestas aleatórias (RF, do inglês *random forest*) (LAVECCHIA, 2014; WELLING, 2011).

O método de RF se baseia na construção de múltiplas árvores de decisão, cujos resultados são combinados para que seja obtida uma predição final. Cada árvore de decisão é gerada em três etapas: (i) construção do conjunto treinamento a partir de uma amostra retirada do conjunto de dados por *bootstrap*, ao passo que os compostos que não estão no conjunto treinamento formam o conjunto *out-of-bag* (OOB); (ii) seleção da melhor divisão dos descritores através do algoritmo de classificação e regressão de árvores (CART, do inglês *classification and regression trees);* (iii) cada árvore é gerada até a máxima extensão possível, sendo os valores preditos originados da média de todas as predições feitas por cada árvore de decisão (Figura 14) (BREIMAN, 2001; SVETNIK et al., 2003).

Os métodos não supervisionados se caracterizam por serem treinados sem definição da variável y, considerando apenas as variáveis x, ou seja, busca-se identificar padrões no conjunto de dados com base apenas nos descritores (x). Esses métodos são muito úteis na visualização, compressão e organização de dados. São exemplos de

métodos não supervisionados: análise de *clusters*, usada para organização dos dados em agrupamentos; e análise de componentes principais (PCA, do inglês *principal component analysis*), que se baseia na redução de dimensionalidade dos dados (WELLING, 2011).

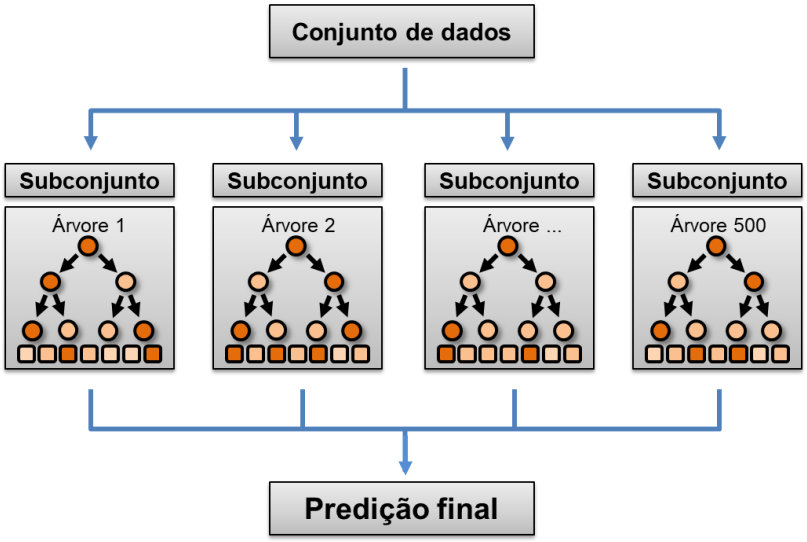


Figura 14 – Esquema geral do método *Random Forest*. O conjunto de dados é separado em subconjuntos contendo linhas e colunas (descritores) selecionadas aleatoriamente. Para cada conjunto, é calculada uma árvore. A combinação das predições de todas as árvores constitui a predição final (Adaptado de NEVES, 2015).

1.3.1.4. Boas práticas de desenvolvimento e validação de modelos de QSAR

Várias diretrizes e recomendações de boas práticas de construção e validação de modelos de QSAR foram disponibilizadas nos últimos anos (CHERKASOV et al., 2014; DEARDEN; CRONIN; KAISER, 2009; TROPSHA, 2010). Dentre elas, se destaca o documento, publicado pela Organização para a Cooperação e Desenvolvimento Econômico (OECD, do inglês *Organization for Economic Co-operation and Development*), o qual sugere que a construção e validação de modelos de QSAR sigam cinco princípios básicos: (i) ter atividade biológica ou propriedade (*endpoint*) definida; (ii) utilizar um algoritmo não ambíguo, ou seja, um algoritmo que apresente clareza; (iii) ter o domínio de aplicabilidade (DA) definido; (iv) ter a robustez e preditividade avaliados de forma apropriada; (v) realização, se possível, da interpretação mecanística dos modelos, a qual consiste em encontrar relações entre os descritores e a atividade biológica, com o objetivo de se compreender o mecanismo de

ação dos compostos ou aprofundar o conhecimento sobre a propriedade em estudo (OECD, 2004, 2007).

1.3.1.4.1. Preparo do conjunto de dados

Nos últimos anos, o avanço em tecnologias como HTS e química combinatória trouxe um aumento significativo no volume de dados químicos e biológicos, disponíveis em bases de dados. Entretanto, acompanhados desse grande volume de dados, surgiram os problemas relacionados à representação correta dessas informações. Dentre os problemas mais comuns se destacam a representação incorreta ou não padronizada de estruturas químicas, a presença de replicatas com dados muito discordantes, erros de anotação dos dados biológicos, dentre outros. Consequentemente, antes de se iniciar a geração de modelos de QSAR, é fundamental que o conjunto de dados seja preparado e padronizado para que os erros não se perpetuem e acabem interferindo na qualidade e preditividade dos modelos gerados a partir desses dados (Figura 15) (TROPSHA, 2010).

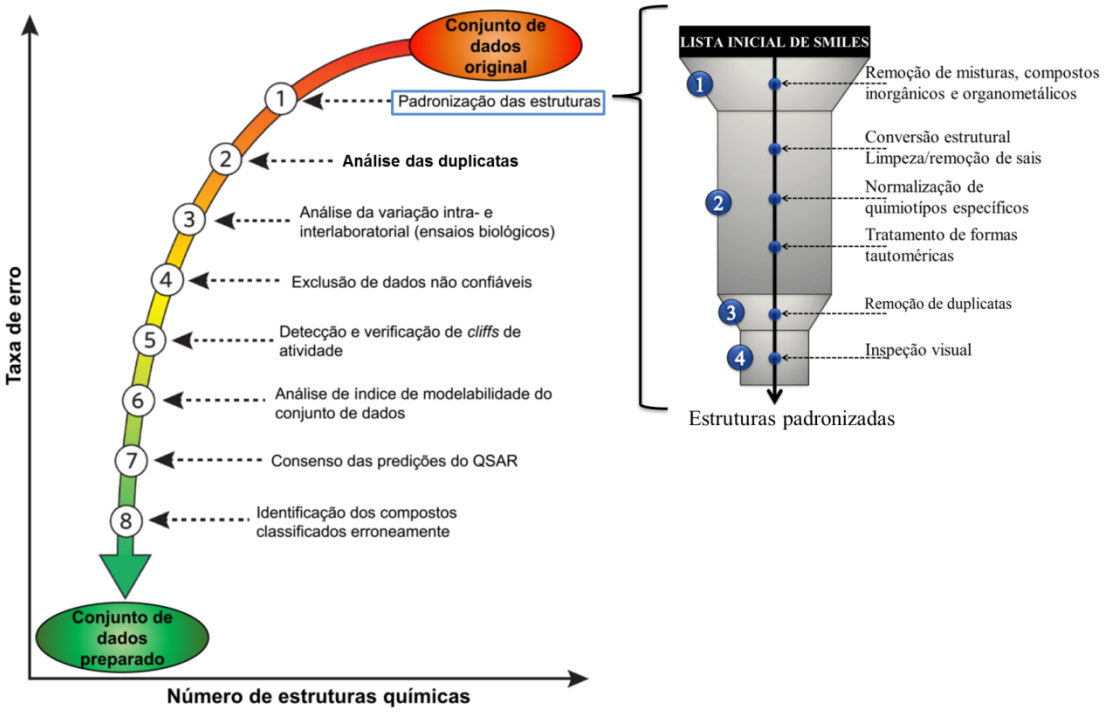


Figura 15 - Representação das etapas para o preparo do conjunto de dados (FOURCHES; MURATOV; TROPSHA, 2015).

Durante o preparo do conjunto de dados, certos compostos precisam ser excluídos por não serem adequadamente tratados pela maioria das técnicas de quimioinformática. Por exemplo, alguns *softwares* não calculam descritores moleculares para compostos inorgânicos e organometálicos, os quais devem ser excluídos durante o preparo do conjunto de dados. Outros aspectos importantes desse processo, dizem respeito à padronização da representação de quimiotipos específicos, como anéis aromáticos e grupos nitro (RNO_2), padronização de formas tautoméricas, remoção de sais e contraíons, dentre outros. Somada a essas etapas, está análise de duplicatas, uma vez que ela pode fornecer informações a respeito da qualidade do conjunto de dados. As duplicatas devem ser identificadas e removidas, a fim de que entradas repetidas não contribuam para geração de modelos com informações equivocadas e que levem a conclusões errôneas. Ao final de todo o processo de preparo, é fundamental que seja feita uma inspeção manual, a fim de se garantir a representação correta das estruturas e dados biológicos (FOURCHES; MURATOV; TROPSHA, 2016, 2010).

1.4. Bases moleculares para a descoberta de novos fármacos contra as leishmanioses

A disponibilidade da sequência genômica completa de espécies de *Leishmania* como, *L. major* (IVENS et al., 2005), *L. infantum* e *L. braziliensis* (PEACOCK et al., 2007), abriu oportunidades para o entendimento da biologia desses organismos. A análise do genoma pode contribuir de forma significante para o planejamento de fármacos, pois permite a identificação de novas enzimas, receptores e, consequentemente, a escolha de alvos exclusivos dos parasitos ou que apresentem diferenças significativas em comparação aos seus homólogos em humanos (CHAWLA; MADHUBALA, 2010; RAJASEKARAN; CHEN, 2015).

Uma etapa importante no planejamento de novos fármacos antiparasitários é a escolha do alvo molecular (e.g. receptor, enzima), o qual deve desempenhar determinada função que seja vital, em alguma via bioquímica do parasito. Vários fatores são importantes para a escolha de um alvo, tais como: identificação do estágio do ciclo biológico do parasito em que o alvo é expresso, investigação das propriedades bioquímicas do alvo molecular e identificação de sítios de ligação para que inibidores específicos possam ser planejados. Além disso, o alvo molecular selecionado deve

permitir a realização de ensaios *in vitro*, utilizando sistemas específicos, que não exijam custos muito elevados (CHAWLA; MADHUBALA, 2010; CRUZ et al., 2009; RAJASEKARAN; CHEN, 2015).

Diversas vias bioquímicas de *Leishmania* têm sido exploradas na busca de alvos promissores para o planejamento de novos fármacos leishmanicidas. Dentre essas vias, se destacam: a via glicolítica, a biossíntese de esteróis de membrana, a via de salvação de purinas, a biossíntese de glicosilfosfatidilinositol, a biossíntese de folatos, o sistema glioxalase, a via da tripanotiona e a via da hipusina (CHAWLA; MADHUBALA, 2010; NAGLE et al., 2014; RAJASEKARAN; CHEN, 2015). Além dessas vias bioquímicas, recentemente tem sido observado um crescente interesse por vias de sinalização relacionadas às proteínas quinases, uma vez que foi demonstrado que os genomas de tripanossomatídeos codificam um grande número dessas proteínas, constituindo cerca de 2% dos genes preditos como codificadores de proteínas nesses organismos (MERRITT et al., 2014).

1.4.1. Via glicolítica

A via glicolítica constitui uma importante via metabólica para a maioria dos organismos, incluindo os tripanossomatídeos. Essa via apresenta papel importante na produção de ATP e, portanto, a exploração das enzimas dessa via pode auxiliar na identificação de alvos promissores para o desenvolvimento de novos fármacos (HARRIS; MITCHELL; MORRIS, 2014) (Figura 16).

As formas amastigotas de *Leishmania* spp., encontradas no interior dos macrófagos do hospedeiro vertebrado, sobrevivem em um ambiente em que os aminoácidos e os ácidos graxos são os nutrientes predominantes para a produção de ATP (OPPERDOES; COOMBS, 2007). Portanto, em contraste ao que ocorre em *T. brucei*, a via glicolítica parece não desempenhar um papel crucial na forma amastigota de *Leishmania* spp. Entretanto, em experimentos de inibição da via glicolítica, foi demonstrada a diminuição no crescimento de parasitos do gênero *Leishmania*, como por exemplo, em estudos de inibição da enzima piruvato quinase (PK, E.C. 2.7.1.40), reportados na literatura (MORGAN et al., 2011a, 2012; NOWICKI et al., 2008). Tais resultados levantaram um paradoxo que, aparentemente, pode ser explicado pela dependência, dos parasitos do gênero *Leishmania*, de uma via altamente relacionada à glicólise: a gliconeogênese (BARROS-ALVAREZ et al., 2014). A importância da gliconeogênese,

em amastigotas de *Leishmania* spp., está relacionada à produção de glicoconjugados e oligossacarídeos de reserva como o manogênio, importantes para a manutenção do parasito (BARROS-ALVAREZ et al., 2014; VERLINDE et al., 2001).

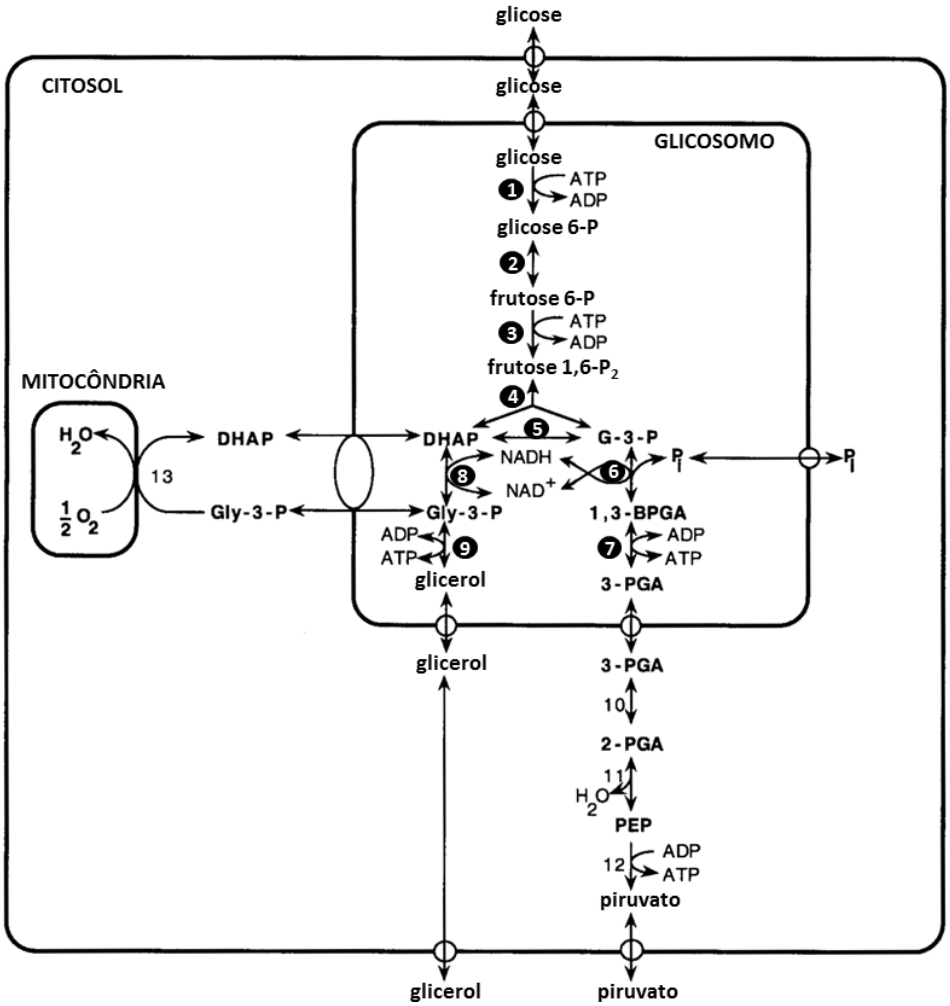


Figura 16 - Esquema geral da via glicolítica em tripanossomatídeos (Adaptado de VERLINDE et al., 2001). Enzimas situadas dentro do glicossomo: (1) hexoquinase; (2) glicose-6-fosfato; (3) fosfofrutoquinase; (4) aldolase; (5) triosefosfato isomerase; (6) gliceraldeído-3-fosfato desidrogenase; (7) fosfoglicerato quinase; (8) glicerol-3-fosfato desidrogenase; (9) glicerol quinase. Enzimas situadas fora do glicossomo: (10) fosfoglicerato mutase; (11) enolase; (12) piruvato quinase; (13) glicerol-3-fosfato oxidase; Abreviações: 1,3-PGA: 1,3-bisfosfoglicerato; DHAP: dihidroxiacetona fosfato; G-3-P: gliceraldeído 3-fosfato; Gly-3-P: glicerol 3-fosfato; PEP: fosfoenolpiruvato; 2-PGA: 2-fosfoglicerato; 3-PGA: 3-fosfoglicerato.

Nos cinetoplastídeos, a maior parte da via glicolítica está situada dentro de organelas especializadas denominadas glicossomos, permitindo a regulação dessa via e

capacitando a célula a superar períodos curtos de anaerobiose. Apesar da compartimentalização das sete primeiras enzimas da via glicolítica, as três últimas enzimas dessa via não são compartimentalizadas e estão presentes no citosol, sendo a última delas, a enzima PK, responsável pela síntese de ATP e piruvato (BARROS-ALVAREZ et al., 2014). Além de abrigar maior parte da via glicolítica, o glicossomo apresenta outras vias importantes, como a via das pentoses-fosfato, β-oxidação, salvação de purinas e vias para biossíntese de pirimidinas, éter-lipídios e esqualenos (GUALDR et al., 2013; MICHELS et al., 2006).

1.4.1.1. Piruvato quinase (PK)

A enzima PK catalisa a reação final e irreversível da glicólise em que, fosfoenolpiruvato (PEP) e ADP são convertidos em piruvato e ATP, respectivamente. Além disso, durante a catálise, participam os íons metálicos K^+ e Mg^{2+} (MORGAN et al., 2011a).

A PK é uma enzima homotetramérica e apresenta, em cada monômero, 498 resíduos de aminoácidos, sendo estes distribuídos em quatro domínios: A, B, C e N-terminal (MORGAN et al., 2011a; RIGDEN et al., 1999). Cada monômero contém um sítio catalítico e um sítio alostérico. O sítio catalítico é constituído pelo sítio de ligação do PEP e pelo sítio de ligação de ADP/ATP. No sítio alostérico ou efetor, se liga a molécula efetora que, nos tripanossomatídeos, é a frutose-2,6-bifosfato (F-2,6-BP) (RIGDEN et al., 1999). A PK apresenta uma grande mobilidade, visto que a ligação da molécula efetora em seu sítio alostérico promove mudanças conformacionais na enzima, alterando assim a afinidade de ligantes que realizam interações em seu sítio catalítico (MORGAN et al., 2011a). De acordo com as mudanças conformacionais observadas, a PK pode assumir dois estados básicos: estado-R, que corresponde ao estado ativo da enzima, no qual o efetor alostérico encontra-se ligado; estado-T, que corresponde ao estado inativo da enzima, na ausência do efetor alostérico (TULLOCH et al., 2008) (Figura 17).

Em 1904, estudos conduzidos por Paul Ehrlich demonstraram a atividade tripanocida, em camundongos infectados, de uma série de corantes de naftaleno relacionados ao azul de tripano e ao vermelho de tripano. Esses estudos serviram como ponto de partida para a síntese da suramina, em 1916, a qual é utilizada até hoje no tratamento da doença do sono, causada por *T. brucei* (MORGAN et al., 2011a). Estudos demonstraram que a suramina foi capaz de inibir sete enzimas glicolíticas purificadas a

partir do glicossomo, além da PK citosólica, com valores de IC_{50} entre 3 e 100 μM (BARROS-ALVAREZ et al., 2014; MORGAN et al., 2011a; WILLSON et al., 1993). Além disso, em estudos de validação de enzimas glicolíticas como alvos biológicos, incluindo a PK, em *T. brucei*, utilizando interferência por RNA (RNAi), foi observada a morte do parasito (MORGAN et al., 2011a). A adoção de *T. brucei* como organismo modelo de tripanossomatídeos para validação de enzimas, pelo método de RNAi, se deve principalmente às tentativas falhas de tornar esse método funcional em *T. cruzi* e *Leishmania* spp. (BRINGAUD; RIVIÈRE; COUSTOU, 2006).

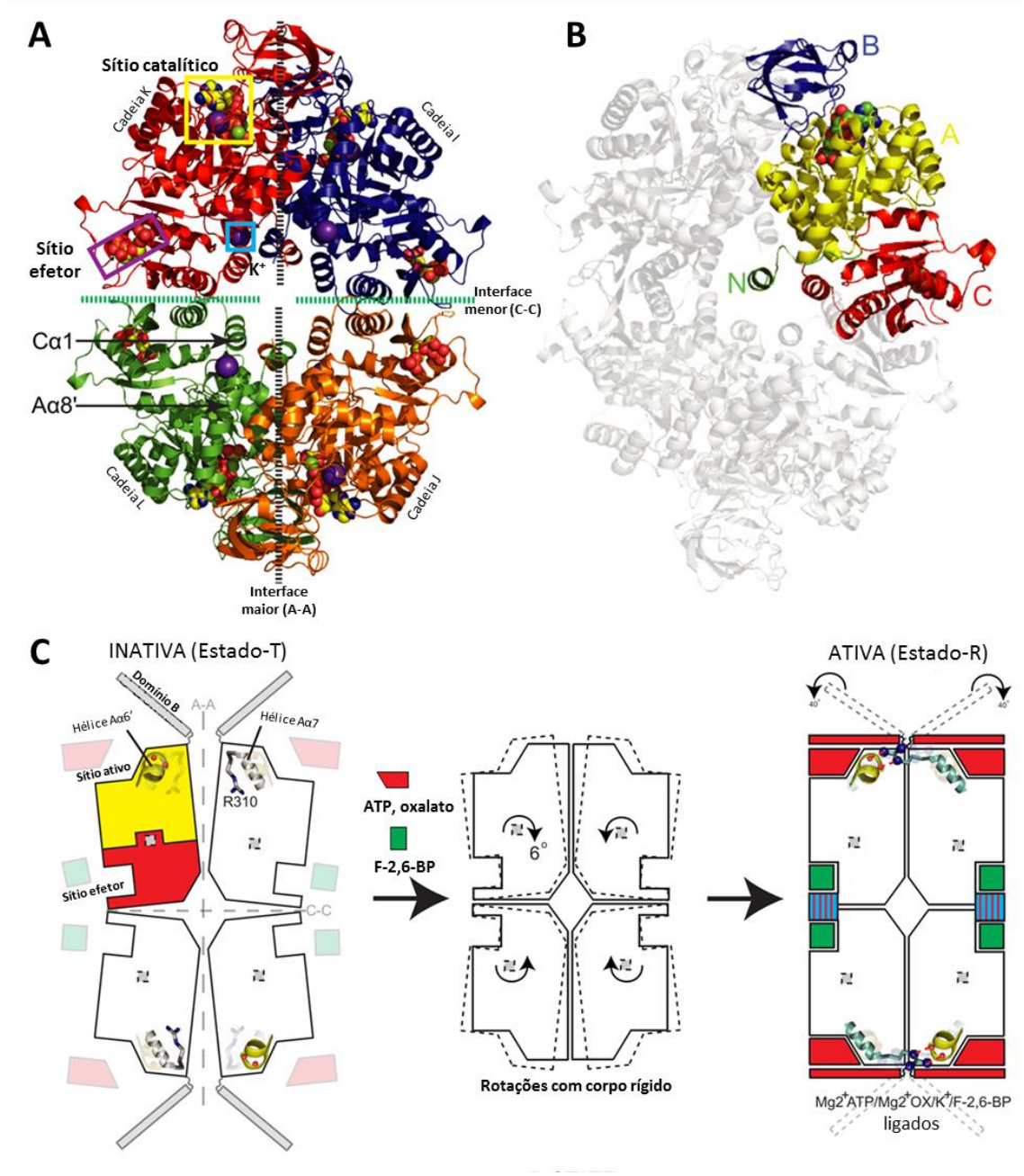


Figura 17 - Estrutura da PK de *L. mexicana* (*Lm*PK) em complexo com ATP, oxalato e F-2,6-BP. (A) Estrutura tetramérica da *Lm*PK evidenciando o sítio catalítico, o sítio

efetor (alostérico) e o potássio (K^+). Cada monômero está representado por uma cor diferente, demostrando as interfaces entre as subunidades. A interface maior (A-A) e a interface menor (C-C) estão representadas por linhas pontilhadas em preto e verde, respectivamente. (B) Destaque de uma das subunidades do tetrâmero, a qual está colorida de forma a mostrar os diferentes domínios que constituem cada subunidade (A, B, C e N-terminal). (C) Representação do estado-T (inativo) da enzima e as mudanças conformacionais, após a ligação do efetor F-2,6-BP e oxalato, levando a enzima a assumir o estado-R (ativo) (Adaptado de MORGAN et al., 2010).

Estruturas cristalográficas da enzima PK de *L. mexicana*, em complexo com a suramina e outros três corantes relacionados, indicam que essas moléculas se ligam ao sítio de ligação de ADP/ATP, localizado próximo ao sítio catalítico da enzima (MORGAN et al., 2011a). O grupo sulfona, presente na suramina e seus análogos, mimetiza o grupo α-fosfato do ATP. Dessa forma, essas moléculas ocupam o sítio de ligação de ADP/ATP, bloqueando a interação do ADP nesse sítio e a conversão do mesmo em ATP (MORGAN et al., 2011a). As estruturas cristalográficas da suramina e seus análogos em complexo com a PK estão disponíveis no *Protein Data Bank* (PDB IDs: 3PP7, 3QV6, 3QV8, 3QV7, 3QV9). Estudos experimentais indicaram um aumento da potência da suramina quando o efetor alostérico (F-2,6-BP) estava presente (K_i =115,6 ± 2,6 μM diminuído para K_i = 57 ± 1,2 μM) indicando a importância das mudanças conformacionais resultantes da regulação alostérica da PK, permitindo uma melhor acomodação da suramina em seu sítio de ligação. A presença do efetor ligado ao sítio alostérico reflete a provável situação observada *in vivo*, uma vez que existe uma saturação de F-2,6-BP em tripanossomatídeos (MORGAN et al., 2011a).

A PK de *Leishmania* spp. apresenta diferenças em comparação à PK de humanos. Todas as isoenzimas da PK encontradas em humanos, com exceção da isoenzima M1, que não possui regulação alostérica, apresentam a frutose-1,6-bifosfato (F-1,6-BP) como molécula efetora, enquanto a PK de *Leishmania* utiliza a F-2,6-BP. Além disso, estudos comparativos entre a PK de *L. mexicana* e a PK de humanos demonstraram diferenças importantes em cinco resíduos do sítio de ligação de ADP/ATP (TULLOCH et al., 2008). Estudos reportados na literatura indicam que enzimas glicolíticas de tripanossomatídeos são pouco conservadas em relação às enzimas correspondentes no hospedeiro. Comparações entre as sequências da PK de *Leishmania* spp., *T.cruzi e T. brucei*, em relação à PK de humanos, demonstraram uma

identidade sequencial de 49%, 48% e 49%, respectivamente (HARRIS; MITCHELL; MORRIS, 2014). Essas diferenças podem ser importantes para o planejamento de inibidores seletivos da PK do parasito.

A enzima PK apresenta alta similaridade entre os tripanosomatídeos (identidade sequencial >70%). Consequentemente, espera-se que inibidores de PK ativos em um tripanossomatídeo tenham grandes chances de serem ativos nos outros. Portanto, a busca de novos inibidores da enzima PK abre possibilidades de se desenvolver novos candidatos a fármacos contra *T. cruzi* e várias espécies do gênero *Leishmania* (NOWICKI et al., 2008).

1.4.2. Biossíntese de esteróis de membrana

Os esteróis são componentes indispensáveis das células eucarióticas desempenhando diversas funções como: manutenção da fluidez e permeabilidade da membrana, modulação de proteínas de membrana e canais iônicos, precursores de moléculas com funções de regulação de vários processos biológicos (LEPESHEVA; WATERMAN, 2011).

A biossíntese de esteróis de membrana em vertebrados, fungos, protozoários e plantas tem início através da condensação da acetilcoenzima A (Acetil-CoA) e continua através da formação de diversos intermediários chegando ao farnesil-pirofosfato, que por sua vez, leva a formação do esqualeno. Na presença de oxigênio molecular e por ação da esqualeno epoxidase, o esqualeno é transformado em 2,3-óxido de esqualeno (LEPESHEVA; VILLALTA; WATERMAN, 2011). Em vertebrados, fungos e protozoários o 2,3-óxido de esqualeno é transformado em lanosterol, enquanto em plantas ele é convertido em fitoesterol a partir do precursor cicloartenol (CHOI; PODUST; ROUSH, 2014). Os produtos finais da biossíntese de esteróis variam entre os eucariotos (Figura 18). A partir do lanosterol os fungos e protozoários (tripanossomatídeos, por exemplo) produzem o ergosterol enquanto os mamíferos produzem o colesterol; a plantas, por sua vez, produzem uma variedade de fitoesteróis a partir do cicloartenol (BUCKNER, 2008). A biossíntese desses diferentes produtos finais requer a remoção do grupo 14α–metil de seus respectivos precursores de esteróis. A reação responsável por essa remoção é catalisada pela enzima 14 α–esterol desmetilase (CYP51) (E.C. 1.14.13.70), que é uma monooxigenase da superfamília citocromo P450 (BUCKNER, 2008; LEPESHEVA et al., 2008).

A complexa reação catalisada pela CYP51 ocorre em três etapas sucessivas que levam à clivagem da ligação C-C, na posição C-14 e remoção do grupo metila do C-32 do lanosterol (Figura 18).

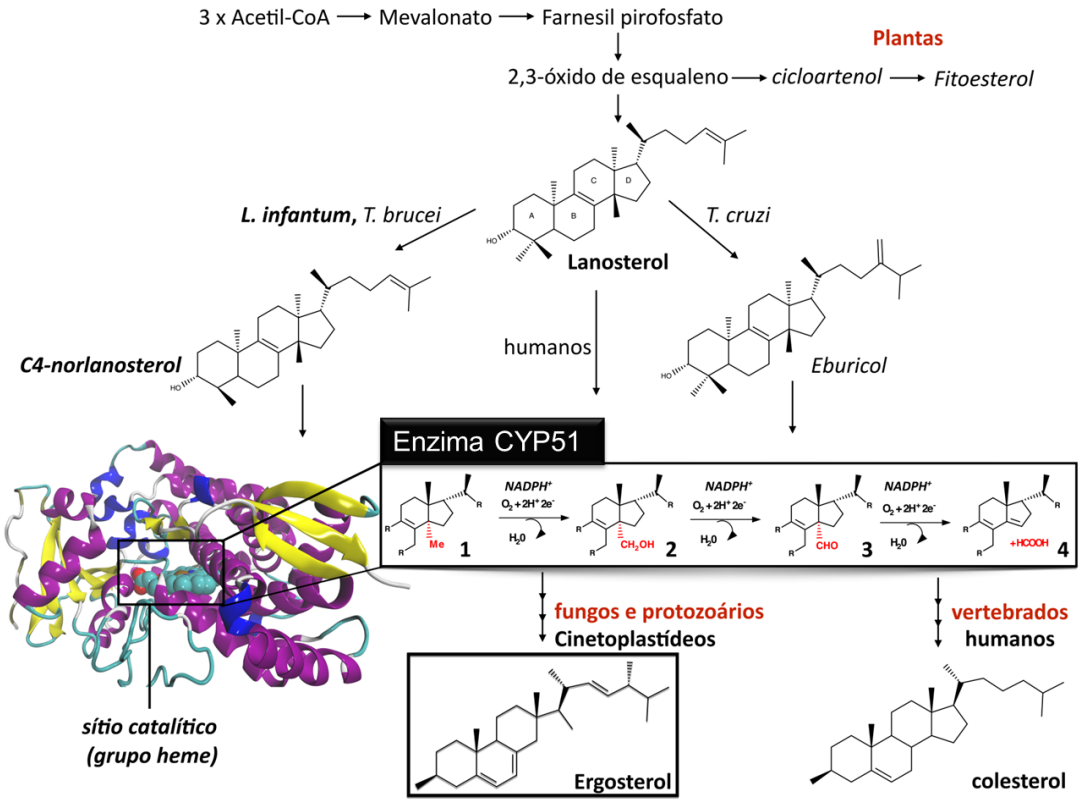


Figura 18- Biossíntese de esteróis a partir do acetil-CoA até a formação de colesterol (vertebrados), ergosterol (fungos e protozoários) e fitoesterol (plantas) (Adaptado de CHOI; PODUST; ROUSH, 2014; HARGROVE et al., 2012)**.**

Em cada etapa, são necessários uma molécula de oxigênio e dois equivalentes redutores derivados de NADPH (LEPESHEVA; WATERMAN, 2007). Na primeira etapa, o substrato se liga no sítio ativo e aceita o primeiro elétron da proteína de transferência de elétrons, reduzindo e ativando o átomo de ferro do grupo heme. Isso permite a ligação de uma molécula de oxigênio na posição axial distal do átomo de ferro, próximo ao grupo metila da posição 14α. Em seguida, o segundo elétron é transferido, formando o carbono radical. Dois prótons causam a cisão do oxigênio molecular e uma molécula de água é liberada, enquanto um dos átomos de oxigênio é transferido ao grupo -CH_2 do lanosterol (**1**), levando à formação do oxiesterol (**2**). Na segunda etapa da catálise, o grupo hidroxila da posição 14α é convertido em aldeído (**3**). A terceira etapa consiste na liberação de ácido fórmico concomitante com a

formação da dupla ligação entre C14-C15 levando ao produto final (**4**) (CHOI; PODUST; ROUSH, 2014; LEPESHEVA; VILLALTA; WATERMAN, 2011; LEPESHEVA; WATERMAN, 2007).

Os fármacos antifúngicos azólicos são conhecidos inibidores da enzima CYP51 de fungos e revolucionaram o tratamento de infecções fúngicas tópicas e sistêmicas. Ao longo dos anos, vários antifúngicos azólicos e seus derivados foram testados contra cinetoplastídeos (LEPESHEVA; VILLALTA; WATERMAN, 2011). Além disso, estudos de reposicionamento desses fármacos para o tratamento da doença de Chagas têm sido executados (FIELD et al., 2017).

Recentemente, o posaconazol e o ravuconazol, dois antifúngicos azólicos, foram testados em ensaios clínicos para o tratamento da fase crônica da doença de Chagas. Ambos demonstraram a capacidade de promover, inicialmente, uma redução importante na quantidade de parasitos. Entretanto, após o fim do tratamento, os pacientes apresentaram recorrência da doença (MOLINA et al., 2014). Tais resultados mostram a dificuldade de se usar os antifúngicos azólicos em monoterapia, mas não descartam o seu possível uso em combinação com outros fármacos (FIELD et al., 2017). Além disso, esses resultados não descartam a CYP51 como um alvo promissor no planejamento de novos fármacos contra cinetoplastídeos, uma vez que novos inibidores podem ser descobertos através de diversas estratégias alternativas como a otimização de compostos líderes, o planejamento baseado nos ligantes (LBDD) e o planejamento baseado estrutura (SBDD). Nos últimos anos, vários compostos líderes promissores foram descobertos (e.g. VNI, VNF, VFV, derivados do fenarimol) e várias estruturas cristalográficas da CYP51 de *T. cruzi*, *T. brucei* e *L. infantum* foram disponibilizadas no PDB (CHOI; PODUST; ROUSH, 2014).

Em estudos recentes, a essencialidade da CYP51 em *L. donovani* foi demonstrada através de estudos de silenciamento de genes e abordagens farmacológicas, em que a inibição de CYP51 causou a inibição do crescimento do parasito (MCCALL et al., 2015). Dessa forma, a CYP51 é um alvo validado para o planejamento de novos agentes leishmanicidas.

2. JUSTIFICATIVA

A doença de Chagas e as leishmanioses são DTNs que afetam milhões de pessoas em países em desenvolvimento. Entretanto, ao longo de décadas, os investimentos na busca de alternativas terapêuticas para as DTNs foram insuficientes quando comparados às demais doenças (PEDRIQUE et al., 2013). Os poucos fármacos disponíveis para essas doenças apresentam diversos problemas relacionados à resistência, toxicidade e eficácia, tornando urgente a necessidade de se buscar novas opções terapêuticas (BERMUDEZ et al., 2016; FREITAS-JUNIOR et al., 2012). Portanto, a busca por novos fármacos contra essas doenças se justifica pelos seguintes fatores:

1. Ambas as doenças são consideradas um grande problema de saúde pública, podendo levar os pacientes a quadros clínicos graves e, consequentemente, à morte;
2. Ambas são doenças endêmicas em diversos países pouco desenvolvidos ou em desenvolvimento, inclusive o Brasil;
3. Os fármacos disponíveis para essas doenças apresentam problemas de resistência, toxicidade, baixa eficácia na fase crônica (Chagas) e, em alguns casos, custo elevado (anfotericina B lipossomal para as leishmanioses);
4. Escassez de opções terapêuticas por via oral contra as leishmanioses (a miltefosina é o único fármaco disponível por via oral);

O renascimento do interesse por triagens fenotípicas de compostos, observado nos últimos anos, combinado ao desenvolvimento de métodos como o HTS, têm contribuído para a geração de um grande volume dados oriundos de ensaios em células, os quais são depositados em bases de dados como o ChEMBL e PubChem BioAssay (GAULTON et al., 2017; WANG et al., 2017; ZHENG; THORNE; MCKEW, 2013). Dentre os dados, se destacam aqueles relacionados a ensaios contra tripanossomatídeos, que podem ser utilizados no desenvolvimento de modelos de QSAR preditivos, permitindo a triagem virtual e validação experimental de novos compostos promissores contra esses protozoários.

Os avanços nas áreas de genômica, proteômica, biologia estrutural e molecular permitiram o sequenciamento do genoma de determinadas espécies de *Leishmania* e, dessa forma, abriram diversas possibilidades na busca de alvos biológicos promissores

para a descoberta de novos fármacos contra as leishmanioses (IVENS et al., 2005; PEACOCK et al., 2007). Dentre os alvos promissores, destacam-se a enzima piruvato quinase (PK) e a enzima esterol 14 α-desmetilase (CYP51), as quais participam das vias glicolítica e de biossíntese de esteróis, respectivamente.

A pesquisa de compostos inibidores das enzimas PK e CYP51 de *Leishmania*, como novos candidatos a fármacos leishmanicidas, se justifica pelos seguintes fatores:

1. A enzima PK apresenta um papel importante na biossíntese de ATP, além de interferir, indiretamente, na gliconeogênese, a qual apresenta papel determinante na manutenção do parasito dentro dos macrófagos;
2. A enzima PK de *Leishmania* apresenta diferenças em relação às isoformas encontradas em humanos, tanto no sítio de ligação do ATP, quanto no sítio alostérico (efetor);
3. Há uma grande quantidade de informações sobre inibidores da PK de *Leishmania* em bases de dados como PubChem BioAssay, permitindo a construção de modelos de QSAR preditivos e modelos baseados no ligante. Além disso, a estrutura tridimensional da PK de *Leishmania mexicana* está disponível, em complexo com o inibidor de PK suramina, o qual é um fármaco já utilizado contra outro tripanossomatídeo, o *T. brucei*. Dessa forma, é possível construir modelos baseados na estrutura, para se identificar potenciais inibidores de PK;
4. A enzima CYP51 apresenta papel importante na manutenção da integridade da membrana do parasito, uma vez que atua na biossíntese do ergosterol. Nos mamíferos o ergosterol está ausente, sendo substituído pelo colesterol;
5. A disponibilidade de informações na literatura sobre inibidores de CYP51 de *Leishmania* e a existência estruturas tridimensionais dessa enzima em complexo com inibidores conhecidos permitem a geração de modelos baseados no ligante e na estrutura, os quais podem ser usados na triagem virtual de novos potenciais inibidores.

3. OBJETIVOS

O objetivo do presente trabalho foi identificar novos candidatos a fármacos contra a doença de Chagas e leishmanioses, através da aplicação de estratégias integradas em química medicinal.

3.1. Objetivos específicos

Para melhor compreensão dos estudos realizados, esta tese foi dividida em três grandes objetivos, relacionados a três trabalhos que integraram diferentes abordagens do planejamento de fármacos auxiliado por computador (CADD) e validação experimental:

3.1.1. Identificação de novos compostos bioativos contra *Trypanosoma cruzi*

- Gerar e validar modelos de QSAR binários (classificatórios) para predição de atividade contra *T. cruzi* e de citotoxicidade em células de mamíferos a partir de conjuntos de dados fenotípicos reportados na literatura;
- Realizar triagem virtual de uma base de dados com mais de 1 milhão de compostos comerciais, utilizando os modelos de QSAR previamente gerados e validados, a fim de selecionar um pequeno grupo de compostos com potencial atividade e seletividade contra *T. cruzi*;
- Realizar avaliação experimental *in vitro* dos compostos selecionados contra estágio intracelular de *T. cruzi* e células de mamíferos;
- Realizar uma análise multiparamétrica dos compostos selecionados, a fim de comparar o balanço entre potência, seletividade, propriedades físico-químicas e farmacocinéticas.

3.1.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp.

- Gerar e validar modelos baseados na forma e volume moleculares (*shape-based*) e modelos baseados na estrutura (SBDD) a partir de conjuntos de dados de inibidores da enzima PK de *Leishmania mexicana*, reportados na literatura;

- Gerar e validar modelos de QSAR binários para predição de atividade contra a enzima PK a partir de conjuntos de dados reportados na literatura;
- Realizar triagem virtual de base de dados com mais de 1 milhão de compostos comerciais, utilizando os modelos *shape-based* e SBDD previamente gerados e validados, a fim de selecionar um pequeno grupo de compostos com potencial inibitório contra a enzima PK de *Leishmania*;
- Realizar avaliação experimental *in vitro* dos compostos selecionados contra formas amastigotas de *Leishmania infantum* e células de mamíferos.

3.1.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp.

- Gerar e validar modelos *shape-based* a partir de conjuntos de dados de inibidores das enzimas PK e CYP51 de *Leishmania* reportados na literatura;
- Realização de triagem virtual *dual-target* com os modelos previamente gerados e validados para as enzimas PK e CYP51, para seleção de um pequeno grupo de compostos com potencial inibitório contra ambas as enzimas;
- Realizar a predição da atividade biológica dos potenciais inibidores *dual-target* através de modelos de QSAR baseados em dados de ensaios fenotípicos contra forma amastigota de *L. infantum*;
- Realizar avaliação experimental *in vitro* dos compostos selecionados contra formas amastigotas de *L. infantum* e células de mamíferos.

4. MÉTODOS

4.1. Identificação de novos compostos bioativos contra *T. cruzi*

4.1.1. Estratégias Computacionais

4.1.1.1. Conjuntos de dados

Foram utilizados dois conjuntos de dados extraídos da base de dados *PubChem Bioassay* (WANG et al., 2017), contendo resultados de ensaios confirmatórios de HTS realizados no *Broad Institute of MIT and Harvard*:

- Conjunto de dados I: contendo compostos testados em ensaio de inibição da replicação intracelular de *T. cruzi* (PubChem Identifier: AID 2044; URL: *https://pubchem.ncbi.nlm.nih.gov/bioassay/2044*). Baseado em um limiar de atividade de 1,15 µM, esse conjunto de dados apresentava 2.044 compostos com $EC_{50} \leq 1{,}15$ µM (inibidores), 2.016 compostos com $EC_{50} > 1{,}15$ µM (não-inibidores) e cinco compostos com resultados inconclusivos;
- Conjunto de dados II: contendo todos os compostos do conjunto de dados I, os quais foram submetidos a um ensaio paralelo para avaliação da citotoxicidade contra fibroblastos embrionários de camundongos (NIH/3T3), a mesma linhagem celular utilizada como célula hospedeira nos ensaios de replicação intracelular de *T. cruzi* (PubChem Identifier: AID 2010; URL: *https://pubchem.ncbi.nlm.nih.gov/bioassay/2010*). Baseado em um limiar de 50 µM, o conjunto de dados apresentava 1.024 compostos com $EC_{50} \leq 50$ µM (citotóxicos), 2.487 compostos com $EC_{50} > 50$ µM (não citotóxicos) e 554 compostos com resultados inconclusivos;

4.1.1.2. Preparo e padronização dos conjuntos de dados

Todas as estruturas químicas e suas informações de atividade biológica foram analisadas e preparadas de acordo com os protocolos desenvolvidos por Fourches e colaboradores (FOURCHES; MURATOV; TROPSHA, 2016, 2010, 2015). Todas as etapas de preparo e padronização foram executadas no protocolo *in house* KSAR 1.6.2, implementado na plataforma KNIME 3.2.2 (BERTHOLD et al., 2009) e integrado aos *softwares* ChemAxon Standardizer 15.10.26 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) e R versão 3.3.2. Resumidamente, hidrogênios explícitos

foram adicionados, sais removidos e quimiotipos específicos foram padronizados. Polímeros, compostos organometálicos e misturas também foram removidos. Além disso, foi realizada a análise e exclusão de duplicatas. Os critérios usados na análise de duplicatas foram: (i) se a atividades reportadas para as duplicatas estivessem em concordância, uma entrada era mantida no conjunto de dados e a outra excluída; (ii) se apresentassem discordância em suas atividades biológicas, ambas entradas eram excluídas. Tanto o conjunto de dados I quanto o II apresentaram elevada concordância entre as duplicatas (99% e 95%, respectivamente). Todos os compostos que apresentaram dados inconclusivos, ou seja, compostos sem curva dose-resposta e/ou apresentando operadores (> ou <) ambíguos, também foram excluídos.

4.1.1.3. Balanceamento dos conjuntos de dados

O conjunto de dados I não foi submetido a nenhum procedimento para balancear os dados, uma vez que o limiar de atividade utilizado foi suficiente para se obter um balanceamento satisfatório (2.035 inibidores e 2.013 não inibidores). Entretanto, como o conjunto de dados II estava bastante desbalanceado, foi aplicado o método de *under-sampling* linear, em que os compostos da classe majoritária (compostos não citotóxicos) tiveram seu número reduzido para se atingir o balançado desejado em relação à classe minoritária (compostos citotóxicos). Diferentemente de outros métodos de *under-sampling*, os quais realizam uma amostragem randômica para o balanceamento, o *under-sampling* linear retém a maioria das estruturas representativas da classe majoritária, permitindo uma alta cobertura do espaço químico original. Inicialmente, foi calculada a distância Euclidiana da matriz contendo os compostos da classe minoritária, representados por *fingerprints* moleculares MACCS (do inglês *Molecular Access System*), em relação aos *fingerprints* MACCS de cada composto da classe majoritária, utilizando o método de *k*NN (do inglês *k-neareast neighbors*) (ALTMAN, 1992). Em seguida, amostras da classe majoritária foram linearmente extraídas ao longo de toda faixa de *k*-distâncias (compostos representativos de todas as faixas de distância), compondo um conjunto final balanceado em relação à classe minoritária (1.012 compostos citotóxicos: 1.012 compostos não citotóxicos). Esse procedimento foi realizado em um *workflow* implementado na plataforma KNIME 3.2.2, integrado ao *software* R 3.3.2.

4.1.1.4. Geração e validação dos modelos de QSAR

Após o preparo, padronização e balanceamento, os conjuntos de dados foram usados para a geração de dois modelos de QSAR binários: (i) um modelo para predição da atividade contra *T. cruzi* intracelular; (ii) modelo para predição de citotoxicidade na célula hospedeira. Os modelos foram gerados e validados de acordo com as boas práticas de modelagem por QSAR descritas na literatura (CHERKASOV et al., 2014; TROPSHA, 2010). Detalhes sobre os modelos estão descritos nos próximos tópicos.

4.1.1.5. Fingerprints moleculares

Os descritores do tipo impressão digital molecular (*fingerprints*) Morgan, uma implementação do *software* RDKit para fingerprints ECFP4 (do inglês *standard-connectivity fringerprints*), utilizando raio igual a 2 e vetores contendo 2048 *bits*, foram calculados para todas as moléculas dos conjuntos de dados (RINIKER; LANDRUM, 2013a; ROGERS; HAHN, 2010). Os *fingerprints* ECFP4 foram calculados a partir do *software* aberto RDKit (http://www.rdkit.org), executado em Python 2.7 (https://www.python.org).

4.1.1.6. Método de aprendizado de máquina

O algoritmo RF (*Random Forest*) foi utilizado para geração dos modelos de QSAR. O RF é baseado na construção de um conjunto de diferentes árvores de decisão. Cada árvore é construída a partir de um subconjunto aleatório obtido a partir do conjunto de dados original. As predições de cada árvore são então combinadas para se formar a predição final (BREIMAN, 2001; SVETNIK et al., 2003). O algoritmo de RF foi executado em Python 2.7, utilizando o pacote scikit-learn 0.18 (http://scikit-learn.org).

4.1.1.7. Validação cruzada externa 5-fold

Os modelos de QSAR foram submetidos à validação cruzada externa *5-fold* (CHERKASOV et al., 2014). Esse procedimento pode ser descrito da seguinte forma: o conjunto de dados inteiro foi randomicamente dividido em 5 subconjuntos de tamanhos iguais. Em seguida, um dos subconjuntos (20% dos compostos) foi separado como conjunto de validação externa, enquanto os quatro subconjuntos restantes formaram o conjunto de modelagem (80% dos compostos), o qual foi usado para geração de um

modelo. Esse procedimento foi repetido cinco vezes, permitindo que cada um dos subconjuntos fosse usado uma vez como conjunto de validação externa. Portanto, todos os modelos foram gerados a partir de seus respectivos conjuntos de modelagem. Os conjuntos de validação externa não participaram da construção dos modelos.

4.1.1.8. Definição do domínio de aplicabilidade (DA)

O DA dos modelos foi calculado a partir da distância Euclidiana dos compostos a serem preditos em relação aos seus vizinhos mais próximos no conjunto treinamento de cada modelo. Essa distância foi então comparada a um limiar de distância previamente definido. Portanto, uma predição seria considerada menos confiável caso a distância calculada fosse maior que o limiar de distância. Nesse estudo, o limiar de distância (D_T) para cada composto predito e seu vizinho mais próximo no conjunto treinamento foi calculado conforme a seguinte equação:

$$\mathrm{D_T} = \bar{y} + \mathrm{Z}\sigma \qquad (\mathrm{Eq.}\,2)$$

Na equação 2, $\bar{y}$ corresponde à distância Euclidiana média em relação aos 5 vizinhos mais próximos do conjunto treinamento, σ representa o desvio padrão das distâncias e Z corresponde a um parâmetro arbitrário para controlar o nível de significância. Nesse trabalho, o valor de Z utilizado foi de 0,5.

4.1.1.9. Avaliação da performance dos modelos de QSAR

A sensibilidade (SE), especificidade (SP), taxa de classificação correta (CCR, do inglês *correct classification rate*), valor preditivo positivo (PPV, do inglês *positive predictive value*) e o valor preditivo negativo (NPV, do inglês *negative predictive value*) foram calculados para se estimar a preditividade dos modelos. Essas métricas foram calculadas conforme as equações 3 a 7, representadas a seguir:

$$\mathrm{SE} = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FN}} \qquad (\mathrm{Eq.}\,3)$$

$$\mathrm{SP} = \frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{FP}} \qquad (\mathrm{Eq.}\,4)$$

$$\mathrm{CCR} = \frac{\mathrm{SE} + \mathrm{SP}}{2} \qquad (\mathrm{Eq.}\,5)$$

$$\mathrm{PPV} = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FP}} \qquad (\mathrm{Eq.}\,6)$$

$$\mathrm{NPV} = \frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{FN}} \qquad (\mathrm{Eq.}\,7)$$

Nas equações 3 a 7, VP e VN correspondem, respectivamente, ao número de verdadeiros positivos (ex.: inibidores classificados corretamente) e verdadeiros negativos (ex.: não-inibidores classificados corretamente). FP e FN representam, respectivamente, o número de falsos positivos (ex.: compostos erroneamente classificados como inibidores) e falsos negativos (ex.: compostos erroneamente classificados como não-inibidores).

4.1.1.10. Triagem virtual

Os modelos de QSAR foram aplicados na triagem virtual da base de dados ChemBridge, contendo 1.063.926 compostos, com o objetivo de se identificar novos possíveis inibidores da replicação intracelular de *T. cruzi* com potencial de serem seletivos contra o parasito, ou seja, não citotóxicos em células de mamíferos. Antes da triagem, todos os compostos da base de dados foram preparados de acordo com o mesmo protocolo aplicado aos compostos usados na geração dos modelos de QSAR. Em seguida, foi utilizada a ferramenta FILTER v. 2.5, implementada no *software* OMEGA v.2.5.1 (OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) (HAWKINS et al., 2010), em que filtros *drug-like* foram aplicados, incluindo o filtro de Veber e a regra dos 5 de Lipinski (LIPINSKI et al., 2001; VEBER et al., 2002). Foi permitido aos compostos violarem no máximo uma das regras de Lipinski. Nessa etapa, também estavam inclusos filtros de solubilidade e de exclusão de agregadores conhecidos. Em seguida, os compostos restantes foram submetidos aos modelos de QSAR para predição de sua atividade biológica contra *T. cruzi* e citotoxicidade. Nessa etapa, os *hits* virtuais foram selecionados de acordo com os seguintes critérios: (i) os compostos deveriam ser preditos como inibidores da replicação intracelular de *T. cruzi* e como não citotóxicos, apresentando concordância

entre as árvores de decisão acima de 60%; (ii) os compostos deveriam estar dentro do DA dos modelos de QSAR. Ao final da triagem virtual, os *hits* virtuais selecionados foram adquiridos e submetidos à avaliação experimental *in vitro*.

4.1.2. Avaliação experimental

A avalição experimental seguiu os protocolos descritos por Moraes e colaboradores (2014) para as culturas *in vitro* das células de mamíferos e de parasitos, preparo dos compostos, avaliação da atividade antiparasitária e análise dos resultados (MORAES et al., 2014). Os experimentos foram realizados em colaboração com o Dr. Lucio H. Freitas-Junior e Dra. Carolina Borsoi Moraes (Laboratório Nacional de Biociências (LNBio) - Campinas-SP).

4.1.2.1. Cultura de células

As células de osteosarcoma humano U2OS e células epiteliais de rim de macaco LLC-MK2 foram cultivadas em meio DMEM alta glicose suplementado com 10% de soro fetal bovino (SFB) inativado por calor, 100 $U.mL^{-1}$ de penicilina e 100 $\mu g.mL^{-1}$ de estreptomicina e mantidas em estufa de atmosfera úmida de 5% de CO_2 a 37 ºC.

A cepa Y de *T. cruzi*, usada nesse estudo, foi doada pela Dra. Andrea R. Ávila (Instituto Carlos Chagas, Fiocruz, Curitiba-PR). As culturas axênicas de epimastigotas foram mantidas a 28°C em meio LIT contendo 10% de SFB e glicose a 2%. As epimastigotas em fase de crescimento exponencial foram coletadas, lavadas em PBS e ressuspendidas a $3x10^7$ parasitos/mL em meio Grace suplementado com 10% de SFB. A diferenciação de epimastigotas para tripomastigotas metacíclicos foi monitorada diariamente de 5 a 14 dias. Quando a diferenciação atingiu um platô (pelo menos 50% da população era composta por tripomastigotas metacíclicos), os parasitos foram coletados, lavados em PBS e transferidos para culturas de células LLC-MK2.

As culturas de LLC-MK2 foram usadas para manutenção do ciclo de *T. cruzi* (cepa Y) *in vitro*. As formas tripomastigotas, usadas para infectar as células U2OS, foram obtidas a partir do sobrenadante das culturas de LLC-MK2, enquanto as culturas de células infectadas U2OS foram mantidas em meio DMEM baixa glicose (Sigma-Aldrich) suplementado como descrito anteriormente para células de mamíferos, mas com 2% de SFB.

4.1.2.2. Preparo dos compostos

Todos os compostos foram adquiridos da empresa ChemBridge (http://www.chembridge.com) e diluídos em DMSO a 20 mM, enquanto o controle positivo, benznidazol, adquirido da Sigma-Aldrich, foi diluído a 40 mM. Todos compostos foram estocados a -80°C. As soluções-estoque dos compostos foram serialmente diluídas em DMSO, transferidas para placas intermediárias com DMSO a 6% e finalmente adicionadas a placas de 10 µL/poço para a concentração de 1% de DMSO, a qual demonstrou previamente não ter efeito aparente no crescimento e viabilidade dos parasitos.

4.1.2.3. Ensaios de atividade antiparasitária

Os ensaios foram realizados conforme descritos previamente na literatura (MORAES et al., 2014). Os compostos foram testados em ensaios de concentração-resposta (10 concentrações), com a concentração mais alta começando em 200 µM, exceto para o benznidazol, que foi iniciado em 400 µM. As culturas infectadas foram expostas aos compostos por 96 h. O experimento foi realizado em duplicata (dois experimentos independentes). Todas as placas incluíam múltiplos poços com infecção tratada apenas com 1% de DMSO, usados como controles negativos e poços não infectados (branco contendo apenas células U2OS, também tratadas com 1% de DMSO) como controles positivos. A análise de HCS (do inglês *high content analysis*) foi realizada no *High Content Imaging System Operetta* (PerkinElmer) com aumento de 20 vezes. As imagens foram analisadas para identificação, segmentação e quantificação do núcleo, citoplasma das células hospedeiras e dos parasitos intracelulares. Os dados foram analisados conforme descrito anteriormente na literatura (MORAES et al., 2014). A curvas concentração-resposta foram processadas no *software* GraphPad Prism v.6 para ajuste não linear das curvas e determinação dos valores de EC_{50} por interpolação. O EC_{50} foi definido como a concentração do composto correspondente à redução de 50% da infecção. O CC_{50} foi definido como a concentração do composto que reduz a quantidade de células pela metade, em comparação ao número de células nos poços de controle negativo. O índice de seletividade (SI) foi calculado de acordo com a razão entre o EC_{50} e o CC_{50} do composto ($SI = CC_{50}/ EC_{50}$).

4.1.3. Avaliação multi-paramétrica *in silico* dos compostos testados

Após a validação experimental, os compostos testados foram submetidos a uma avaliação multi-paramétrica *in silico* utilizando o módulo ADME QSAR do *software* StarDrop (Optibrium Ltd., Cambridge, UK). Um conjunto de propriedades físico-químicas e ADMET foram preditas por modelos de QSAR disponíveis no *software*. Essas propriedades incluíam: LogP (coeficiente de partição octanol/água); LogS (solubilidade em água); HIA (absorção intestinal humana, do inglês *human intestinal absorption*); susceptibilidade ao efluxo pela P-glicoproteína (P-gp); predição do bloqueio do canal de K^+ hERG (pIC_{50}); predição da afinidade com o citocromo P450 (isoformas CYP2C9 e CYP2D6); Predição da ligação a proteínas plasmáticas. Outras propriedades importantes também foram analisadas no *software*: área de superfície polar topológica (TPSA, do inglês *Topological Polar Surface Area*); número de aceptores e ligação de hidrogênio; número de doadores e ligação de hidrogênio; peso molecular. Além disso, a potência (EC_{50}) dos compostos contra a cepa Y de *T. cruzi*, a citotoxicidade em células U2OS (CC_{50}) e o índice de seletividade (SI), obtidos após a validação experimental, foram convertidos para escala logarítmica negativa e incorporados na avaliação multi-paramétrica.

4.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp.

4.2.1. Conjunto de dados

Foi utilizado um conjunto de dados contendo inibidores e não inibidores da enzima piruvato quinase (PK) de *L. mexicana*, depositado na base de dados *PubChem Bioassay* contendo 293.196 compostos com dados de concentração necessária para inibir 50% da atividade enzimática (IC_{50}) (PubChem Identifier: AID 1721; URL:*https://pubchem.ncbi.nlm.nih.gov/bioassay/1721*). Um limiar de 10 μM, conforme proposto por Katsuno e colaboradores para seleção de *hits* em ensaios biológicos contra *Leishmania*, foi usado para discriminação de inibidores ($IC_{50} \leq 10$ μM) e não inibidores ($IC_{50} > 10$ μM) (KATSUNO et al., 2015). Portanto, 309 compostos foram considerados inibidores e 290.884 compostos não inibidores de PK. Além disso, um total de 2.003 compostos com resultados inconclusivos foram descartados.

4.2.1.1. Preparo e padronização do conjunto de dados

Todas as estruturas químicas e suas informações de atividade biológica foram analisadas e preparadas de acordo com os protocolos desenvolvidos por Fourches e colaboradores (FOURCHES; MURATOV; TROPSHA, 2016, 2010, 2015). Todas as etapas de preparo e padronização foram executadas no pacote KSAR 1.6.2, implementado na plataforma KNIME 3.2.2 (BERTHOLD et al., 2009) e integrado aos *softwares* ChemAxon Standardizer 15.10.26 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) e R versão 3.3.2. Todas as etapas de preparo, padronização e análise de duplicatas foram realizadas de acordo com o mesmo protocolo descrito no item 4.1.1.2. A concordância entre as duplicatas desse conjunto de dados foi de 99.89%. O conjunto de dados, após o preparo, apresentava 307 inibidores e 288.777 não inibidores.

4.2.1.2. Balanceamento do conjunto de dados

Nesse estudo, dois níveis de balanceamento foram utilizados: balanceamento de 1:36 (307 inibidores : 11.052 não inibidores), o qual foi usado para validação dos modelos baseados na forma e volume moleculares (*shape-based*, LBDD) e do modelo baseado em *docking* (SBDD); balanceamento 1:1, que foi utilizado para geração e validação dos modelos de QSAR. Os balanceamentos foram executados por meio do método de *under-sampling* linear, conforme descrito anteriormente no item 4.1.1.3.

4.2.1.3. Preparo da estrutura tridimensional dos compostos

Antes de serem utilizados na validação dos modelos LBDD e SBDD, os compostos balanceados na proporção 1:36 foram preparados conforme descrito a seguir: (i) as estruturas tridimensionais (3D) dos compostos foram geradas e um número máximo de 200 confôrmeros foi calculado por molécula, utilizando o programa OMEGA v. 2.5.1 (OpenEye Scientific Software, Santa Fe, NM, http://www.eyesopen.com, HAWKINS et al., 2010). Esse programa gera uma série de conformações iniciais para cada composto, baseando-se em um banco de dados de fragmentos pré-calculados. Ao final, as estruturas são otimizadas utilizando-se o campo de força MMFF94 (HAWKINS et al., 2010); (ii) o estado de ionização dos compostos foi calculado em pH = 7,4, através da função '*fixpka'* disponível no programa QUACPAC v.1.6.3 (OpenEye Scientific Software, Santa Fe, NM,

http://www.eyesopen.com); (iii) o estado tautomérico mais estável foi calculado para todos os compostos, utilizando o mesmo programa; (iv) cargas AM1-BCC foram calculadas para todos os compostos, também utilizando o programa QUACPAC v.1.6.3.

4.2.2. Modelo baseado em *docking* (SBDD)

4.2.2.1. Preparo da proteína

A estrutura da PK de *L. mexicana* em complexo com o inibidor suramina, ligado ao sítio de ATP, foi extraída do *Protein Data Bank* (PDB ID: 3PP7, resolução: 2,35 Å) (MORGAN et al., 2011b). Em seguida, a estrutura da proteína foi importada para o programa Maestro v. 10.0 (Schrödinger, LLC, New York, NY, 2014) e preparada através da ferramenta *Protein Preparation Wizard*, conforme descrito a seguir: (i) átomos de hidrogênio foram adicionados (pH = 7,4 ± 0,5) de acordo com o programa Epik v. 2.7 (Schrödinger, LLC, New York, NY, 2014); (ii) foram definidos os estados de protonação e tautomérico dos resíduos em pH 7,4; (iii) as moléculas de água localizadas a uma distância superior a 5 Å do inibidor co-cristalizado ou realizando menos de 3 ligações de hidrogênio com o complexo proteína-ligante (águas não-estruturais) foram removidas; (iv) a estrutura do complexo ligante-enzima foi refinada através do campo de força OPLS-2005 (BANKS et al., 2005; SHELLEY et al., 2007). O estado de protonação do ligante co-cristalizado foi também definido através do programa Epik 2.7 (Schrödinger, LLC, New York, NY, 2014).

Ao final do preparo, a estrutura da proteína foi processada na ferramenta *Make Receptor*, disponível no pacote OEDocking v.3.0.1(OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com), em que foram definidas as coordenadas de *docking*, gerando-se uma caixa (*grid*) de dimensões 16.20 Å x 17.24 Å x 23.95 Å e volume de 6,692 $Å^3$, ao redor do inibidor co-cristalizado.

4.2.2.2. Docking

A estrutura do inibidor co-cristalizado (suramina) foi extraída do arquivo PDB e processada utilizando-se o mesmo protocolo aplicado aos inibidores e não-inibidores de PK, descrito no item 4.2.1.3. Em seguida, a suramina foi submetida à docagem molecular utilizando o programa FRED, disponível no pacote OEDocking v.3.0.1(OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com),

utilizando a opção de *docking* em alta precisão e a função de *score ChemGauss4*. O desvio médio quadrático (RMSD, do inglês *Root Mean Square Deviation*) entre a pose obtida no *docking* e a conformação co-cristalizada foi de 1,14 Å. Utilizando o mesmo protocolo, todos os confôrmeros dos 307 inibidores e 11.052 não-inibidores de PK, previamente preparados, foram submetidos ao *docking*. Ao final, a melhor pose de cada composto, ou seja, aquela com melhor *ChemGauss4 score*, foi armazenada em uma lista ordenada, a qual foi utilizada para o cálculo das métricas para avaliação do modelo SBDD. O cálculo das métricas foi executado no workflow *in house* "KSAR VS toolkits - metrics" implementado no programa KNIME v. 3.2.2.

4.2.3. Modelos baseados na forma e volume moleculares (modelos *shape-based*)

Os modelos baseados na forma e volume moleculares (modelos *shape-based*) foram gerados e validados no programa ROCS v. 3.2.1 (OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) (HAWKINS; SKILLMAN; NICHOLLS, 2007). A melhor pose de cada um dos 307 inbidores de PK, obtida no *docking*, foi submetida à opção *Ligand Model Builder*, disponível no programa ROCS. Dessa forma, foi realizado o alinhamento rígido de todas as poses e, para cada um dos inibidores, foi gerado um modelo hipotético, o qual foi avaliado de acordo com a função de pontuação *TanimotoCombo*. A função *TanimotoCombo* é uma combinação das funções *ShapeTanimoto* (compara as moléculas de acordo com a melhor sobreposição do volume molecular) e *ColorTanimoto* (relacionada à sobreposição apropriada de grupos que apresentam características como doadores e aceptores de ligação de hidrogênio, hidrofóbicos, cátions, ânions e anéis). O valor de *TanimotoCombo* pode variar de 0 a 2. Valores mais próximos de 2 indicam uma boa sobreposição das moléculas (HAWKINS; SKILLMAN; NICHOLLS, 2007). Portanto, os cinco melhores modelos *shape-based* hipotéticos, ou seja, aqueles com maiores valores de *TanimotoCombo*, foram selecionados para validação.

Na validação, todos os confôrmeros gerados para os compostos do conjunto de dados (307 inibidores e 11.052 não-inibidores) foram sobrepostos aos modelos. O melhor confôrmero de cada molécula foi selecionado de acordo com os valores de *TanimotoCombo* e, após esse processo, foi gerada uma lista ordenada com todas as moléculas (inibidores e não-inibidores), que foi submetida ao cálculo de métricas para avaliação dos modelos. O cálculo das métricas foi executado no workflow *in house*

"KSAR VS toolkits - metrics" implementado no programa KNIME v. 3.2.2. Ao final, o melhor modelo foi escolhido para ser usado posteriormente na triagem virtual.

4.2.4. Avaliação dos modelos SBDD e *shape-based*

A fim de se avaliar os modelos SBDD e *shape-based*, foram calculados os seguintes parâmetros: área sob a curva ROC (AUC), o fator de enriquecimento (EF), a discriminação aprimorada de Boltzmann (BEDROC), sensibilidade (SE) e especificidade (SP). Os dois últimos parâmetros foram descritos no item 4.1.1.9.

A análise ROC (do inglês *receiver operating characteristic*) é um método de avaliação de modelos amplamente utilizado, uma vez que fornece um resultado visual e numérico. A curva obtida nessa análise, conhecida como curva ROC, é construída a partir da variação de limiares (*thresholds*), apresentando no eixo *y* os valores de sensibilidade (taxa de verdadeiros positivos) e no eixo *x* as taxas de falsos positivos (1-especificidade), correspondentes a cada limiar (SONEGO; KOCSOR; PONGOR, 2008). A partir desses resultados, pode ser calculada a área sob a curva ROC (AUC), que é um parâmetro que avalia o poder de classificação global do modelo (Figura 19) (EMPEREUR-MOT et al., 2015).

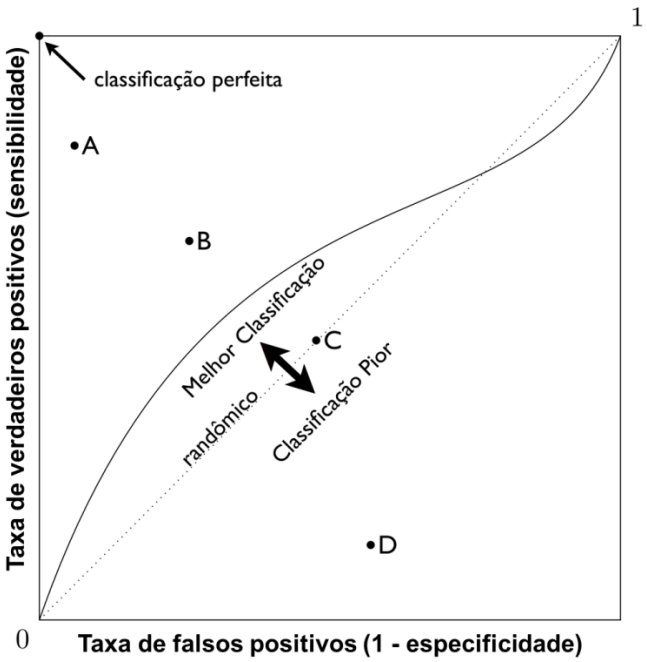


Figura 19 – Representação do desempenho de diferentes classificadores (pontos A, B, C e D) e da curva ROC (linha curva). O resultado da classificação A é melhor do que B,

C e D, pois apresenta maior sensibilidade e uma menor taxa de falsos positivos. A curva ROC para um classificador aleatório (C) é apresentada como uma linha pontilhada. Um classificador perfeito tem uma taxa de falsos positivos igual a zero e a sensibilidade (taxa de verdadeiros positivos) igual a um, conforme indicado pelo ponto no canto superior esquerdo do gráfico (Adaptado de BRAGA; ANDRADE, 2013).

Uma classificação perfeita corresponde ao ponto situado no canto superior esquerdo do gráfico, gerando uma curva com AUC = 1. Na prática, boas curvas se situam entre a linha diagonal pontilhada e a classificação perfeita, de modo que, quanto maior a distância da curva à linha diagonal e mais próxima do canto superior esquerdo, melhor é o modelo (BRAGA; ANDRADE, 2013). Entretanto, um valor alto de AUC não é garantia de que somente compostos ativos estarão no topo da lista ordenada, uma vez que essa métrica não é sensível à posição relativa dos compostos ativos e inativos situados no topo da lista (SONEGO; KOCSOR; PONGOR, 2008).

O fator de enriquecimento (EF) indica o quão melhor é o modelo em ordenar os compostos (ativos e inativos) quando comparado a uma lista ordenada aleatoriamente. O cálculo é feito a partir da fração de compostos ativos encontrados no topo da lista ordenada na seleção de *x%* do conjunto de dados, a qual é comparada com a fração de compostos ativos encontrados na lista ordenada contendo todos os compostos (ativos e inativos) (HECKER et al., 2002; JACOBSSON et al., 2003). O cálculo do EF pode ser definido pela equação 8, representada a seguir:

$$EF^{x\%} = \frac{Hits_{selecionado}^{x\%} / N_{selecionado}^{x\%}}{Hits_{total} / N_{total}} \qquad \text{(Eq. 8)}$$

Na equação 8, $Hits_{selecionado}^{x\%}$corresponde ao número de *hits* (compostos ativos) no percentual selecionado (*x%*) da lista ordenada dos compostos; $N_{selecionado}^{x\%}$corresponde ao número total de compostos no percentual selecionado (*x%*) da lista; $Hits_{total}$corresponde ao número total de *hits* em toda a lista ordenada; N_{total}corresponde ao número total de compostos da lista (BRAGA; ANDRADE, 2013; WEI et al., 2002).

A discriminação aprimorada de Boltzmann da curva ROC (BEDROC) é uma forma generalizada da AUC com decréscimo exponencial de uma função ponderada que foca em como os compostos ativos são ordenados no início da lista. Em outras palavras,

a BEDROC pode ser interpretada como a probabilidade de selecionar um composto ativo antes daqueles provenientes da seleção aleatória de compostos exponencialmente distribuídos para valores de α (TRUCHON; BAYLY, 2007). A BEDROC é calculada a partir da equação 9, representada a seguir:

$$BEDROC = RIE\ x\ \frac{R_\alpha \sinh(\alpha/2)}{\cosh\left(\frac{\alpha}{2}\right) - \cosh(\frac{\alpha}{2} - \alpha R_a)} + \frac{1}{1 - e^{\alpha(1-R_\alpha)}} \approx \frac{RIE}{\alpha} + \frac{1}{1 - e^{\alpha}} \quad (\text{Eq. } 9)$$

, se $\alpha R_a \ll 1$ e $\alpha \neq 0$

Na equação 9, *RIE* corresponde ao enriquecimento inicial robusto, R_a à fração de ativos e α ao parâmetro que controla a distribuição exponencial.

4.2.5. Modelos de QSAR

O conjunto de dados de inibidores e não-inibidores de PK previamente preparado, padronizado e balanceado (1:1) foi utilizado para a geração de modelos de QSAR binários (inibidores = 1; não-inibidores = 0) utilizando diferentes *fingerprints* moleculares. Os modelos foram gerados, validados e combinados em um modelo consenso, usado posteriormente na triagem virtual de potenciais inibidores da PK de *Leishmania*. Todos os modelos foram gerados e validados em workflows *in house* desenvolvidos na plataforma KNIME v. 3.2.2 (BERTHOLD et al., 2009), integrada com R *software* v. 3.3.2 e ao programa RDKit (http://www.rdkit.org).

4.2.5.1. Fingerprints moleculares

Foram calculados cinco *fingerprints* moleculares, disponíveis no RDKit: i) *fingerprints* Morgan (raio=2, vetor com 1024 *bits*), que são uma implementação do *software* RDKit para fingerprints ECFP4 (LANDRUM, 2017; ROGERS; HAHN, 2010); ii) *fingerprints* FeatMorgan (raio=2, vetor com 1024 *bits*), que são uma variação dos *fingerprints* Morgan, sendo baseados em características (*feature-based*), equivalentes aos fingerprints FCFP4 (*functional-class fingerprints*) (LANDRUM, 2017; ROGERS; HAHN, 2010); (iii) *fingerprints* MACCS (*Molecular Access System*) (ANDERSON, 1984; DILL et al., 1981; DURANT et al., 2002); (iv) Atom-Pairs, que são uma implementação em RDKit dos *atom pairs* de Carhart, com vetores de 1024 *bits*

(CARHART; SMITH; VENKATARAGHAVAN, 1985); (v) *fingerprints* Avalon, que são similares aos *fingerprints Daylight* (GEDECK; ROHDE; BARTELS, 2006).

4.2.5.2. Método de aprendizado de máquina

O algoritmo RF (*random forest*) foi aplicado para geração dos modelos de QSAR (BREIMAN, 2001; SVETNIK et al., 2003). O algoritmo de RF foi executado no programa R v. 3.3.2, integrado ao KNIME v. 3.2.2. Uma breve descrição desse algoritmo está disponível nos itens 1.3.1.3 e 4.1.1.6.

4.2.5.3. Validação cruzada externa 5-fold

Os modelos de QSAR foram submetidos à validação cruzada externa *5-fold* (CHERKASOV et al., 2014). O procedimento foi executado conforme descrito anteriormente no item 4.1.1.7.

4.2.5.4. Definição do DA

O DA dos modelos foi calculado a partir da distância Euclidiana dos compostos a serem preditos em relação aos seus vizinhos mais próximos no conjunto treinamento de cada modelo. Essa distância foi então comparada a um limiar de distância previamente definido. A definição do limiar de distância e os detalhes do cálculo estão disponíveis no item 4.1.1.8.

4.2.5.5. Modelos consenso

Após a geração de cinco modelos de QSAR, a partir de diferentes *fingerprints* moleculares e utilizando o algoritmo RF, os mesmos foram combinados em um modelo consenso. O modelo consenso foi obtido através do cálculo da média aritmética das predições de cada modelo individual (inibidor = 1; não-inibidor = 0). Consequentemente, a predição da atividade de cada novo composto pelo modelo consenso se encontrava na faixa de 0 a 1. Portanto, compostos com predição igual ou acima de 0,6 (ativos em pelo menos três dos cinco modelos que formavam o consenso) foram classificados como inibidores, enquanto compostos com predição igual ou abaixo de 0,4 (ativos no máximo em dois dos cinco modelos que formavam o consenso) foram classificados como não-inibidores. Quanto mais próxima de 0 ou 1 fosse a predição,

maior seria concordância entre os cinco modelos e maior a confiabilidade na classificação como não-inibidor ou inibidor.

4.2.5.6. Avaliação da robustez e performance dos modelos de QSAR

Os modelos de QSAR (individuais e consenso) foram avaliados através do cálculo das seguintes métricas, descritas anteriormente no item 4.1.1.9: sensibilidade (SE), especificidade (SP), taxa de classificação correta (CCR, do inglês *correct classification rate*). Além disso, foi utilizado o coeficiente *kappa* (*k*) para se estimar a concordância entre as predições dos modelos e os dados experimentais de atividade das moléculas do conjunto de dados (COHEN, 1960). Esse coeficiente é calculado a partir das equações 10, 11 e 12, apresentadas a seguir:

$$\mathrm{Pr}(a) = \frac{\mathrm{VP} + \mathrm{VN}}{\mathrm{N}} \qquad (\mathrm{Eq.}\,10)$$

$$\mathrm{Pr}(e) = \frac{(\mathrm{VP} + \mathrm{FP})\ \mathrm{x}\ (\mathrm{VP} + \mathrm{FN})\ \mathrm{x}\ (\mathrm{VN} + \mathrm{FN})\ \mathrm{x}\ (\mathrm{VN} + \mathrm{FP})}{\mathrm{N}} \qquad (\mathrm{Eq.}\,11)$$

$$k = \frac{\mathrm{Pr}(a) - \mathrm{Pr}(e)}{1 - \mathrm{Pr}(e)} \qquad (\mathrm{Eq.}\,12)$$

Nessas equações, VP e VN correspondem, respectivamente, ao número de verdadeiros positivos e verdadeiros negativos. FP e FN representam, respectivamente, o número de falsos positivos e falsos negativos. N corresponde ao número total de compostos. O Pr(a) corresponde à concordância relativa observada entre a predição (classificação) do modelo e a classificação conhecida experimentalmente. O Pr(e) representa a probabilidade hipotética de concordância ao acaso. O *k*, definido pela equação 12, pode apresentar valores entre -1,0 (ausência de concordância) e 1,0 (completa concordância). Valores entre 0,6 e 1,0 indicam modelos preditivos.

A robustez dos modelos foi avaliada através do teste de permutação da variável dependente Y (*Y-randomization*), executado em 10 rodadas, utilizando um protocolo *in house*, desenvolvido na plataforma KNIME. Este teste foi usado para se avaliar se a correlação entre a estrutura e a atividade ocorreu ou não ao acaso.

4.2.6. Triagem virtual

A base de dados comercial ChemBridge, contendo 1.063.926 compostos, foi submetida à triagem virtual de novos potenciais inibidores da PK de *Leishmania*. Antes da triagem, todos os compostos da base de dados foram preparados de acordo com o mesmo protocolo aplicado aos compostos usados na geração e validação dos modelos SBDD, *shape-based* e de QSAR (seções 4.2.1.1 e 4.2.1.3). Em seguida, foi utilizada a ferramenta FILTER v. 2.5, implementada no programa OMEGA v.2.5.1 (OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) (HAWKINS et al., 2010), em que filtros *drug-like* foram aplicados, incluindo o filtro de Veber e a regra dos 5 de Lipinski (LIPINSKI et al., 2001; VEBER et al., 2002). Foi permitido aos compostos violarem no máximo uma das regras de Lipinski. Nessa etapa, também estavam inclusos filtros de solubilidade e de exclusão de agregadores conhecidos. Após essa filtragem, os compostos restantes foram submetidos ao melhor modelo *shape-based*. Nesse passo, todos os compostos que apresentaram o *score TanimotoCombo* acima de 1,0 foram selecionados. Esses compostos foram então submetidos ao modelo SBDD, em que foi realizado o *docking* dos mesmos no sítio de ligação de ATP da PK de *L. mexicana*, o mesmo sítio de ligação do inibidor conhecido (suramina). Os compostos melhor ranqueados, baseando-se no *score ChemGauss4*, prosseguiram para a próxima etapa. Os compostos então foram submetidos ao modelo de QSAR consenso, para predição e seleção de potenciais inibidores da enzima PK. Nessa etapa, foram selecionados os compostos que: i) foram preditos como ativos em pelo menos 80% dos modelos de QSAR que compunham o consenso; ii) estavam dentro do DA de mais que 50% dos modelos de QSAR. Ao final da triagem, os *hits* virtuais selecionados foram adquiridos e submetidos à avaliação experimental.

4.2.7. Avaliação experimental

Os ensaios de atividade leishmanicida *in vitro* em formas promastigotas e amastigotas de *L. infantum* e de citotoxicidade em células de mamíferos foram realizados no Centro de Parasitologia e Micologia do Instituto Adolfo Lutz, São Paulo, sob a coordenação do Prof. Dr. André G. Tempone Cardoso.

Foram utilizados hamsters dourados (*Mesocricetus auratus*) para manutenção da cultura de *L.* (L.) *infantum* (MHOM/BR/1972/LD). Fêmeas jovens de camundongos BALB/c foram utilizadas como fonte de macrófagos peritoneais. Os procedimentos

envolvendo animais foram realizados de acordo com os "*Guidelines for the Care and Use of Laboratory Animals*" da *National Academy of Sciences*, USA. Os procedimentos foram executados com aprovação do Comitê de Ética em Pesquisa do Instituto Adolfo Lutz/Instituto Pasteur (CEUA-IAL/Pasteur 04/2016).

4.2.7.1. Culturas de L. infantum, macrófagos peritoneais e células NCTC

As promastigotas de *L.* (L.) *infantum* (MHOM/BR/1972/LD) foram mantidas em meio M-199 suplementado com 10% de soro fetal bovino (SFB) inativado por calor (Gibco, Thermo Fisher Scientific, São Paulo, Brasil) e 0,25% de hemina (Sigma-Aldrich) a 24 °C. As amastigotas foram obtidas por centrifugação diferencial do baço de hamsters previamente infectados. Macrófagos residentes foram coletados da cavidade peritoneal de camundongos BALB/c por meio de lavagem com meio RPMI-1640 suplementado com 10% de SFB, mantidos a 37 °C em estufa de CO_2 a 5%. Os fibrolastos NCTC (clone L929 ATCC) foram usados para avaliação da citotoxicidade dos compostos testados. As células NCTC foram mantidas em meio M-199 suplementado com 10% de SFB, 20 µg/mL de gentamicina e a 37 °C em estufa de CO_2 a 5%.

4.2.7.2. Avaliação da atividade contra L. infantum e citotoxicidade contra células de mamíferos

4.2.7.2.1. Avaliação da atividade em promastigotas

Os compostos foram dissolvidos em DMSO e serialmente diluídos em meio M-199 em placas de 96 poços, com a maior concentração de 150 µM. Promastigotas de *L.* (L.) *infantum* em fase de crescimento tardia (5-8 dias) foram adicionadas na quantidade de 1 x 10^6 por poço e incubadas com os compostos por 48 horas em estufa a 24 °C. Após a incubação, a viabilidade das células foi determinada pelo ensaio de MTT (TADA et al., 1986). Um grupo controle interno foi usado com 0,5% DMSO (concentração máxima). A miltefosina foi usada como fármaco padrão (controle positivo).

4.2.7.2.2. Avaliação da atividade em amastigotas intracelulares

Os macrófagos peritoneais de camundongos (residentes) foram adicionados na quantidade de 1 x 10^5 por poço em lâminas de 16 poços, que foram incubados por 24 horas em estufa a 37 °C com 5% de CO_2. Em seguida, os macrófagos foram infectados com amastigotas purificadas de baço de hamster infectado na proporção 1:10 (macrófagos:amastigotas), ou seja, 1 x 10^6 amastigotas por poço, e mantidos a 37 °C em estufa de CO_2 a 5% por 24 h. Posteriormente, os macrófagos infectados foram tratados com os compostos que foram seriamente diluídos e incubados por 120 horas sob as mesmas condições. Ao final do ensaio, os poços foram retirados e a lâmina foi fixada em metanol, corada com Giemsa (Merck KGaA, Darmstadt, Germany) e observada ao microscópio óptico. O IC_{50} foi determinado por meio da contagem de 500 macrófagos/poço, avaliando-se o número de macrófagos infectados (YARDLEY; CROFT, 2000). Como controle do ensaio (controle negativo), foram utilizados macrófagos infectados não tradados. O controle negativo foi utilizado para se avaliar a taxa de infecção dos macrófagos (n° de macrófagos infectados/500 macrófagos). A taxa de infecção no controle negativo deveria ser maior que 90% para que o ensaio fosse validado. Como controle positivo, foram utilizados macrófagos infectados tratados com miltefosina, que é um fármaco em uso clínico contra as leishmanioses.

4.2.7.2.3. Avaliação da citotoxicidade

As células NCTC foram adicionadas na quantidade de 6 x 10^4 em placas de 96 poços incubadas com os compostos por 48h. Um grupo controle interno foi usado com DMSO a 0,5% (máxima concentração). A quantificação de células viáveis foi determinada por ensaio de MTT (TADA et al., 1986). O índice de seletividade (SI) foi determinado a partir da seguinte equação: SI = CC_{50} em células NCTC / IC_{50} contra amastigotas.

4.2.7.3. Análise estatística

O processamento dos dados e análise estatística foram realizados no programa GraphPad Prism 5.0 (GraphPad Software, La Jolla, CA, USA). Para determinação da atividade leishmanicida e citotoxicidade em células NCTC, os valores de IC_{50} e CC_{50} foram calculados após normalização utilizando curvas dose-resposta sigmoides. A

avaliação da significância estatística foi feita pelo teste de Tukey. Os experimentos foram realizados em triplicata, em pelo menos dois ensaios independentes.

4.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp.

4.3.1. Conjuntos de dados

Foram utilizados três conjuntos de dados reportados na literatura:

- Conjunto de dados contendo inibidores e não inibidores da enzima piruvato quinase (PK) de *L. mexicana*: depositado na base de dados *PubChem Bioassay* (PubChem Identifier: AID 1721; URL: *https://pubchem.ncbi.nlm.nih.gov/bioassay/1721*). Os detalhes sobre esse conjunto de dados estão descritos no item 4.2.1.
- Conjunto de dados de inibidores da enzima CYP51: contendo compostos provenientes de um ensaio depositado na base dados ChEMBL (código de entrada CHEMBL2444665) (SURYADEVARA et al., 2013) e compostos obtidos na literatura (ANDRIANI et al., 2013; BUCKNER et al., 2012; CALVET et al., 2014; CHOI; PODUST; ROUSH, 2014; GUNATILLEKE et al., 2012; HARGROVE et al., 2012; KRAUS et al., 2010; SOEIRO et al., 2013; SURYADEVARA et al., 2009; VIEIRA et al., 2014). No total, 404 compostos com atividade reportada contra CYP51 foram obtidos. Todos os compostos apresentavam atividade expressa em EC_{50} e, baseando-se em um limiar de atividade de 10 μM, foram identificados 399 inibidores ($EC_{50} \leq 10$ μM) e 5 não-inibidores ($EC_{50} > 10$ μM). Apesar de não haver um valor específico definido na literatura para o limiar de atividade contra CYP51, a escolha se baseou no valor recomendado por Katsuno e colaboradores, que definem o limiar de 10 μM como critério para seleção de *hits* em ensaios biológicos contra *Leishmania* (KATSUNO et al., 2015).
- Conjunto de dados de compostos testados contra amastigotas de *L. infantum* (MHOM/MA/BE/67): depositado na base de dados ChEMBL (código de entrada ChEMBL2093840) e contendo 740 compostos avaliados *in vitro*. Baseando-se em um limiar de 10 μM ($IC_{50} \leq 10$ μM = ativo; $IC_{50} > 10$ μM =

inativo) proposto por Katsuno e colaboradores, esse conjunto de dados apresentava 220 compostos ativos e 520 inativos (KATSUNO et al., 2015). Esse conjunto de dados foi preparado, padronizado e balanceado para ser usado na geração de modelos de QSAR.

4.3.1.1. Preparo e padronização dos conjuntos de dados

Todas as estruturas químicas e suas informações de atividade biológica, dos três conjuntos de dados, foram analisadas e preparadas de acordo com os protocolos desenvolvidos por Fourches e colaboradores (FOURCHES; MURATOV; TROPSHA, 2016, 2010, 2015). Todas as etapas de preparo e padronização foram executadas no pacote KSAR 1.6.2, implementado na plataforma KNIME 3.2.2 (BERTHOLD et al., 2009) e integrado aos *softwares* ChemAxon Standardizer 15.10.26 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) e R versão 3.3.2. Todas as etapas de preparo, padronização e análise de duplicatas foram realizadas de acordo com o mesmo protocolo descrito no item 4.1.1.2.

Os detalhes sobre o conjunto de dados de inibidores de PK, após o seu preparo, foram descritos no item 4.2.1.1. O conjunto de dados de inibidores de CYP51 apresentou 281 compostos após a remoção de duplicatas, dos quais 276 eram inibidores da enzima (ativos) e 5 não-inibidores (inativos). A concordância entre as duplicatas desse conjunto de dados foi de 100%. O conjunto de dados de compostos testados contra amastigotas de *L. infantum* não apresentou duplicatas, porém um composto foi excluído por apresentar uma anotação ambígua, impossibilitando o seu uso na geração de modelos de QSAR. Ao final, esse conjunto apresentou 219 compostos ativos e 520 inativos.

4.3.1.2. Balanceamento dos conjuntos de dados

Os balanceamentos foram executados por meio do método de *under-sampling* linear, conforme descrito anteriormente no item 4.1.1.3. O conjunto de dados de inibidores de PK foi balanceado na proporção de 1:36, conforme descrito no item 4.2.1.2. Esse conjunto de dados foi usado para validação dos modelos *shape-based* descritos no item 4.2.3. O conjunto de dados de inibidores de CYP51, também usado para validação de modelos *shape-based*, apresentou um número muito superior de compostos ativos (276 ativos e apenas 5 inativos). Portanto, para esse conjunto de dados

foram gerados *decoys*, que consistiam em compostos com probabilidade de serem inativos. Os detalhes sobre a geração dos *decoys* estão descritos na próxima seção. O conjunto de dados de compostos testados contra *L. infantum*, que foi usado na geração de modelos de QSAR, foi balanceado na proporção de 1:1 (219 ativos e 219 inativos).

4.3.1.3. Geração de decoys para o conjunto de dados de inibidores de CYP51

Um conjunto de compostos supostamente inativos contra CYP51 (*decoys*) foi gerado a partir da base de dados ZINC "*drug-now"* contendo mais de 10 milhões de compostos com características *drug-like* (IRWIN; SHOICHET, 2005). O algoritmo de seleção de *decoys* foi utilizado a partir do protocolo *in-house* "KSAR VS toolkits – decoys", implementado no programa KNIME v. 3.2.2. Esse protocolo se baseia no mesmo protocolo utilizado para construir o diretório de *decoys* úteis (DUD, do inglês *directory of useful decoys*) (HUANG; SHOICHET; IRWIN, 2006). Os *fingerprints* MACCS e cinco propriedades físico-químicas foram utilizados para se calcular e avaliar a similaridade entre cada composto ativo do conjunto de dados de inibidores de CYP51 e seus respectivos potenciais *decoys*. Para cada composto ativo, foram selecionados 36 *decoys*.

Para serem considerados *decoys*, os compostos deveriam obedecer às seguintes características: i) similaridade aos compostos ativos de acordo com as seguintes propriedades físico-químicas: massa molecular (± 25 Da em relação ao composto ativo), número de ligações rotacionáveis (± 1), número de doadores de ligação de hidrogênio (± 1), número de aceptores de ligação de hidrogênio (± 2) e coeficiente de partição óleo-água (logP ± 1); ii) o coeficiente de Tanimoto entre um *decoy* e seu respectivo composto ativo não poderia ser maior que 0,75; iii) o coeficiente de Tanimoto entre um *decoy* e outro *decoy* previamente selecionado não poderia ser maior que 0,9.

Os *decoys* selecionados passaram pelas mesmas etapas de preparo e padronização que os compostos ativos contra CYP51.

4.3.1.4. Preparo da estrutura tridimensional dos inibidores de CYP51 e decoys

Antes de serem utilizados na validação dos modelos *shape-based*, os compostos ativos contra CYP51 e os *decoys*, balanceados na proporção 1:36, foram preparados seguindo o mesmo protocolo descrito no item 4.2.1.3, usado para o preparo das

estruturas 3D dos inibidores e não-inibidores de PK (cálculo de confôrmeros, estado de protonação, tautômeros e cargas atômicas parciais).

4.3.2. Modelos baseados na forma e volume moleculares (*shape-based*)

Os modelos baseados na forma e volume moleculares (modelos *shape-based*) foram gerados e validados no programa ROCS v. 3.2.1 (OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) (HAWKINS; SKILLMAN; NICHOLLS, 2007). Os modelos *shape-based* gerados a partir de inibidores conhecidos da enzima PK e a validação dos mesmos foi descrita em detalhes no item 4.2.3.

Os modelos *shape-based* baseados em inibidores conhecidos da enzima CYP51 foram gerados seguindo-se os seguintes passos: i) pesquisa no PDB de inibidores co-cristalizados com a enzima CYP51 de tripanossomatídeos; ii) geração dos modelos *shape-based* a partir da conformação co-cristalizada desses inibidores; iii) validação dos modelos.

Os inibidores de CYP51 de tripanossomatídeos selecionados para a geração dos modelos foram: VNF (PDB ID: 3KSW) ; VNT (PDB ID: 4G3J); VNI (PDB ID:4G7G); fluconazol, sendo este último o único inibidor co-cristalizado com a enzima CYP51 de um parasito do gênero *Leishmania* (co-cristalizado com a CYP51 de *L. infantum*; PDB ID: 3L4D).

Na validação, semelhantemente ao protocolo descrito para os modelos *shape-based* para PK (item 4.2.3), todos os confôrmeros gerados para os compostos do conjunto de dados de CYP51 (ativos e *decoys*) foram sobrepostos aos modelos. O melhor confôrmero de cada molécula foi selecionado de acordo com os valores de *TanimotoCombo* e, após esse processo, foi gerada uma lista ordenada com todas as moléculas (ativos e *decoys*), que foi submetida ao cálculo de métricas para avaliação dos modelos. O cálculo das métricas foi executado no workflow *in house* "KSAR VS toolkits - metrics" implementado no programa KNIME v. 3.2.2.

Ao final, o melhor modelo *shape-based* para enzima PK, juntamente ao melhor modelo para a enzima CYP51, foram selecionados e aplicados na triagem virtual *dual-target*.

4.3.3. Modelos de QSAR para predição da atividade contra amastigotas de *L. infantum*

O conjunto de dados de contendo compostos ativos e inativos contra amastigotas de *L. infantum*, previamente preparado, padronizado e balanceado (1:1) foi utilizado para a geração de modelos de QSAR binários (ativos = 1; inativos = 0) utilizando diferentes descritores moleculares. Os modelos foram gerados, validados e combinados em um modelo consenso, usado posteriormente na última etapa da triagem virtual. Todos os modelos foram gerados e validados em *workflows in-house* desenvolvidos na plataforma KNIME v. 3.2.2 (BERTHOLD et al., 2009), integrada ao *software* R v. 3.3.2 e ao programa RDKit (http://www.rdkit.org).

4.3.3.1. Descritores moleculares

Foram calculados cinco descritores moleculares: i) *fingerprints* Morgan (raio=2, vetor com 1024 bits), que são uma implementação do software RDKit para *fingerprints* ECFP4 (LANDRUM, 2017; ROGERS; HAHN, 2010); ii) *fingerprints* FeatMorgan (raio=2, vetor com 1024 bits), equivalentes aos *fingerprints* FCFP4 (functional-class fingerprints) (LANDRUM, 2017; ROGERS; HAHN, 2010); (iii) Atom-Pairs, que são uma implementação em RDKit dos *atom pairs* de Carhart, com vetores de 1024 bits (CARHART; SMITH; VENKATARAGHAVAN, 1985); (iv) *fingerprints* Avalon, que são similares aos *fingerprints Daylight* (GEDECK; ROHDE; BARTELS, 2006); (v) descritores CDK (*Chemistry Development Kit*), que constituem uma variedade de descritores de diversas classes como topológicos, constitucionais, eletrônicos, híbridos e geométricos. A matriz de descritores foi normalizada. Foram excluídos, antes da geração dos modelos de QSAR, todos descritores altamente correlacionados ($r > 0,9$) e aqueles que apresentavam valores constantes (BEISKEN et al., 2013; WILLIGHAGEN et al., 2017).

4.3.3.2. Método de aprendizado de máquina

O algoritmo RF (*random forest*) foi aplicado para geração dos modelos de QSAR (BREIMAN, 2001; SVETNIK et al., 2003). O algoritmo de RF foi executado no programa R v. 3.3.2, integrado ao KNIME v. 3.2.2. Uma breve descrição desse algoritmo está disponível nos itens 1.4.1.3 e 4.1.1.6.

4.3.3.3. Validação cruzada externa 5-fold

Os modelos de QSAR foram submetidos à validação cruzada externa *5-fold* (CHERKASOV et al., 2014). O procedimento foi executado conforme descrito anteriormente no item 4.1.1.7.

4.3.3.4. Definição do DA

O DA dos modelos foi calculado a partir da distância Euclidiana dos compostos a serem preditos em relação aos seus vizinhos mais próximos no conjunto treinamento de cada modelo. Essa distância foi então comparada a um limiar de distância previamente definido. A definição do limiar de distância e os detalhes do cálculo estão disponíveis no item 4.1.1.8.

4.3.3.5. Modelos consenso

Após a geração de cinco modelos de QSAR, a partir de diferentes descritores moleculares e utilizando o algoritmo RF, os mesmos foram combinados em um modelo consenso. O modelo consenso foi obtido através do cálculo da média aritmética das predições de cada modelo individual (ativo = 1; inativo = 0). Consequentemente, a predição da atividade de cada novo composto pelo modelo consenso se encontrava na faixa de 0 a 1. Portanto, compostos com predição igual ou acima de 0,6 (ativos em pelo menos três dos cinco modelos que formavam o consenso) foram classificados como ativos, enquanto compostos com predição igual ou abaixo de 0,4 (ativos no máximo em dois dos cinco modelos que formavam o consenso) foram classificados como inativos. Quanto mais próxima de 0 ou 1 fosse a predição, maior seria concordância entre os cinco modelos e maior a confiabilidade na classificação como inativo ou ativo.

4.3.3.6. Avaliação da robustez e performance dos modelos de QSAR

Os modelos de QSAR (individuais e consenso) foram avaliados através do cálculo das seguintes métricas, descritas anteriormente no item 4.1.1.9: sensibilidade (SE), especificidade (SP), taxa de classificação correta (CCR, do inglês *correct classification rate*). Além disso, foi utilizado o coeficiente *kappa* (*k*) para se estimar a concordância entre as predições dos modelos e os dados experimentais de atividade das moléculas do conjunto de dados (COHEN, 1960). O coeficiente *k* foi descrito anteriormente no item 4.2.5.6. A robustez dos modelos foi avaliada através do teste de

permutação da variável dependente Y (*Y-randomization*), executado em 10 rodadas, utilizando um protocolo *in house*, desenvolvido na plataforma KNIME. Este teste foi usado para se avaliar se a correlação entre a estrutura e a atividade ocorreu ou não ao acaso.

4.3.4. Triagem virtual

A base de dados comercial ChemBridge, contendo 1.063.926 compostos, foi submetida à triagem virtual *dual-target* de novos potenciais inibidores das enzimas PK e CYP51 de *Leishmania*. Antes da triagem, todos os compostos da base de dados foram preparados de acordo com o mesmo protocolo aplicado aos compostos usados na geração e validação dos modelos *shape-based* e de QSAR (seções 4.3.1.1 e 4.3.1.4). Em seguida, foi utilizada a ferramenta FILTER v. 2.5, implementada no programa OMEGA v.2.5.1 (OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) (HAWKINS et al., 2010), em que filtros *drug-like* foram aplicados, incluindo o filtro de Veber e a regra dos 5 de Lipinski (LIPINSKI et al., 2001; VEBER et al., 2002). Foi permitido aos compostos violarem no máximo uma das regras de Lipinski. Nessa etapa, também estavam inclusos filtros de solubilidade e de exclusão de agregadores conhecidos.

Após a filtragem de propriedades *drug-like*, os compostos restantes foram submetidos ao melhor modelo *shape-based* correspondente a cada enzima. Após a sobreposição aos modelos *shape-based*, foram geradas duas listas ordenadas de acordo com o *score TanimotoCombo*, cada uma correspondente a uma enzima. Em seguida, os compostos melhor ranqueados de cada lista (*top* 10% de cada lista) foram selecionados e formaram duas listas menores, as quais foram comparadas entre si, permitindo a seleção dos compostos presentes em ambas. Portanto, os compostos melhor ranqueados (*top* 10%) concomitantemente, nos dois modelos *shape-based* (PK e CYP51), foram selecionados.

Os potenciais *hits dual-target* foram então submetidos ao modelo de QSAR consenso, usado para predição da atividade contra a forma amastigota de *L. infantum*. Nessa etapa, foram selecionados os compostos que: i) foram preditos como ativos em pelo menos 80% dos modelos de QSAR que compunham o consenso; ii) estavam dentro do DA de mais que 50% dos modelos de QSAR. Ao final da filtragem por QSAR, foram selecionados potenciais *hits dual-target* e com potencial de serem ativos contra

amastigotas de *L. infantum*. Após a filtragem por QSAR, os *hits* virtuais selecionados foram adquiridos e submetidos à validação experimental.

4.3.5. *Docking* dos *hits* virtuais nas enzimas PK e CYP51

4.3.5.1. Preparo das proteínas

O preparo da estrutura da enzima PK (PDB ID: 3PP7, resolução: 2,35 Å) (MORGAN et al., 2011b) foi realizado conforme descrito anteriormente no item 4.2.2.1.

A estrutura da CYP51 de *L. infantum*, em complexo com o inibidor fluconazol (PDB ID: 3L4D, resolução 2,75 Å) foi preparada através da ferramenta *Protein Preparation Wizard*, implementada no programa Maestro v. 10.0 (Schrödinger, LCC, New York, 2014). Os passos seguidos no preparo foram: (i) átomos de hidrogênio foram adicionados (pH = 7,4 ± 0,5) de acordo com o programa Epik v. 2.7 (Schrödinger, LLC, New York, NY, 2014); (ii) foram definidos os estados de protonação e tautomérico dos resíduos em pH 7,4; (iii) as moléculas de água localizadas a uma distância superior a 5 Å do inibidor co-cristalizado ou realizando menos de 3 ligações de hidrogênio com o complexo proteína-ligante (águas não-estruturais) foram removidas; (iv) a estrutura do complexo ligante-enzima foi refinada através do campo de força OPLS-2005 (BANKS et al., 2005; SHELLEY et al., 2007). O estado de protonação do ligante co-cristalizado foi também definido através do programa Epik 2.7 (Schrödinger, LLC, New York, NY, 2014). Ao final do preparo, a estrutura da proteína foi processada na ferramenta *Make Receptor*, disponível no pacote OEDocking v.3.0.1(OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com), em que foram definidas as coordenadas de *docking*, gerando-se uma caixa (*grid*) de dimensões 18,00 Å x 18,17 Å x 18,33 Å ao redor do inibidor co-cristalizado.

4.3.5.2. Preparo dos ligantes

Foram utilizadas as estruturas 3D dos *hits* previamente preparadas de acordo com o procolo descrito no item 4.2.1.3 (cálculo de até 200 confôrmeros por ligante; estado de protonação em pH 7,4; tautômeros; cargas AM1-BCC).

4.3.5.3. Docking

Os potenciais *hits* selecionados ao final da triagem virtual foram submetidos a estudos de *docking* nas enzimas PK e CYP51, utilizando o programa FRED, disponível no pacote OEDocking v.3.0.1(OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com), utilizando a opção de *docking* em alta precisão e a função de *score ChemGauss4*. A validação do protocolo de *docking* foi realizada através da docagem molecular do ligante co-cristalizado, preparado previamente de acordo com o mesmo protocolo aplicado aos demais ligantes. O RMSD entre a pose obtida no *docking* e a conformação co-cristalizada do fluconazol foi de 1,66 Å.

5. RESULTADOS

5.1. Identificação de novos compostos bioativos contra *Trypanosoma cruzi* utilizando triagem virtual baseada em QSAR

Artigo:

Discovery of New Potent Hits Against intracellular *Trypanosoma cruzi* by QSAR-Based Virtual Screening

Autores: Cleber C. Melo-Filho, Rodolpho C. Braga, Eugene N. Muratov, Caio Haddad Franco, Carolina B. Moraes, Lucio H. Freitas-Junior & Carolina Horta Andrade

(Artigo submetido à revista *European Journal of Medicinal Chemistry – under review*)

Discovery of New Potent Hits Against intracellular *Trypanosoma cruzi* by QSAR-Based Virtual Screening

Cleber C. Melo-Filho[1,£], Rodolpho C. Braga[1,£], Eugene N. Muratov[2,3], Caio Haddad Franco[4], Carolina B. Moraes[4,5], Lucio H. Freitas-Junior[4,5] & Carolina Horta Andrade[1]*

[1] LabMol - Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmacia, Universidade Federal de Goiás - UFG, Rua 240, Qd.87, Goiania, GO 74605-510, Brazil;

[2] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA;

[3] Department of Chemical Technology, Odessa National Polytechnic University, 1. Shevchenko Ave., Odessa, 65000, Ukraine;

[4]National Laboratory of Biosciences (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, SP 13083-970, Brazil.

[5]Deparment of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP 05508-900, Brazil.

[£]*These authors have equally contributed to this work.*

*Corresponding Author: Tel: + 55 62 3209 6451; Fax: +55 62 3209 6037; E-mail: carolina@ufg.br

Abstract

Chagas disease is a neglected tropical disease (NTD) caused by the protozoan parasite *Trypanosoma cruzi* and is primarily transmitted to humans by the feces of infected Triatominae insects during their blood meal. The disease affects 6-8 million people, mostly in Latin America countries, and kills more people in the region each year than any other parasite-born disease, including malaria. Moreover, patient numbers are currently increasing in non-endemic, developed countries, such as Australia, Japan, Canada, and the United States. The treatment is limited to one drug, benznidazole, which is only effective in the acute phase of the disease and is very toxic. Thus, there is an urgent need to develop new, safer, and effective drugs against the chronic phase of Chagas disease. Using a QSAR-based virtual screening followed by *in vitro* experimental evaluation, we report herein the identification of novel potent and selective hits against *T. cruzi* intracellular stage. We developed and validated binary QSAR models for prediction of anti-trypanosomal activity and cytotoxicity against mammalian cells using the best practices for QSAR modeling. These models were then used for virtual screening of a commercial database, leading to the identification of 39 virtual hits. Further *in vitro* assays showed that seven compounds were potent against intracellular *T. cruzi* at submicromolar concentrations (EC_{50} < 1 μM) and were very selective (SI > 30). Furthermore, other six compounds were also inside the hit criteria for Chagas disease, which presented activity at low micromolar concentrations (EC_{50} < 10 μM) against intracellular *T. cruzi* and were also selective (SI > 15). Moreover, we performed a multi-parameter analysis for the comparison of tested compounds regarding their balance between potency, selectivity, and predicted ADMET properties. In the next studies, the most promising compounds will be submitted to additional *in vitro* and *in vivo* assays in acute model of Chagas disease, and can be further optimized for the development of new promising drug candidates against this important yet neglected disease.

Keywords: Chagas disease; QSAR; virtual screening; high content screening; validated hits; ADMET.

1. Introduction

Chagas disease, or American trypanosomiasis, is caused by the flagellate kinetoplastid protozoan parasite *Trypanosoma cruzi* and belongs to the group of neglected tropical diseases (NTDs), affecting mostly low-income populations of developing countries [1,2]. According to the World Health Organization, 6 to 8 million people are infected with *T. cruzi* in 21 endemic Latin America countries, with more than 70 million people at risk. Moreover, the disease causes about 12,000 deaths annually, killing more people in the region each year than any other parasite-born disease, including malaria [3]. Surprisingly, the number of infected patients is currently increasing in non-endemic, developed countries, such as Australia, Japan, Canada and the United States, mainly due to the increased migration of infected individuals from endemic areas [4,5]. The disease is primarily transmitted to humans by the feces of infected Triatominae insects, also known as "the kissing bugs", during their blood meal [6]. Other routes of transmission such as ingestion of contaminated food, congenital transmission, blood transfusion, and organ transplantation have also been reported [7–10]. Chagas disease can lead to life-long morbidity associated to cardiomyopathy and gastrointestinal problems, and then to death [5].

The treatment is extremely limited to only one drug, benznidazole, which is only effective at the acute phase of the disease and present limitations regarding its toxicity and efficacy in chronic infections in adults [11,12]. Thus, the development of new effective and safer drugs to treat the chronic phase of Chagas disease is urgently needed [13]. In this context, the integration of computer-assisted drug design (CADD) approaches with experimental validation can be a valuable strategy for the discovery of new drug candidates against Chagas disease, as observed with other NTDs [14–17]. Among various approaches of CADD, we highlight here the quantitative structure-activity relationships (QSAR), which have been extensively used over the past years for virtual screening (VS) and for lead optimization [18–20]. Previous studies reported by our group have also demonstrated the potential of QSAR for filtering large databases, leading to the discovery of potent *in vitro* antiparasitic compounds [15,16].

Therefore, the main goal of this study was the discovery of both potent and selective hits against intracellular *T. cruzi* using QSAR-based VS approach integrated with experimental evaluation. We have

generated and rigorously validated binary QSAR models for the prediction of anti-trypanosome activity and cytotoxicity against mammalian cells. These models were used for a VS of ChemBridge chemical database of 1M compounds to identify novel promising compounds. Then, the selected virtual hits were purchased and submitted to *in vitro* evaluation against Y strain of *T. cruzi* and cytotoxicity assays against human osteosarcoma cell line (U2OS), using High Content Screening (HCS). Finally, all tested compounds were submitted to an *in silico* multi-parameter analysis, aiming to highlight those with better balance between potency, selectivity, and predicted ADMET properties, to move forward to the *in vivo* studies.

2. Materials and Methods

The critical steps of this work are summarized in Figure 1. All recommended best practices of QSAR modeling and CADD were used in our workflow [18,21]. The main steps can be described as follows: (i) data compilation and integration; (ii) data curation; (iii) QSAR model generation and validation; (iv) virtual screening (VS) and selection of compounds; (v) *in vitro* experimental evaluation against intracellular *T. cruzi* Y strain and human osteosarcoma cell line (U2OS) using HCS; (vi) multi-parameter analysis of the tested compounds.

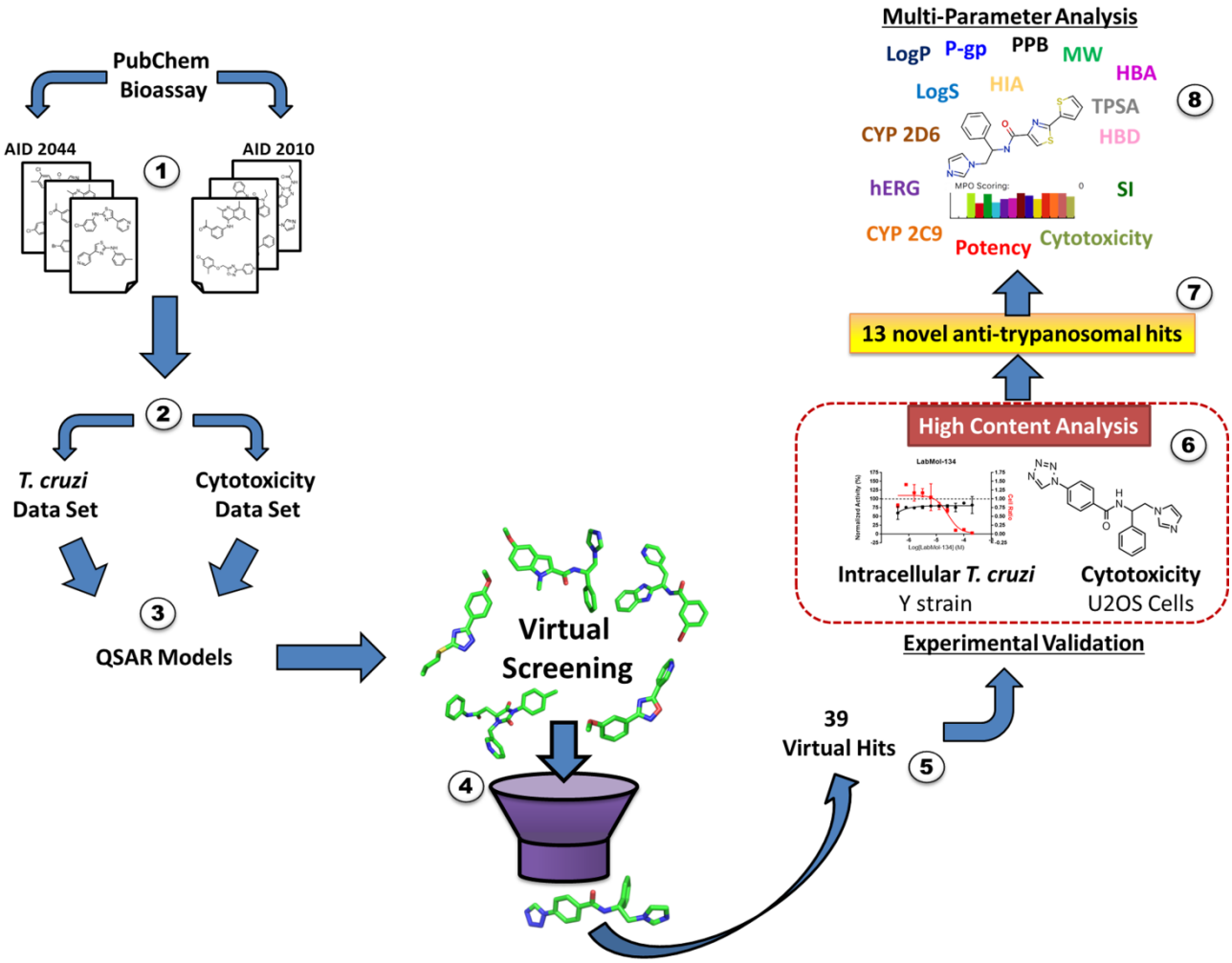


Fig. 1. General workflow of this study. (1) Data compilation and integration; (2) Data curation; (3) Generation and validation of QSAR models; (4) VS of ChemBridge database; (5) Final selected virtual hits; (6) *in vitro* experimental evaluation using HCS; (7) 13 hits were identified after experimental validation; (8) Multi-parameter analysis of tested compounds.

2.1. Computational

2.1.1. Data sets

In this study, we used two data sets extracted from PubChem Bioassay, containing results of high-throughput screening (HTS) confirmatory experiments performed at The Broad Institute of MIT and Harvard University [22,23]. A brief description of the data sets is presented below.

97 ***Data set I.*** Data set of compounds tested for inhibition of intracellular *T. cruzi* replication (PubChem
98 AID: 2044). Based on a threshold of 1.15 μM, it consisted of 2,044 compounds with $EC_{50} \leq 1.15$ μM
99 (inhibitors), 2,016 compounds with $EC_{50} > 1.15$ μM (non-inhibitors), and five compounds with
100 inconclusive results.
101 ***Data set II.*** Data set containing all compounds from data set I (PubChem AID: 2010), which were
102 submitted to a counter screen to evaluate the cytotoxicity against NIH/3T3 cells (mouse embryonic
103 fibroblast cells), the same cell line used as host cell for the *T. cruzi* replication assays. Based on a 50 μM
104 threshold, the data set presented 1,024 compounds with $EC_{50} \leq 50$ μM (cytotoxic), 2,487 compounds
105 with $EC_{50} > 50$ μM (non-cytotoxic), and 554 compounds with inconclusive results.
106
107 *2.1.2. Data curation*
108 All chemical structures and correspondent activity information were analyzed according to the
109 data curation protocols proposed by Fourches et al. [24–26]. All steps of data curation were performed in
110 our in-house KSAR v.1.6.2 workflow [15,27], implemented on KNIME 3.2.2 [28], integrated with
111 ChemAxon Standardizer v.15.10.26 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) and
112 R software v.3.3.2 [29]. In summary, explicit hydrogens were added, salts were removed, and specific
113 chemotypes were normalized. Polymers, inorganic salts, organometallic compounds, and mixtures were
114 also removed. Furthermore, we performed the analysis and exclusion of duplicates. The criteria for
115 exclusion of duplicates were: (i) if the reported activity of the duplicates were the same (i.e., in
116 concordance), only one entry will be retained in the dataset; (ii) if duplicates presented discordance in
117 biological activity, all entries will be excluded. Further analysis showed high concordance between
118 duplicate records for data set I (99%), and data set II (95%). Additionally, the inconclusive records, i.e.,
119 compounds without dose-response curve, and/or presenting ambiguous operators (> or <), were
120 removed.

121 *2.1.3. Data set balancing*

122 After data curation, the data set I (*T. cruzi* data set) was not submitted to any procedure for
123 balancing, as the defined threshold, *per se*, was sufficient for data balancing (2,035 inhibitors and 2,013
124 non-inhibitors). However, as data set II (cytotoxicity data set) was very unbalanced, a linear under-
125 sampling method was used. Briefly, the majority class (non-cytotoxic) was reduced to match the desired
126 balancing to the minority class (cytotoxic). In contrast to the traditional under-sampling methods which
127 randomly balance the data set, the linear under-sampling approach retains most of the representative
128 structures of the majority class, thus ensuring a high coverage of original chemical space. The basic
129 principle of this approach is to measure the Euclidian distance of the matrix containing the minority
130 class compounds, represented by the MACCS key fingerprints, to the MACCS key fingerprints of each
131 compound of the majority class, using a *k*-nearest neighbors (*k*NN) method [30]. Then, the samples of
132 majority class are linearly extracted over the whole range of *k*-distances, and used to generate the
133 balanced data sets. This approach was executed in KNIME v.3.2.2 [28], integrated with R software
134 v.3.3.2 [29]. After data curation and balancing, data set II presented 1,012 cytotoxic and 1,012 non-
135 cytotoxic compounds.

136

137 *2.1.4. QSAR model generation and validation*

138 The data sets were used for generation of two binary QSAR models: (i) a model for prediction of
139 intracellular *T. cruzi* inhibition (*T. cruzi* model); and (ii) a model for prediction of cytotoxicity
140 (cytotoxicity model). The models were generated and validated according to the best practices for QSAR
141 modeling [18,21]. Details about model generation are described below.

142 *2.1.5. Molecular fingerprints*

143 The Morgan fingerprints, a RDKit implementation [31] of the extended-connectivity fingerprints
144 (ECFP4) [32], with radius 2 and bit vector of 2,048 bits, was used for QSAR modeling. The ECFP4
145 fingerprints were calculated in the open-source cheminformatics software RDKit (http://www.rdkit.org)
146 [31] executed on Python 2.7 (https://www.python.org) [33].

147 *2.1.6. Machine learning method*

148 In this study, the random forest (RF) algorithm was used for model generation. The RF is based
149 on the construction of a set of different decision trees. Each tree is based on a random subset from the
150 original dataset. The predictions of each decision tree are combined to form the final prediction [34].
151 The RF classifier was executed on Python 2.7, using the open-source package scikit-learn 0.18
152 (http://scikit-learn.org) [35].

153 *2.1.7. 5-fold external cross-validation*

154 According to the best practices of QSAR modeling [18], we have chosen five-fold external cross-
155 validation for the estimation of predictivity of developed models. The procedure can be described as
156 follows: the entire data set of compounds with known experimental activities was randomly divided into
157 five subsets of equal size; then one of these subsets (20% of all compounds) was set aside as an external
158 validation set, and the remaining four sets together formed the modeling set (80% of the full set). This
159 procedure was repeated five times allowing each of the five subsets to be used as external validation set.
160 All models were built using the modeling set only, and the compounds in momentary external set (fold)
161 were not used to build or select the models.

162

163 *2.1.8. Applicability Domain (AD)*

164 The AD was estimated based on the Euclidean distances among the training set of each QSAR
165 model generated in the external 5-fold cross-validation procedure. The distance of a test set compound to
166 its nearest neighbor in the training set was compared to the predefined AD threshold level. The
167 prediction was considered to be less reliable if the distance was greater than the threshold level [36]. In
168 our study, the AD was defined as a distance threshold D_T between a compound under prediction and the
169 closest nearest neighbors in training set. The following equation was used for calculation of distance
170 threshold:

$$D_T = \bar{y} + Z\sigma \quad (1),$$

Where $\bar{y}$ is the average Euclidean distance of the k nearest neighbors of each compound from the training set, σ represents the standard deviation of the Euclidian distances, and Z is an arbitrary parameter to control the significance level. The parameter Z was set to a default value of 0.5. Therefore, if the distance of an external compound to all nearest neighbors in the training set is greater than this threshold, the prediction is considered unreliable.

2.1.9. Robustness and performance of QSAR models

Sensitivity (SE), specificity (SP), correct classification rate (CCR), positive predictive value (PPV), and negative predictive value (NPV) statistical metrics were used to estimate the predictive performance of the QSAR models. These metrics were calculated as follows:

$$SE = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$CCR = \frac{SE + SP}{2} \quad (4)$$

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

$$NPV = \frac{TN}{TN + FN} \quad (6)$$

In these equations, TP and TN correspond to the number of true positives (correct classifications of inhibitors) and true negatives (correct classifications of non-inhibitors), respectively, while FP and FN

189 represent the number of false positives (compounds incorrectly classified as inhibitors) and false
190 negatives (compounds incorrectly classified as non-inhibitors), respectively.
191
192 *2.1.10. Virtual Screening*
193 The generated and validated QSAR models were used for virtual screening of a commercial
194 database, ChemBridge (http://www.chembridge.com), containing 1,063,926 compounds, aiming to
195 identify new potential inhibitors of intracellular *T. cruzi* replication, which could be potentially selective
196 against the parasite, i.e., non-toxic to mammalian cells. Prior to screening, all compounds were prepared
197 using the same protocol previously applied for compounds used for QSAR modeling (for more details,
198 see Data curation section). Then, using the FILTER application v.2.5 [37] implemented in OMEGA
199 v.2.5.1 [38,39], some drug-likeness filters were applied, such as Veber filter [40] and Lipinski's rule of
200 five [41]. The compounds were allowed to violate only one of the Lipinski's rules. Other filters like
201 solubility and exclusion of known experimental aggregators were also applied. Then, the remaining
202 compounds had their activity against intracellular *T. cruzi* replication and cytotoxicity against
203 mammalian cells predicted by QSAR models. In the next step, all compounds predicted as inhibitors of
204 *T. cruzi* intracellular replication and as non-toxic to mammalian cells were selected and inspected. At
205 this point, compounds predicted as active and non-cytotoxic, but presenting concordance (agreement)
206 among decision trees below 60% were excluded. Additionally, to evaluate the structural novelty of the
207 potential hits, we calculated pairwise Tanimoto coefficients (using MACCS keys fingerprints) between
208 each virtual hit and its nearest neighbor from the full data set of known *T. cruzi* inhibitors (data set I).
209 Finally, the selected virtual hits were purchased and submitted to *in vitro* experimental evaluation.
210
211 **2.2. Experimental evaluation**
212 The experimental evaluation followed the protocols described by Moraes et al. [42] for *in vitro*
213 mammalian and parasite cell cultures, compound preparation, evaluation of antiparasitic activity, and
214 data analysis.

215

216 *2.2.1. In vitro* culture of cells

217 The human osteosarcoma cell line U2OS and the monkey kidney epithelial cell LLC-MK2 were
218 cultured in DMEM high-glucose medium supplemented with 10% heat-inactivated fetal bovine serum,
219 100 U/mL penicillin and 100 µg/mL streptomycin in a humid atmosphere of 5% CO_2 at 37 ºC. LLC-
220 MK2 tissue cultures were used to support the mammalian cycle of *T. cruzi in vitro*. Trypomastigote
221 forms were obtained from the supernatant of LLC-MK2 tissue cultures and used to infect U2OS cells.
222 All infected cultures, containing intracellular parasites, were maintained in low glucose DMEM media
223 (Sigma-Aldrich) supplemented as described for mammalian cells, but with 2% FBS.

224

225 *2.2.2. Compound preparation*

226 All compounds were purchased from ChemBridge (http://www.chembridge.com) and dissolved
227 in DMSO at 20 mM. Benznidazole (standard drug) was purchased from Sigma-Aldrich and prepared at
228 40 mM.

229 *2.2.1. Anti-trypanosomal activity assays*

230 Assays were performed as described previously [42]. The tested compounds were assayed in
231 dose-response (10 concentration points – 2-fold dilution), with the highest concentration starting at 200
232 µM, except for benznidazole, which started at 400 µM. Infected cultures were exposed to compounds for
233 96 h. The experiment was performed in duplicates (i.e., two independent experiments). All plates
234 included multiple wells of infection treated only with 1% DMSO as negative controls and non-infected
235 wells (blank containing only U2OS cells, also treated with 1% DMSO) as positive controls. High
236 Content Analysis was performed with the High Content Imaging System Operetta (PerkinElmer) at 20X
237 magnification. Images were analyzed for identification, segmentation and quantification of host cell
238 nucleus, cytoplasm and intracellular parasites. High Content data was analyzed as described previously
239 [42]. Concentration-response curves were processed with the Graphpad Prism software, version 6, for

generation of sigmoidal dose-response (variable slope) nonlinear curve fitting and determination of EC_{50} values by interpolation. The EC_{50} was defined as the compound concentration corresponding to 50% reduction of infection (50% normalized activity). The CC_{50} is defined as the compound concentration that reduced the cell ratio by half, when compared to the average number of cells in the negative control wells. Selectivity index (SI) was calculated based on the ratio between the compound EC_{50} and CC_{50} values ($SI = CC_{50}/EC_{50}$).

2.3. Multi-parameter analysis

After the experimental evaluation, all tested compounds were submitted to a multi-parameter analysis, using the ADME QSAR module of StarDrop software (Optibrium Ltd., Cambridge, UK) [43]. A set of important ADMET and physicochemical properties were predicted by the software built-in QSAR models. These properties included: LogP (Octanol/Water); LogS (aqueous solubility); HIA (Human Intestinal Absorption); P-glycoprotein (P-gp) efflux liability; Prediction of pIC_{50} for hERG K^+ channel blockage; Prediction of affinity to Cytochrome P450 (CYP2C9 and CYP2D6 isoforms); Plasma Protein Binding (PPB) estimation. The software also provided other important properties, such as: Topological Polar Surface Area (TPSA); Hydrogen Bond Acceptors (HBA); Hydrogen Bond Donors (HBD); and Molecular Weight (MW). Furthermore, the potency against *T. cruzi* Y strain (EC_{50}), cytotoxicity in U2OS cells (CC_{50}), and the selectivity index (SI), obtained after experimental evaluation, were converted to the negative logarithmic scale, and incorporated in the multi-parameter analysis. The list of parameters and their respective level of importance, which was manually set before the analysis, are presented in Figure S3, Supplementary Material.

3. Results and Discussion

3.1. QSAR models

263 ECFP4 molecular fingerprints (Morgan fingerprints) [31,32] and RF machine learning classifier were
264 used for the generation of the binary QSAR models: (i) model for prediction of intracellular *T. cruzi*
265 inhibition (*T. cruzi* model); (ii) model for prediction of cytotoxicity against mammalian host cells
266 (cytotoxicity model). The statistical characteristics of the best models, using 5-fold external cross-
267 validation, are summarized in Table 1. The correct classification rate (CCR) of *T. cruzi* and cytotoxicity
268 models was 0.74 and 0.69, respectively. The use of the applicability domain (AD) resulted in increase of
269 CCR by 1-2% (0.76 and 0.70, respectively), but at the expense of coverage (0.62 and 0.64, respectively).
270 The predictive positive and negative values (PPV and NPV) were calculated to estimate the probability
271 of an accurate prediction of a new compound as inhibitor/non-inhibitor, and as cytotoxic/non-cytotoxic.
272 As observed in Table 1, the models presented PPV of 0.76 and 0.69, with an increase of 4-5% when AD
273 was considered (0.81 and 0.73). The NPV was 0.73 and 0.68, with a slight decrease on *T. cruzi* model
274 (0.73 to 0.71) when considering the AD. The NPV of cytotoxicity model was not affected by the use of
275 AD. The *T. cruzi* model correctly predicted 72% of the inhibitors, and 77% of non-inhibitors (SE=0.72;
276 SP=0.77); while the cytotoxicity model correctly predicted 67% of cytotoxic, and 70% of non-cytotoxic
277 compounds (SE=0.67; SP=0.70). Both models had increased SE (0.72 – 0.78, and 0.67 – 0.72) when AD
278 was applied. However, the SP slightly decreased (0.77 – 0.74, and 0.70 – 0.69) when AD was used.
279 Additional information about model's performance can be found in Figures S1 and S2 from
280 Supplementary Material.

281 **Table 1.** Statistical characteristics of best developed QSAR models after 5-fold external cross-
282 validation.

Model	SE	PPV	SP	NPV	CCR	Coverage
T. cruzi model	0.72	0.76	0.77	0.73	0.74	1.00
T. cruzi model - AD	0.78	0.81	0.74	0.71	0.76	0.62
Cytotoxicity model	0.67	0.69	0.70	0.68	0.69	1.00
Cytotoxicity model - AD	0.72	0.73	0.69	0.68	0.70	0.64

283 SE: sensitivity; PPV: positive predictive value; SP: specificity; NPV: negative predictive value;
284 CCR: correct classification rate; AD: applicability domain.

285

286 *3.2. Virtual Screening*

287 The results of VS are summarized in Figure 2. The entire ChemBridge database containing
288 1,063,926 compounds was screened for identification of potentially active compounds against
289 intracellular *T. cruzi*, and non-cytotoxic to mammalian cells. Firstly, drug-likeness filters were applied
290 (Veber filter and Lipinski's rule of five). Other filters like aqueous solubility and exclusion of
291 experimentally confirmed aggregators were also used. After these steps of filtering, 338,511 compounds
292 were excluded. In the next step, the 725,415 remaining compounds were submitted to QSAR models for
293 prediction of the activity against intracellular *T. cruzi* and the cytotoxicity against mammalian cells.
294 After this step, 158 virtual screening hits were visually inspected to exclude compounds with potential
295 unwanted effects, undesired structural moieties, or other issues. This resulted in the final choice of 39
296 hits for subsequent experimental validation (Table S1). The criteria for selection of virtual hits can be
297 summarized as follows: i) the compounds should be predicted as active (inhibitor of intracellular *T.*
298 *cruzi*) and non-cytotoxic; ii) the predictions should be accepted (considered reliable) when the
299 agreement between decision trees was higher than 60%; iii) The compounds should be inside the AD of
300 the QSAR models. Additionally, a similarity analysis between the virtual hits and their nearest neighbors
301 on data set I (*T. cruzi* data set) was executed to evaluate their structural novelty. The structure of the
302 virtual hits and the similarity analysis results are presented on Table S1, Supplementary Material.

303 As observed on Table S1, four virtual hits presented Tanimoto coefficient (Tc) between 0.66-
304 0.70 in comparison to their nearest neighbors from data set I (*LabMol-115*, *-96*, *-135*, and *-111*). These
305 compounds could be a potential source of new scaffolds with activity against *T. cruzi*. Other 29 virtual
306 hits presented Tc between 0.71-0.85, and despite being more similar to already known active
307 compounds, they could represent new potential hits against *T. cruzi*. The six remaining virtual hits
308 presented high similarity to their nearest neighbors (Tc between 0.86-0.94).

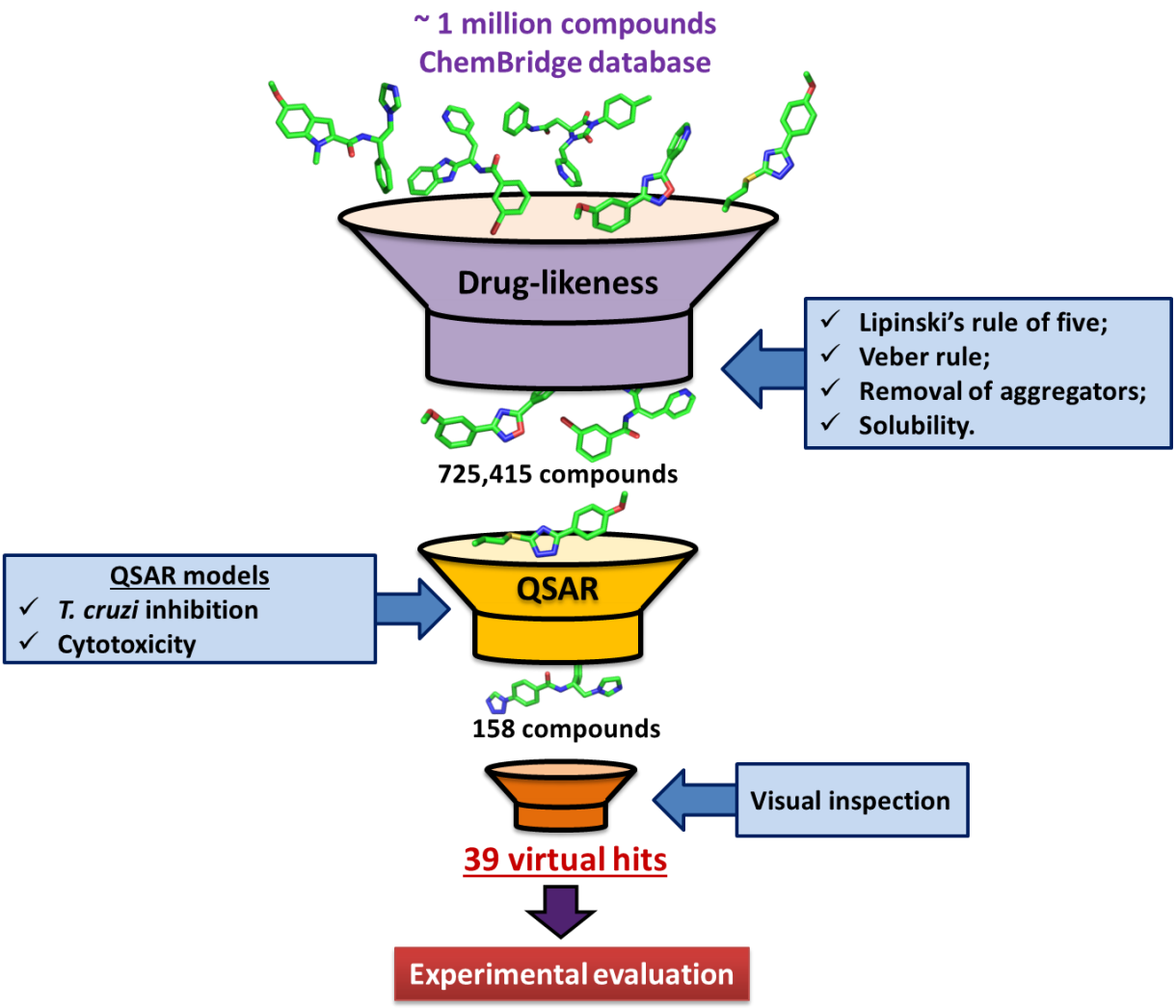


Fig. 2. Workflow summarizing the steps of VS. At the end, 39 virtual hits of *T. cruzi* were selected and submitted to experimental evaluation.

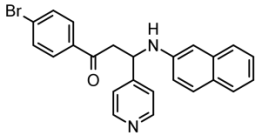
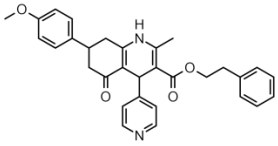
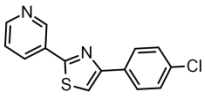
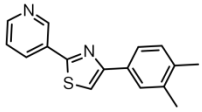
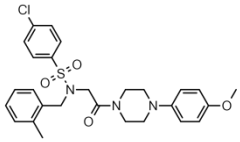
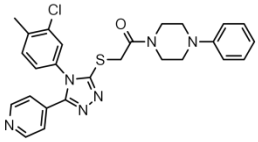
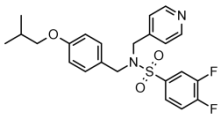
3.3. Anti-trypanosomal activity assays

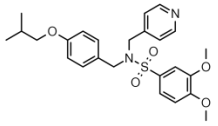
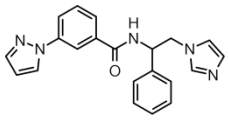
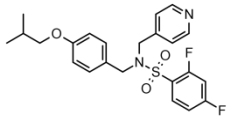
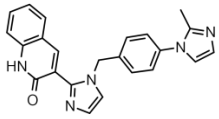
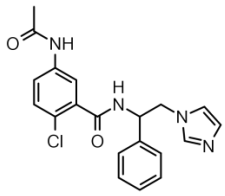
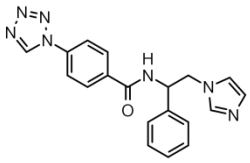
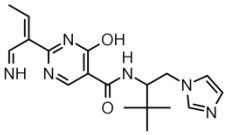
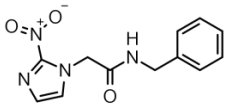
The selected 39 virtual hits were tested against the Y strain of *T. cruzi*, infecting human osteosarcoma cell line (U2OS), using a HCS assay. The best results of the experimental evaluation against *T. cruzi*, cytotoxicity in U2OS cells, and selectivity index (SI) are presented in Table 2. For results of all tested compounds, please refer to Table S2 (Supplementary Material).

Seven compounds presented high potency against intracellular *T. cruzi* with EC_{50} values in the submicromolar range (0.3-0.8 μM). Furthermore, *LabMol-129* presented EC_{50} very close to 1 μM. (EC_{50} = 1.2 μM). Other six compounds showed promising results (EC_{50} varying from 2.7 to 7.0 μM). It is

important to highlight that 11 compounds presented higher potency than the standard drug benznidazole (EC_{50} = 3.1 μM), the only drug currently in use for Chagas disease. In general, the compounds presented good selectivity against the parasite, i.e., 13 compounds with SI higher than 15. Among them, two compounds (*LabMol-96* and *Labmol-128*) were more potent and showed better SI than benznidazole. Compounds *LabMol-102* and *LabMol-103* presented very similar activity against the parasite, if we consider the standard deviation of the experiments. However, the presence of a chlorine atom (*LabMol-102*) instead of two methyl groups (*LabMol-103*), in the benzene ring, may be responsible for the high reduction in cytotoxicity against U2OS cells. Further comparisons of potency, selectivity, and other predicted ADMET properties, between the compounds, will be presented in the next section.

Table 2. List of the best validated hits against intracellular Y strain of *T. cruzi* and human U2OS cells.

Compound	Structure	EC_{50} *T. cruzi* (μM)	CC_{50} U2OS cells (μM)	SI
LabMol-96		**0.8 ± 0.1**	**> 200**	**> 258**
LabMol-97		**0.4 ± 0.1**	**19.3 ± 10.6**	**46.5**
LabMol-102		4.2 ± 1.0	> 200	> 48
LabMol-103		2.7 ± 0.5	40.8 ± 5.4	15.1
LabMol-104		2.7 ± 0.4	73.9*	27.7
LabMol-110		7.0 ± 1.4	21.1 ± 7.2	3.0
LabMol-125		**0.4 ± 0.02**	**11.5 ± 10.0**	**30.6**

LabMol-126		**0.5 ± 0.02**	**21.7 ± 3.4**	**45.7**
LabMol-127		**0.3 ± 0.2**	**11.8 ± 1.3**	**46.3**
LabMol-128		**0.7 ± 0.1**	**> 200**	**> 274**
LabMol-129		1.2 ± 0.1	52.9 ± 5.1	43.9
LabMol-132		2.7 ± 0.3	> 200	> 72.8
LabMol-134		**0.3 ± 0.2**	**33.6 ± 7.7**	**136.8**
LabMol-135		5.8 ± 0.7	> 200	> 34.6
benznidazole		3.1 ± 0.6	> 200	> 127.6

*Obtained in a single experiment; EC_{50}: compound concentration that causes a reduction of 50% in infection, compared to infected controls; CC_{50}: compound concentration that causes reduction in 50% in the number of human osteosarcoma cells; SI: selectivity index. All compounds presenting EC_{50} below 1 µM are highlighted in bold font.

3.4. Multi-parameter analysis of tested compounds

All tested compounds were submitted to an *in silico* multi-parameter analysis, using the ADME QSAR module of StarDrop software. The potency against *T. cruzi* Y strain (EC_{50}), cytotoxicity in U2OS cells (CC_{50}), and the selectivity index (SI) obtained after the experiments were combined to a set of ADMET and physicochemical properties predicted by the models available in the software. The compounds were manually divided in six different groups (clusters) for the analysis: (i) sulfonamides;

343 (ii) benzylbenzamides; (iii) sulfanylacetamides; (iv) oxadiazoles; (v) tetrahydroquinolines; and (vi)
344 thiazoles. All the clusters were analyzed and discussed separately (see below).
345
346 *3.4.1. Sulfonamides*
347 In this cluster (Figure 3), *LabMol-128* presented the best balance between potency against *T.*
348 *cruzi* (pEC_{50} = 6.14), cytotoxicity against U2OS cells (pCC_{50} = 3.70) and selectivity (SI >274).
349 However, *LabMol-126* had the best balance between all analyzed properties, reflected in the best multi-
350 parameter optimization scoring (MPO scoring = 0.08). *LabMol*-126 and *LabMol-128* have very similar
351 structures, but present some differences in the benzene ring, linked to the sulfone group. While *LabMol-*
352 *126* presents two methoxy groups on *meta* and *para* positions, *LabMol-128* has two fluorine atoms on
353 *ortho* and *para* positions. Our medicinal chemistry expertise pinpoint that a benzomethoxy moiety is a
354 well-known group responsible to reduce the metabolic stability. In this way, new analogs will be studied
355 and presented latter on. Compounds *LabMol-124* and *LabMol-101* had the worst performances of the
356 sulfonamides series presenting no selectivity (MPO scoring of zero). Despite the high structural
357 similarity between *LabMol-128*and *LabMol-124*, the presence of a *para* fluorine at the benzene moiety
358 of *LabMol-128* led to a significant increase in potency (from pEC_{50} = 3.70 to 6.14), thus placing this
359 compound among the most potent and selective hits. As demonstrated in the predicted probability maps
360 from Figure S6 of Supplementary Material, the *para* fluorine has a positive contribution to biological
361 activity. Moreover, all compounds were predicted as susceptible to P-gp (P-glycoprotein) transport, and
362 this could impair their intestinal absorption. In general, the compounds were also predicted to have an
363 intermediate affinity to hERG channels and CYP2D6 and CYP2C9. It is important to mention that these
364 ADMET properties were predicted by *in silico* models, and further experimental validation is needed.

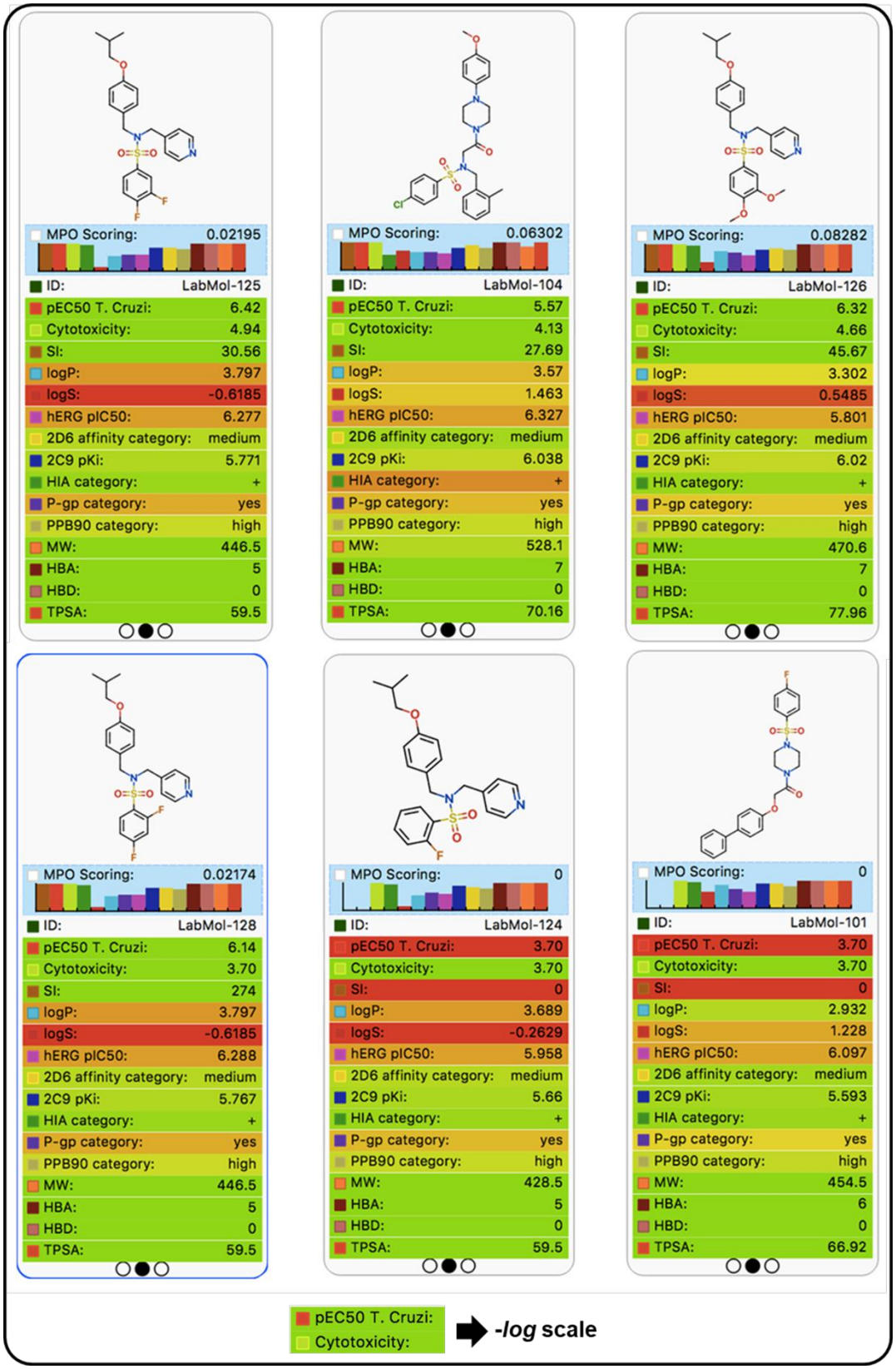


Fig. 3. Multi-parameter analysis of sulfonamides. Colors near green indicate properties with appropriate values, and colors near red indicate properties that need improvement. Intermediate results are presented between yellow and orange colors.

3.4.2. Benzylbenzamides

Compounds *LabMol-127* and *LabMol-134* presented the best MPO scoring (0.30 and 0.34, respectively, Figure 4). Both compounds had very close performance in terms of potency and selectivity. However, *LabMol-134* performed slightly better as it presented superior physicochemical properties and ADMET profile. Compounds *LabMol-123* and *LabMol-133* presented a poor performance in the multi-parameter analysis and they also presented lower potency and selectivity against *T. cruzi*. Compound *LabMol-134* was predicted as liable to P-gp transport and it presented an intermediate predicted affinity to hERG, CYP2D6, and CYP2C9. An additional analysis of this cluster, based on the predicted probability maps, is presented on figure S7, Supplementary Material.

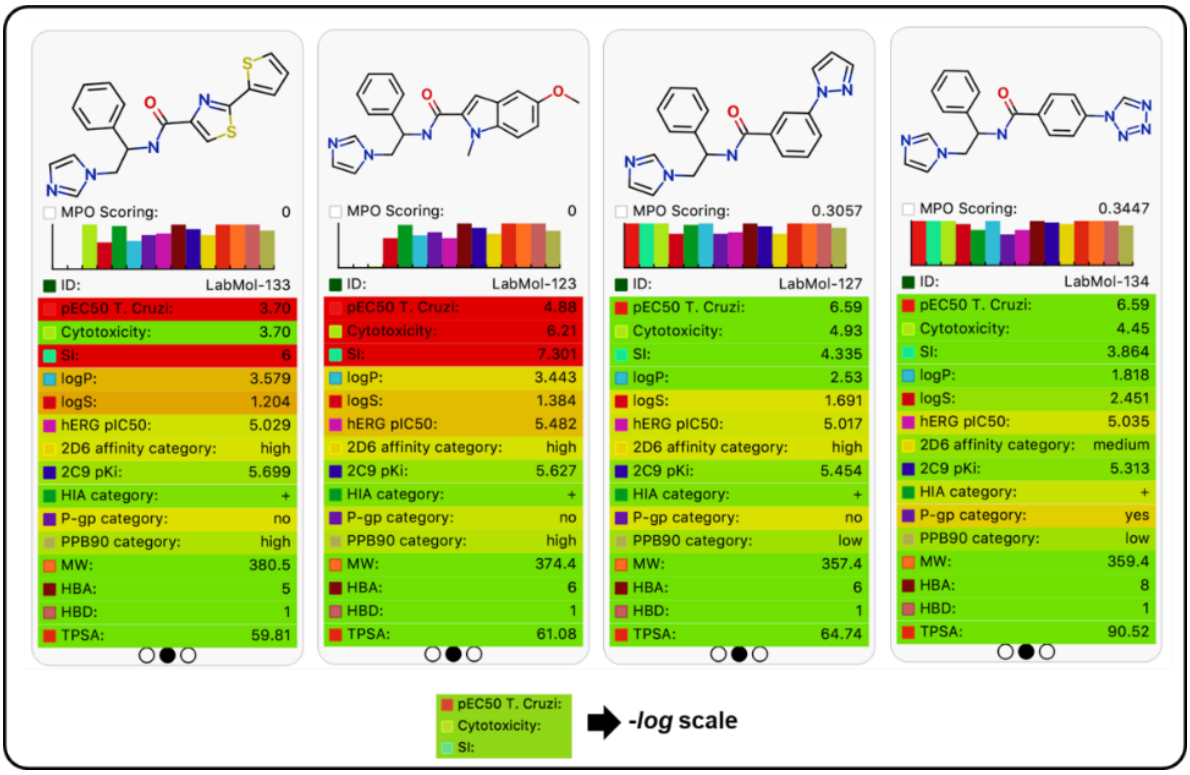


Fig. 4. Multi-parameter analysis of benzylbenzamides. Colors near green indicate properties with adequate values, and colors near red indicate properties that need improvement. Intermediate results are presented between yellow and orange colors.

385

386 *3.4.3. Sulfanylacetamides*

387 In this cluster, all compounds presented low MPO scoring (Figure 5). In general, the compounds in this
388 group presented very low potency against *T. cruzi* (Figure S4 and Table S2, Supplementary Material).
389 The only exception was *LabMol-110* (pEC_{50} = 5.2, Figure S4; EC_{50} = 7.0 ± 1.4 μM, Table 2), that
390 showed a moderate potency. However, all compounds of the *sulfanylacetamides* group presented low
391 selectivity. Two subgroups were observed inside the cluster: compounds containing the indoline group
392 (*LabMol-114* and *LabMol-113*); and compounds with piperazine ring (*LabMol-107, -110, -117, -118).* In
393 indoline subgroup, *LabMol-113*, which contains a triazole ring linked to a methyl group, presented
394 higher potency than *LabMol-114*, which contains a tetrazole ring at the same position (Figure S8,
395 Supplementary Material). In terms of predicted ADMET properties, *LabMol-114* performed slightly
396 better than *LabMol-113*. In piperazine subgroup, *LabMol-110* was the most potent compound, and
397 showed a benzene ring containing methyl and chlorine at *para* and *meta* positions, respectively, linked
398 to the triazole ring. The predicted probability maps (Figure S8, Supplementary Material) highlight the
399 importance of those fragments to biological activity. For *LabMol-117*, the absence of an additional ring
400 attached to the triazole ring may be responsible for the decrease of the potency (Figure S8,
401 Supplementary Material), but it improved the predicted ADMET profile in comparison to *LabMol-110*
402 (Figure S4, Supplementary Material).

403

404 *3.4.4. Oxadiazoles*

405 Three compounds presenting the oxadiazole ring were compared in the multi-parameter analysis
406 (*LabMol-111, -112* and *-115*, Figure S5A, Supplementary Material). All compounds in this group
407 presented a low MPO scoring and they presented very low potency and selectivity against intracellular
408 *T. cruzi* on experiments (Figure S5A, and Table S2, Supplementary Material). Compound *LabMol-111*
409 presented better selectivity and balance of predicted ADMET properties, with low predicted affinity to
410 hERG channel and CYP2C9. It also presented no liability for P-gp transport and low probability for

411 plasma protein binding (PPB). An additional analysis of the oxadiazoles, based on the predicted
412 probability maps, is presented on figure S9, Supplementary Material.
413
414 *3.4.5. Tetrahydroquinolines*
415 In this group of compounds, *LabMol-97* presented higher potency (pEC_{50} = 6.38), good
416 selectivity against *T. cruzi*, and MPO scoring of 0.01. Meanwhile, *LabMol-98*, which is very similar to
417 *LabMol-97*, presented low potency and selectivity (Figure S5B, Supplementary Material). The only
418 difference between these two compounds is the presence of a methoxy group in *LabMol-97*. Despite the
419 high increase in potency and selectivity after addition of the methoxy group, no improvement on the
420 predicted ADMET properties was observed. Further studies must be executed to improve the ADMET
421 profile of *LabMol-97*. The predicted probability maps for this group of compounds are presented in
422 Figure S10, Supplementary Material.
423
424 *3.4.6. Thiazoles*
425 Both compounds (*LabMol-102* and *LabMol-103*) from this group were very similar in terms of
426 structure, potency, and predicted ADMET properties (Figure S5C, Supplementary Material). This
427 resulted in practically the same MPO scoring of 0.22 for both compounds. However, as observed in
428 Table 2, the presence of a chlorine atom in *LabMol-102*, instead of two methyl groups (*LabMol-103*) in
429 the benzene ring, may have contributed for the reduction in cytotoxicity against U2OS cells, and
430 consequently, for improved selectivity against the parasite. Both structures have potential for future
431 optimization, aiming to develop new compounds with improved potency, selectivity, and good balance
432 of ADMET properties. The information obtained from the predicted probability maps (Figure S11,
433 Supplementary Material) can be used for future optimization studies.
434

435 **4. Conclusions**

We succeed to develop QSAR models for virtual screening and identification of new potential inhibitors of intracellular *T. cruzi* replication. As a result, 39 potential inhibitors were selected and tested *in vitro* against intracellular Y strain of *T. cruzi* and against mammalian (U2OS) cells. Seven compounds presented high potency at submicromolar range (EC_{50} = 0.3-0.8 μM) and good selectivity against the parasite (SI = 31-274). Among them, *LabMol-128* and *LabMol-134* were the best candidates from the benzylbenzamide and sulfonamide series, respectively. Moreover, *LabMol-134* (EC_{50} = 0.3 μM, SI = 136.8, and logP = 1.82) presented all the necessary requirements to go directly for *in vivo* experimental evaluation. Furthermore, the sulfonamide analogs *LabMol-104*, *LabMol-125, LabMol-126*, and *LabMol-128* are also very promising scaffolds. Three of them (all but *LabMol-104)* presented submicromolar EC_{50} values ranging from 0.4 μM to 0.7 μM with associated selectivity indices ranging from 12 to greater than 274. However, the sulfonamide series present small solubility issues (logP ranging from 3.3 - 3.8) and future hit-to-lead optimization will be performed to overcome this solubility issue and to move forward with this promising series. Other compounds also presented good potency against *T. cruzi* (EC_{50} = 2.7-7.0 μM) and could be also further optimized. Finally, we performed an *in silico* multi-parameter analysis of tested compounds aiming to identify those with an adequate balance between potency, selectivity, and ADMET properties, which comprehends an essential step to develop new, safe, and effective drugs. In the next studies, the most promising compounds will be submitted to *in vivo* assays in acute model of Chagas disease and experimental ADMET evaluation, and can be further optimized for the development of new promising drug candidates against this important yet neglected disease.

Notes

The authors declare no competing financial interest.

Acknowledgments

The authors would like to thank Brazilian funding agencies, CNPq, CAPES and FAPEG for financial support and fellowships. E.M. acknowledges NIH for his financial support (grants 1U01CA207160 and

462 R01-GM114015). We are also grateful to OpenEye Scientific Software Inc. and ChemAxon for
463 providing us with academic licenses for their software.

464 **REFERENCES**

465 [1] C. Bern, Chagas' Disease, N. Engl. J. Med. 373 (2015) 456–466. doi:10.1056/NEJMra1410150.

466 [2] WHO, Integrating neglected tropical diseases into global health and development: fourth WHO
467 report on neglected tropical diseases, (2017) 278.

468 [3] DNDi, What is Chagas Disease?, (2018).

469 [4] WHO, Chagas disease (American trypanosomiasis) Fact Sheet, (2018).

470 [5] WHO, Investing to overcome the global impact of neglected tropical diseases: third WHO report
471 on neglected diseases, (2015) 118–126.

472 [6] Center for Disease Control and Prevention, Triatomine Bug FAQs, (2016).

473 [7] M.A. Shikanai-Yasuda, N.B. Carvalho, Oral Transmission of Chagas Disease, Clin. Infect. Dis.
474 54 (2012) 845–852. doi:10.1093/cid/cir956.

475 [8] T. Grinnage-Pulley, B. Scott, C.A. Petersen, A Mother's Gift: Congenital Transmission of
476 Trypanosoma and Leishmania Species, PLoS Pathog. 12 (2016) 1–7.
477 doi:10.1371/journal.ppat.1005302.

478 [9] M. Flores-Chavez, B. Fernandez, S. Puente, P. Torres, M. Rodriguez, C. Monedero, I. Cruz, T.
479 Garate, C. Canavate, Transfusional Chagas Disease: Parasitological and Serological Monitoring
480 of an Infected Recipient and Blood Donor, Clin. Infect. Dis. 46 (2008) e44–e47.
481 doi:10.1086/527448.

482 [10] E.P. Kransdorf, P.C. Zakowski, J.A. Kobashigawa, Chagas disease in solid organ and heart
483 transplantation, Curr. Opin. Infect. Dis. 27 (2014) 418–424.
484 doi:10.1097/QCO.0000000000000088.

485 [11] J.B. Rodriguez, B.N. Falcone, S.H. Szajnman, Detection and treatment of Trypanosoma cruzi: a
486 patent review (2011-2015)., Expert Opin. Ther. Pat. 26 (2016) 993–1015.
487 doi:10.1080/13543776.2016.1209487.

488 [12] K. Salomao, R.F.S. Menna-Barreto, S.L. de Castro, Stairway to Heaven or Hell? Perspectives and
489 Limitations of Chagas Disease Chemotherapy., Curr. Top. Med. Chem. 16 (2016) 2266–89.

490 [13] M.C. Field, D. Horn, A.H. Fairlamb, M.A.J. Ferguson, D.W. Gray, K.D. Read, M. De Rycker,
491 L.S. Torrie, P.G. Wyatt, S. Wyllie, I.H. Gilbert, Anti-trypanosomatid drug discovery: an ongoing
492 challenge and a continuing need., Nat. Rev. Microbiol. 15 (2017) 217–231.
493 doi:10.1038/nrmicro.2016.193.

494 [14] B.J. Neves, E. Muratov, R.B. Machado, C.H. Andrade, P.V.L. Cravo, Modern approaches to
495 accelerate discovery of new antischistosomal drugs., Expert Opin. Drug Discov. 11 (2016) 557–
496 67. doi:10.1080/17460441.2016.1178230.

497 [15] B.J. Neves, R.F. Dantas, M.R. Senger, C.C. Melo-Filho, W.C.G. Valente, A.C.M. de Almeida,
498 J.M. Rezende-Neto, E.F.C. Lima, R. Paveley, N. Furnham, E. Muratov, L. Kamentsky, A.E.
499 Carpenter, R.C. Braga, F.P. Silva-Junior, C.H. Andrade, Discovery of New Anti-Schistosomal
500 Hits by Integration of QSAR-Based Virtual Screening and High Content Screening., J. Med.
501 Chem. 59 (2016) 7075–88. doi:10.1021/acs.jmedchem.5b02038.

502 [16] C.C. Melo-Filho, R.F. Dantas, R.C. Braga, B.J. Neves, M.R. Senger, W.C.G. Valente, J.M.
503 Rezende-Neto, W.T. Chaves, E.N. Muratov, R.A. Paveley, N. Furnham, L. Kamentsky, A.E.

504 Carpenter, F.P. Silva-Junior, C.H. Andrade, QSAR-Driven Discovery of Novel Chemical
505 Scaffolds Active against Schistosoma mansoni., J. Chem. Inf. Model. 56 (2016) 1357–72.
506 doi:10.1021/acs.jcim.6b00055.

507 [17] M.N. Gomes, L.M. Alcântara, B.J. Neves, C.C. Melo-Filho, L.H. Freitas-Junior, C.B. Moraes, R.
508 Ma, S.G. Franzblau, E. Muratov, C.H. Andrade, Computer-aided discovery of two novel
509 chalcone-like compounds active and selective against Leishmania infantum, Bioorg. Med. Chem.
510 Lett. 27 (2017) 2459–2464. doi:10.1016/j.bmcl.2017.04.010.

511 [18] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P.
512 Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'Min, R. Cramer, R. Benigni, C.
513 Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR modeling: Where
514 have you been? Where are you going to?, J. Med. Chem. 57 (2014) 4977–5010.
515 doi:10.1021/jm4004285.

516 [19] T. Fujita, D.A. Winkler, Understanding the Roles of the "Two QSARs," J. Chem. Inf. Model. 56
517 (2016) 269–274. doi:10.1021/acs.jcim.5b00229.

518 [20] T. Wang, M.-B. Wu, J.-P. Lin, L.-R. Yang, Quantitative structure–activity relationship: promising
519 advances in drug discovery platforms, Expert Opin. Drug Discov. 10 (2015) 1283–1300.
520 doi:10.1517/17460441.2015.1083006.

521 [21] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, Mol.
522 Inform. 29 (2010) 476–488. doi:10.1002/minf.201000061.

523 [22] Luminescence Cell-Based/Microorganism Dose Confirmation HTS to Identify Inhibitors of
524 T.Cruzi Replication, (2018).

525 [23] Luminescence Cell-Based Dose Response HTS Screen to Identify Cytotoxic Compounds of

NIH3T3 cells, (2018).

[24] D. Fourches, E. Muratov, A. Tropsha, Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research., J. Chem. Inf. Model. 50 (2010) 1189–204. doi:10.1021/ci100176x.

[25] D. Fourches, E.N. Muratov, A. Tropsha, Trust, But Verify II: A Practical Guide to Chemogenomics Data Curation, J. Chem. Inf. Model. (2016) DOI:10.1021/acs.jcim.6b00129. doi:10.1021/acs.jcim.6b00129.

[26] D. Fourches, E. Muratov, A. Tropsha, Curation of chemogenomics data., Nat. Chem. Biol. 11 (2015) 535. doi:10.1038/nchembio.1881.

[27] R.C. Braga, V.M. Alves, M.F.B. Silva, E. Muratov, D. Fourches, A. Tropsha, C.H. Andrade, Tuning hERG out: Antitarget QSAR Models for Drug Development., Curr. Top. Med. Chem. 14 (2014) 1399–1415. http://www.ncbi.nlm.nih.gov/pubmed/24805060 (accessed May 13, 2014).

[28] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner., SIGKDD Explor. 11 (2009) 26–31.

[29] R Development Core Team, R: A Language and Environment for Statistical Computing., (2008).

[30] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat. 46 (1992) 175–185. doi:10.1080/00031305.1992.10475879.

[31] S. Riniker, G. a Landrum, Open-source platform to benchmark fingerprints for ligand-based virtual screening., J. Cheminform. 5 (2013) 26. doi:10.1186/1758-2946-5-26.

[32] D. Rogers, M. Hahn, Extended-connectivity fingerprints., J. Chem. Inf. Model. 50 (2010) 742–54.

[33] Python Software Foundation. Python, version 2.7., (n.d.).

[34] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling., J. Chem. Inf. Comput. Sci. 43 (2003) 1947–58. doi:10.1021/ci034160g.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2012) 2825–2830. doi:10.1007/s13398-014-0173-7.2.

[36] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models., J. Chem. Inf. Model. 46 (2006) 1984–95. doi:10.1021/ci060132x.

[37] FILTER 2.1.1: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com., (n.d.).

[38] OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com., (n.d.).

[39] P.C.D. Hawkins, A.G. Skillman, G.L. Warren, B.A. Ellingson, M.T. Stahl, Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database., J. Chem. Inf. Model. 50 (2010) 572–84. doi:10.1021/ci100031x.

[40] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates., J. Med. Chem. 45 (2002) 2615–23. http://www.ncbi.nlm.nih.gov/pubmed/12036371.

[41] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings., Adv. Drug Deliv. Rev. 46 (2001) 3–26. doi:10.1016/S0169-409X(00)00129-0.

[42] C.B. Moraes, M. a Giardini, H. Kim, C.H. Franco, A.M. Araujo-Junior, S. Schenkman, E. Chatelain, L.H. Freitas-Junior, Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against Trypanosoma cruzi: implications for Chagas disease drug discovery and development., Sci. Rep. 4 (2014) 4703. doi:10.1038/srep04703.

[43] Optibrium, ADME QSAR models: Predict key ADME and physicochemical properties prior to synthesis., (2017).

Supplementary Material

Discovery of New Potent Hits Against intracellular *Trypanosoma cruzi* by QSAR-Based Virtual Screening

Cleber C. Melo-Filho[1,£], Rodolpho C. Braga[1,£], Eugene N. Muratov[2,3], Caio Haddad Franco[4], Carolina B. Moraes[4,5], Lucio H. Freitas-Junior[4,5] & Carolina Horta Andrade[1]*

LabMol - Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmacia, Universidade Federal de Goiás - UFG, Rua 240, Qd.87, Goiania, GO 74605-510, Brazil;

[2] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA;

[3] Department of Chemical Technology, Odessa National Polytechnic University, 1. Shevchenko Ave., Odessa, 65000, Ukraine;

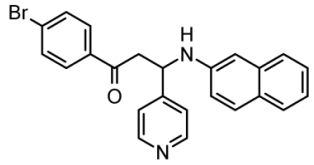
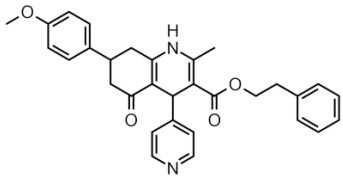
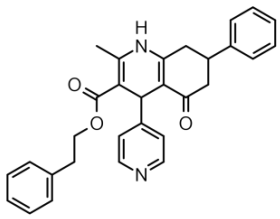
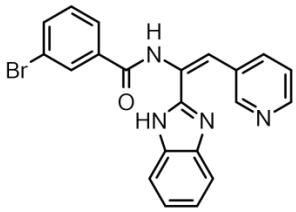
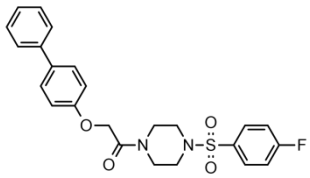
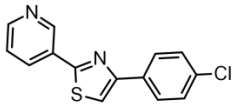
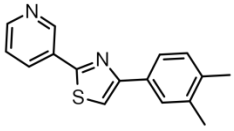
[4]National Laboratory of Biosciences (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, SP 13083-970, Brazil.

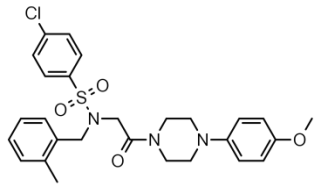
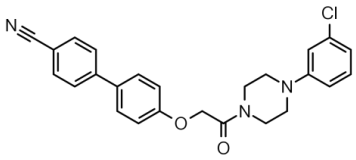
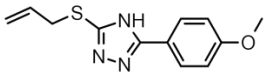
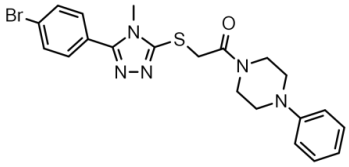
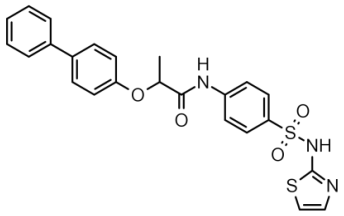
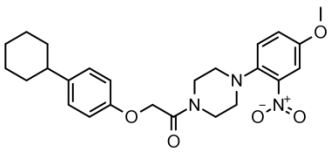
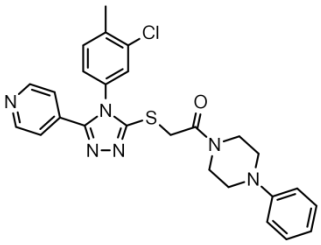
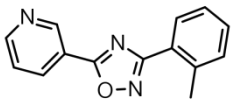
[5]Deparment of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP 05508-900, Brazil.

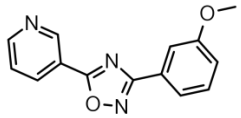
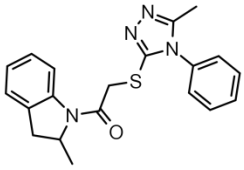
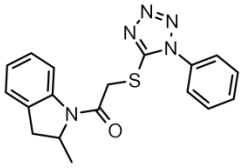
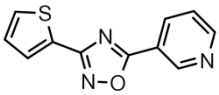
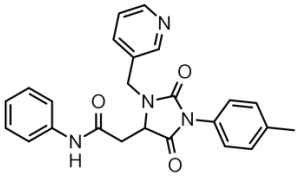
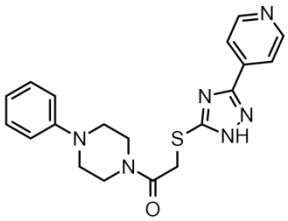
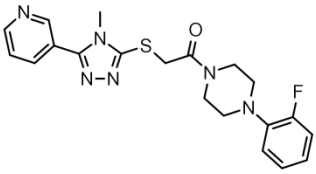
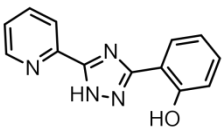
[£]*These authors have equally contributed to this work.*

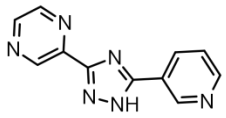
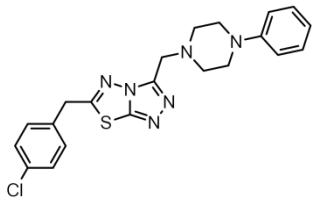
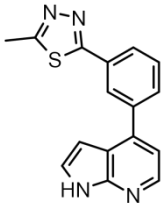
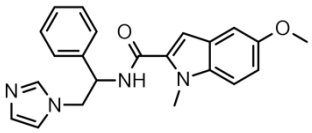
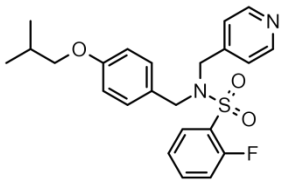
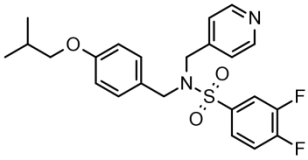
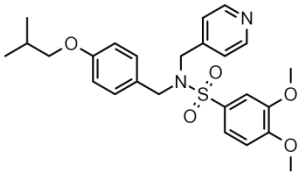
*Corresponding Author: Tel: + 55 62 3209 6451; Fax: +55 62 3209 6037; E-mail: carolina@ufg.br.

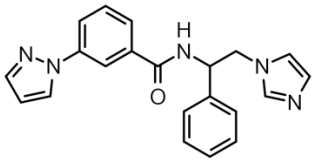
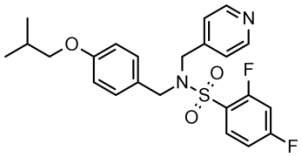
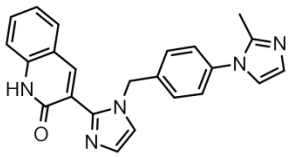
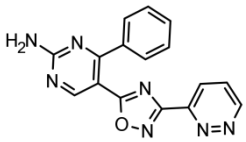
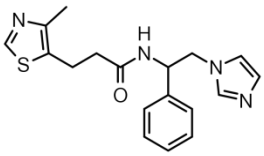
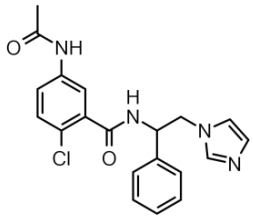
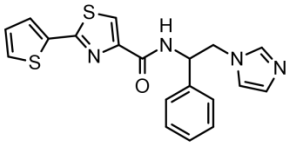
Table S1. List of the 39 virtual hits predicted as inhibitors of intracellular *T. cruzi*, and non-cytotoxic against mammalian cells after virtual screening.

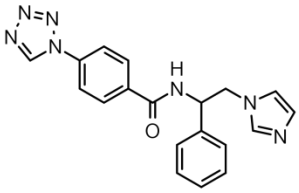
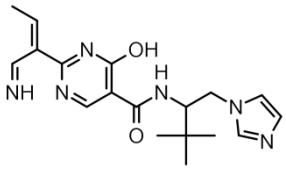
Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-96		0.67
LabMol-97		0.74
LabMol-98		0.76
LabMol-100		0.77
LabMol-101		0.94
LabMol-102		0.77
LabMol-103		0.71

Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-104		0.93
LabMol-105		0.85
LabMol-106		0.75
LabMol-107		0.79
LabMol-108		0.82
LabMol-109		0.90
LabMol-110		0.84
LabMol-111		0.70

Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-112		0.75
LabMol-113		0.89
LabMol-114		0.86
LabMol-115		0.66
LabMol-116		0.80
LabMol-117		0.81
LabMol-118		0.80
LabMol-119		0.83

Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-120		0.79
LabMol-121		0.91
LabMol-122		0.79
LabMol-123		0.84
LabMol-124		0.74
LabMol-125		0.74
LabMol-126		0.78

Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-127		0.80
LabMol-128		0.74
LabMol-129		0.83
LabMol-130		0.73
LabMol-131		0.76
LabMol-132		0.80
LabMol-133		0.74

Compound	Structure	Similarity to nearest neighbor in modeling set[a]
LabMol-134		0.77
LabMol-135		0.67

[a] Similarity to the nearest neighbor in data set I (*T. cruzi* inhibition data set) expressed as Tanimoto coefficient.

Table S2. Results of *in vitro* experimental evaluation of virtual hits against intracellular Y strain of *T. cruzi* and human U2OS cells.

Compound	EC_{50} *T. cruzi* (µM)	CC_{50} U2OS cells (µM)	SI
LabMol-96	**0.8 ± 0.1**	**> 200**	**> 258**
LabMol-97	**0.4 ± 0.1**	**19.3 ± 10.6**	**46.5**
LabMol-98	> 200	> 200	-
LabMol-100	18.9 ± 8.3	47.4 ± 5.5	2.5
LabMol-101	> 200	> 200	-
LabMol-102	**4.2 ± 1.0**	**> 200**	**> 48**
LabMol-103	**2.7 ± 0.5**	**40.8 ± 5.4**	**15.1**
LabMol-104	**2.7 ± 0.4**	**73.9***	**27.7**
LabMol-105	> 200	> 200	-
LabMol-106	148.8*	195.3 ± 3.7	1.3
LabMol-107	> 200	> 200	-
LabMol-108	> 200	> 200	-
LabMol-109	38.3 ± 1.1	49.6 ± 5.6	1.3
LabMol-110	**7.0 ± 1.4**	**21.1 ± 7.2**	**3.0**
LabMol-111	18.6 ± 6.9	> 200	> 10.8
LabMol-112	86.0 ± 8.5	106.1 ± 0.7	1.2
LabMol-113	36.9 ± 5.1	154.8*	4.2
LabMol-114	108.9*	> 200	> 1.8
LabMol-115	> 200	66.3 ± 16.1	< 0.3
LabMol-116	190.6 ± 9.9	> 200	> 1.0
LabMol-117	20.1 ± 3.7	> 200	> 9.9
LabMol-118	> 200	198.1*	< 1.0

Compound	EC_{50} *T. cruzi* (μM)	CC_{50} U2OS cells (μM)	SI
LabMol-119	> 200	> 200	-
LabMol-120	**3.4 ± 0.5**	**1.5 ± 0.0**	**0.4**
LabMol-121	> 200	> 200	-
LabMol-122	> 200	127.0 ± 28.3	< 0.6
LabMol-123	13.3 ± 1.4	0.62 ± 0.3	0.1
LabMol-124	> 200	> 200	-
LabMol-125	**0.4 ± 0.02**	**11.5 ± 10.0**	**30.6**
LabMol-126	**0.5 ± 0.02**	**21.7 ± 3.4**	**45.7**
LabMol-127	**0.3 ± 0.2**	**11.8 ± 1.3**	**46.3**
LabMol-128	**0.7 ±0.1**	**> 200**	**> 274**
LabMol-129	**1.2 ± 0.1**	**52.9 ± 5.1**	**43.9**
LabMol-130	> 200	99.7 ± 1.7	< 0.5
LabMol-131	> 200	> 200	-
LabMol-132	**2.7 ± 0.3**	**> 200**	**> 72.8**
LabMol-133	> 200	> 200	-
LabMol-134	**0.3 ± 0.2**	**33.6 ± 7.7**	**136.8**
LabMol-135	**5.8 ± 0.7**	**> 200**	**> 34.6**
benznidazole	3.1 ± 0.6	> 200	> 127.6

* Value obtained in a single experiment; EC_{50}: compound concentration that causes a reduction of 50% in infection, compared to infected controls; CC_{50}: compound concentration that causes reduction in 50% in the number of human osteosarcoma cells; SI: selectivity index.

T. cruzi model

True class \ Predicted class	Inhibitors	Non-inhibitors
Inhibitors	**1471** (TP)	**564** (FN)
Non-inhibitors	**463** (FP)	**1550** (TN)

Number of compounds: 4,048
(2,035 inhibitors : 2,013 non-inhibitors)

T. cruzi model - AD

True class \ Predicted class	Inhibitors	Non-inhibitors
Inhibitors	**1132** (TP)	**329** (FN)
Non-inhibitors	**273** (FP)	**776** (TN)

Number of compounds: 2,510
(1,461 inhibitors : 1,049 non-inhibitors)

Fig. S1. Confusion matrices for *T. cruzi* model and *T. cruzi* model-AD indicating the actual *versus* predicted classes of compounds from data set I. (TP: number of true positives; FP: number of false positives; FN: number of false negatives; TN: number of true negatives).

Cytotoxicity model

True class \ Predicted class	Cytotoxic	Non-cytotoxic
Cytotoxic	678 (TP)	334 (FN)
Non-cytotoxic	304 (FP)	708 (TN)

Number of compounds: 2,024
(1,012 cytotoxic : 1,012 non-cytotoxic)

Cytotoxicity model - AD

True class \ Predicted class	Cytotoxic	Non-cytotoxic
Cytotoxic	483 (TP)	188 (FN)
Non-cytotoxic	188 (FP)	436 (TN)

Number of compounds: 1,295
(671 cytotoxic : 624 non-cytotoxic)

Fig. S2. Confusion matrices for cytotoxicity model and cytotoxicity model-AD indicating the actual *versus* predicted classes of compounds from data set II. (TP: number of true positives; FP: number of false positives; FN: number of false negatives; TN: number of true negatives).

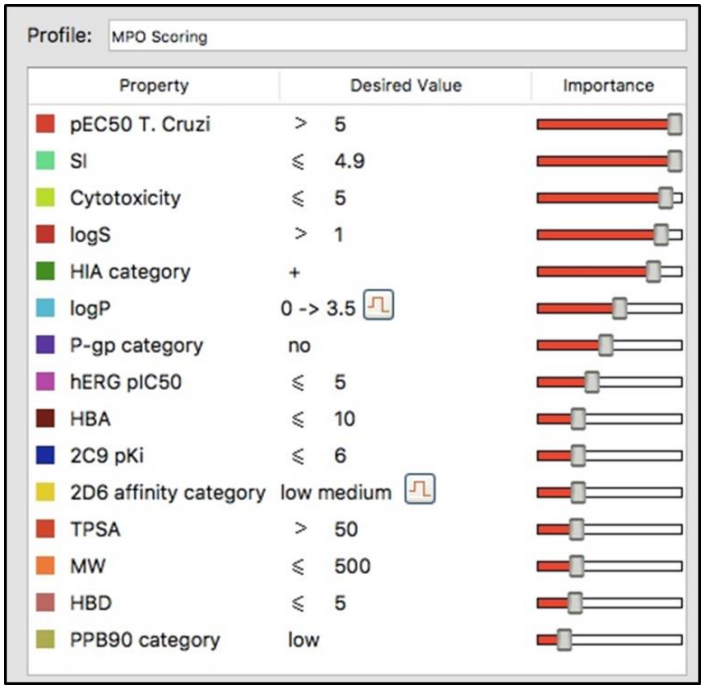


Fig. S3. List of properties used for the multi-parameter analysis. The desired cut-off value for each property is represented in the center column. The level of importance of each parameter, which was manually set before the analysis, is represented by the red bars on the right column. The EC_{50} against *T. cruzi* Y strain, the cytotoxicity on U2OS cells, and SI were converted to the negative logarithmic scale before the analysis. The ADMET and physicochemical properties were predicted by StarDrop built-in QSAR models. Additional parameters were also calculated, such as: Topological Polar Surface Area (TPSA); Hydrogen Bond Acceptors (HBA); Hydrogen Bond Donors (HBD); and Molecular Weight (MW).

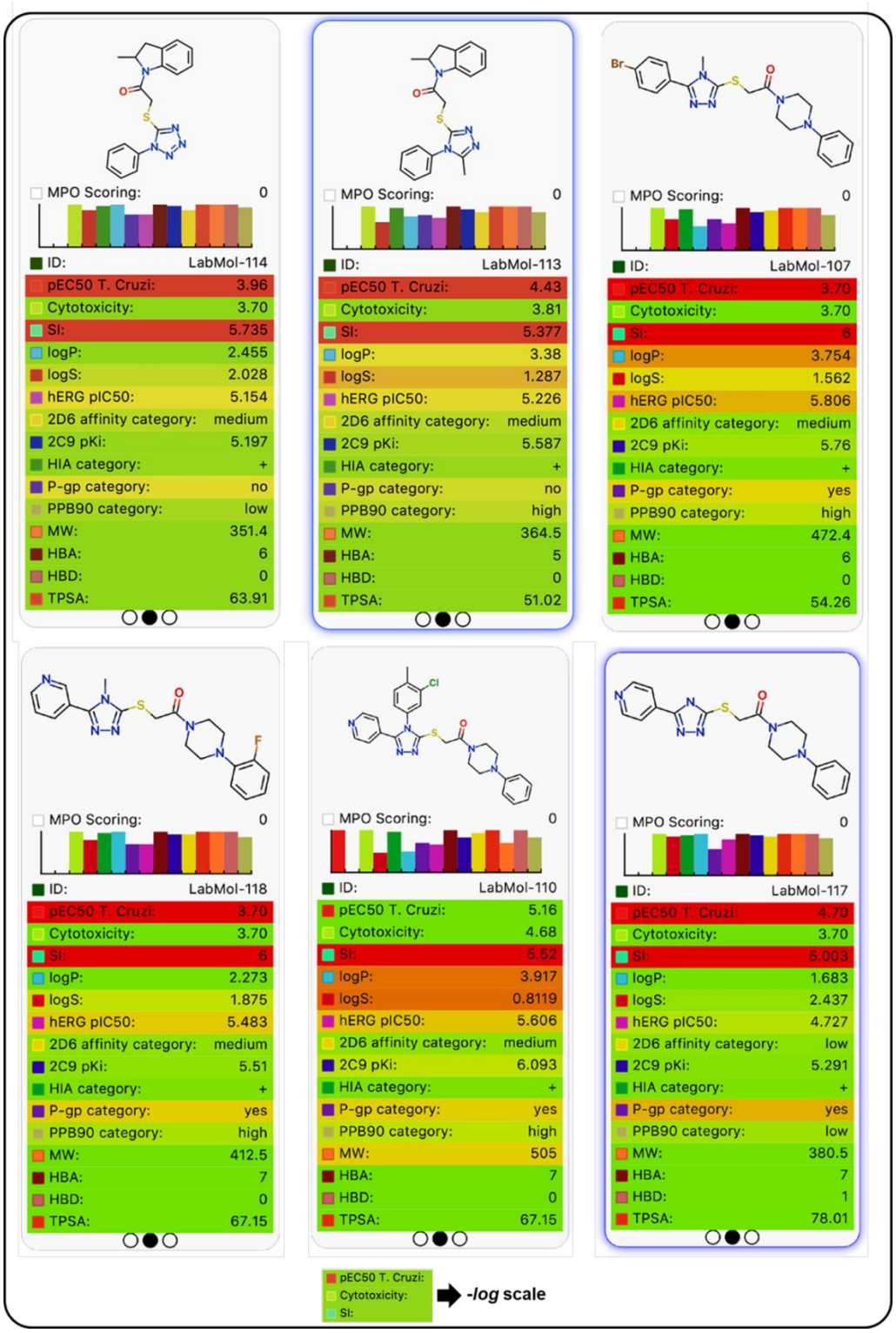


Fig. S4. Multi-parameter analysis of sulfanylacetamides. Colors near green indicate properties with adequate values, and colors near red indicate properties that need improvement. Intermediate results are presented between yellow and orange colors.

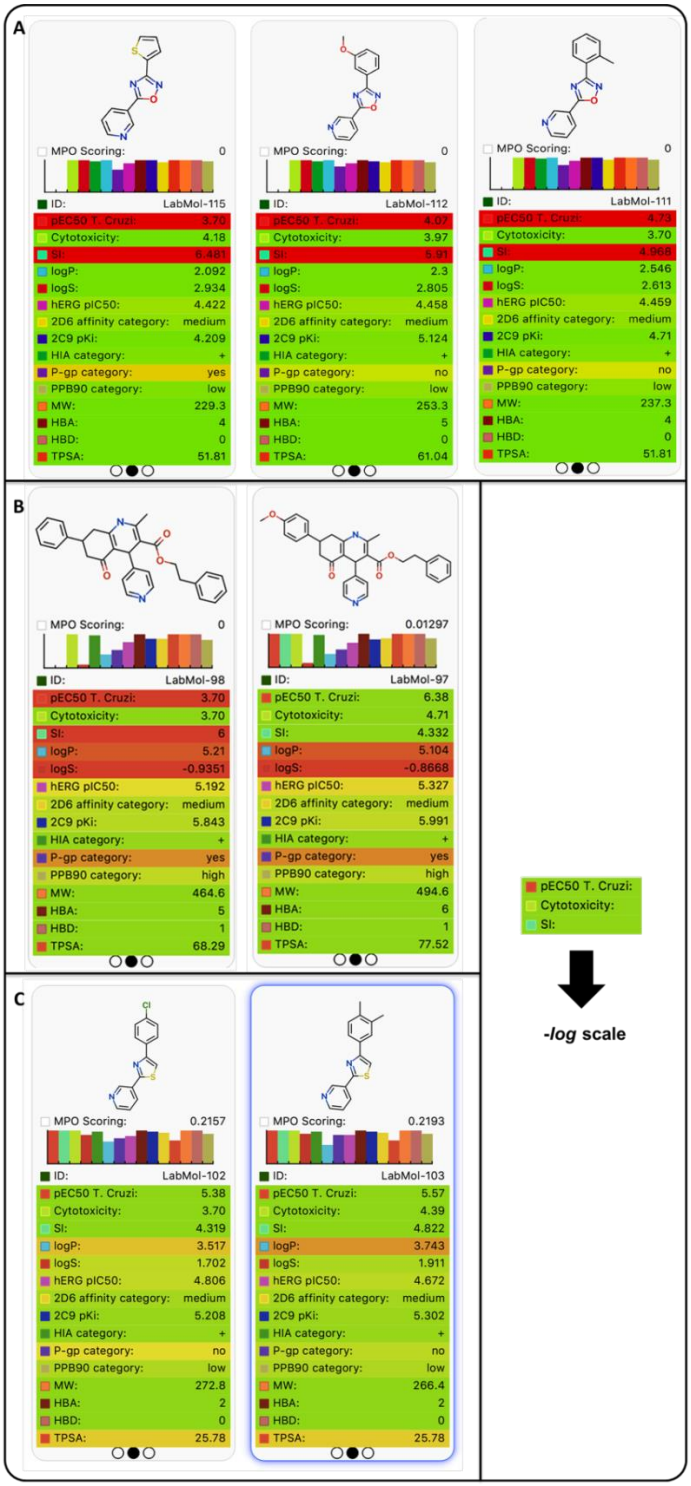


Fig. S5. Multi-parameter analysis of oxadiazoles (A), tetrahydroquinolines (B), and thiazoles (C). Colors near green indicate properties with adequate values, and colors near red indicate properties that need improvement. Intermediate results are presented between yellow and orange colors.

Predicted probability maps. The predicted probability maps were calculated for all compounds submitted to the *in silico* multi-parameter analysis, which were previously tested *in vitro* against intracellular *T. cruzi*. These maps were based on the predictions from *T. cruzi* QSAR model (ECFP4 fingerprints: radius 2; bit vector = 2,048 bits; and Random Forest algorithm), calculated in the open-source packages RDKit (http://www.rdkit.org) and scikit-learn 0.18 (http://scikit-learn.org). The figures below represent the maps for each cluster of compounds and highlight the fragments with positive (green) and negative (purple) contribution to biological activity. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale. The information from the maps can be useful in future compound optimization studies.

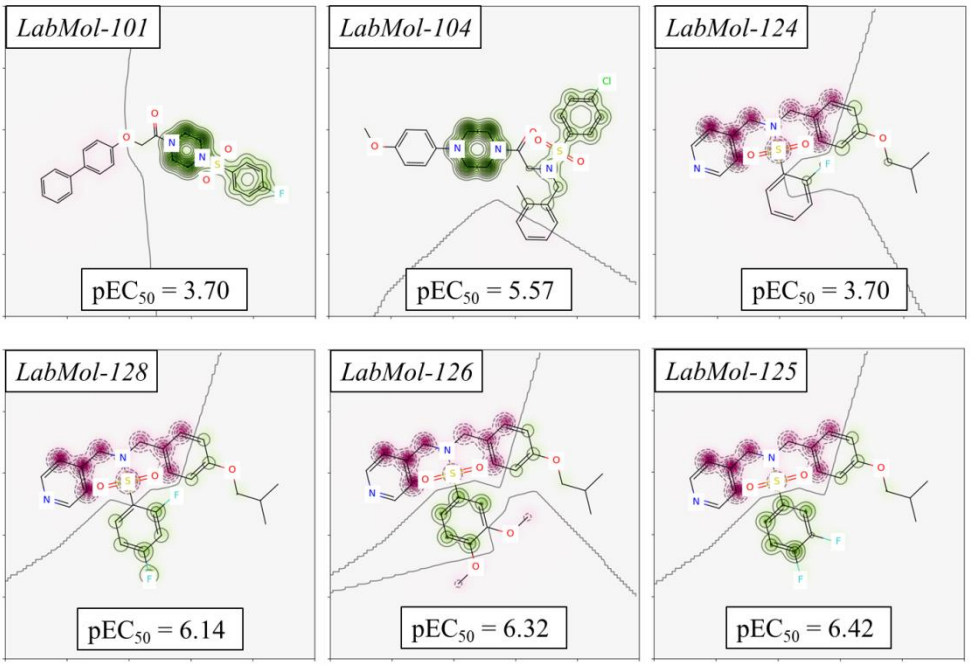


Fig. S6. Predicted probability maps calculated for the sulfonamides. The fragments with positive contribution to biological activity are highlighted in green and those with negative contribution are highlighted in purple. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

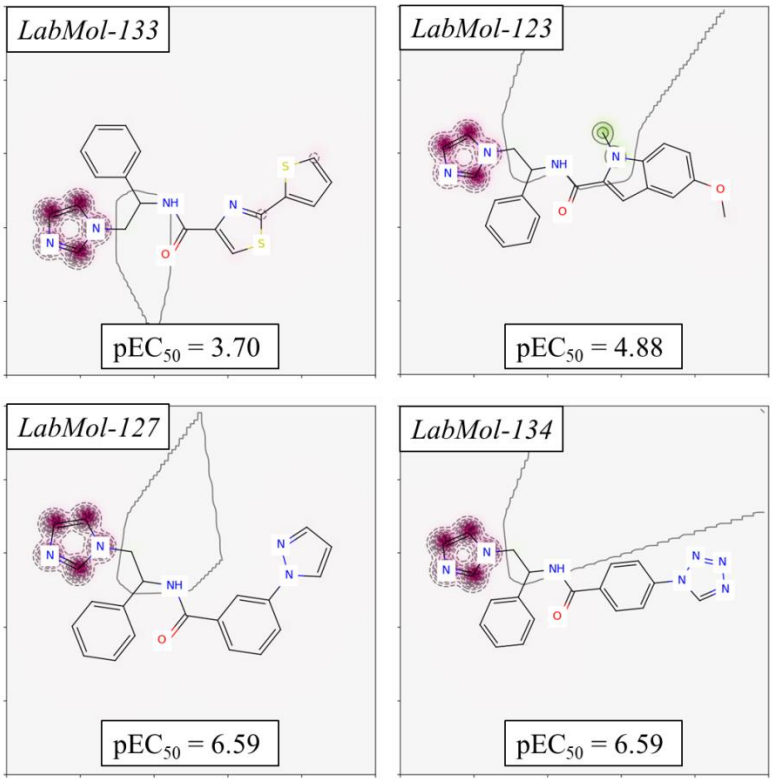


Fig. S7. Predicted probability maps calculated for the benzylbenzamides. These results suggest that the imidazole ring, present in all compounds from this cluster, has negative contribution to biological activity (highlighted in purple). The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

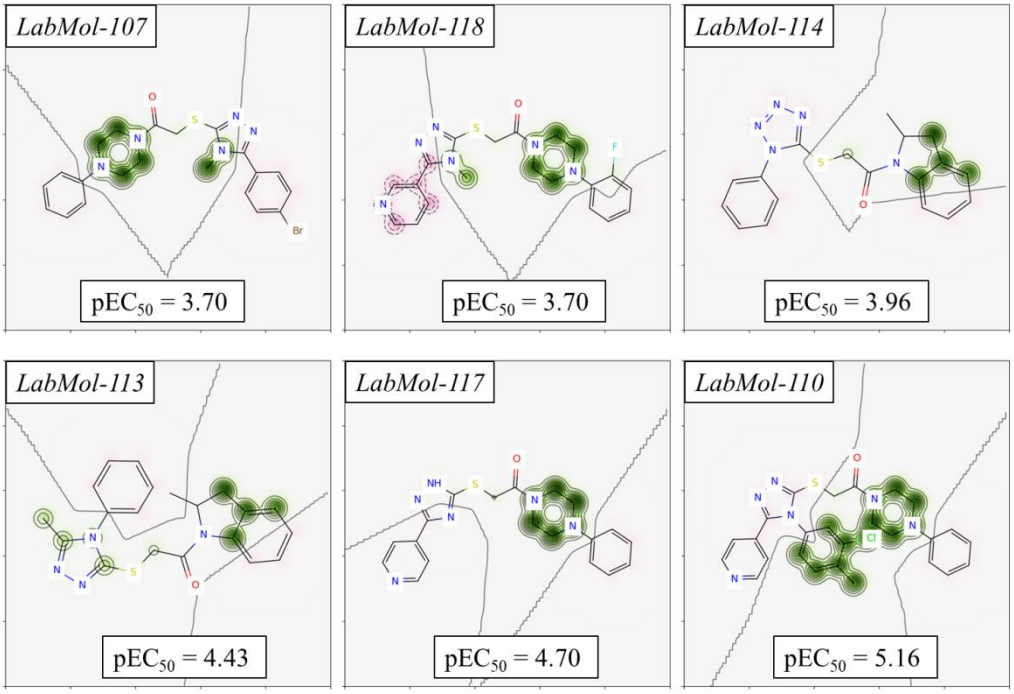


Fig. S8. Predicted probability maps calculated for the sulfanylacetamides. The fragments with positive contribution to biological activity are highlighted in green and those with negative contribution are highlighted in purple. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

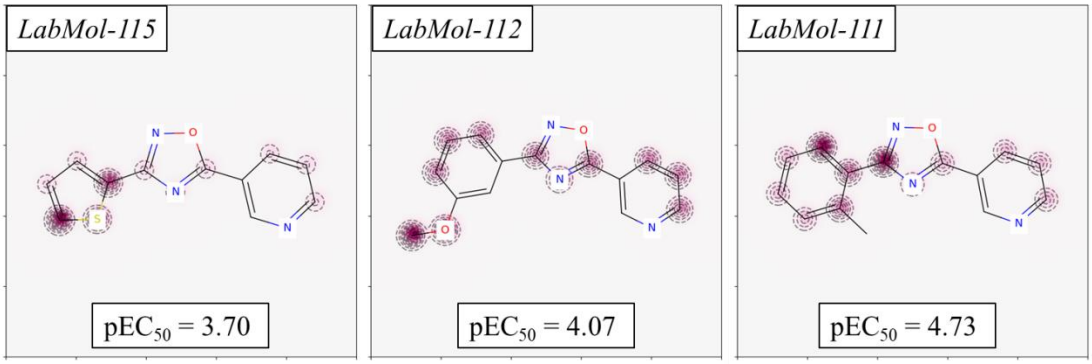


Fig. S9. Predicted probability maps calculated for the oxadiazoles. The fragments with negative contribution to biological activity are highlighted in purple. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

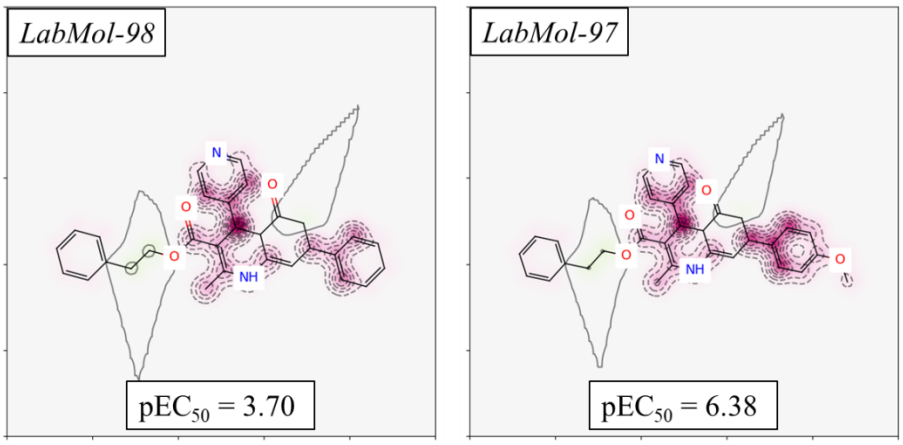


Fig. S10. Predicted probability maps calculated for the tetrahydroquinolines. The fragments with negative contribution to biological activity are highlighted in purple. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

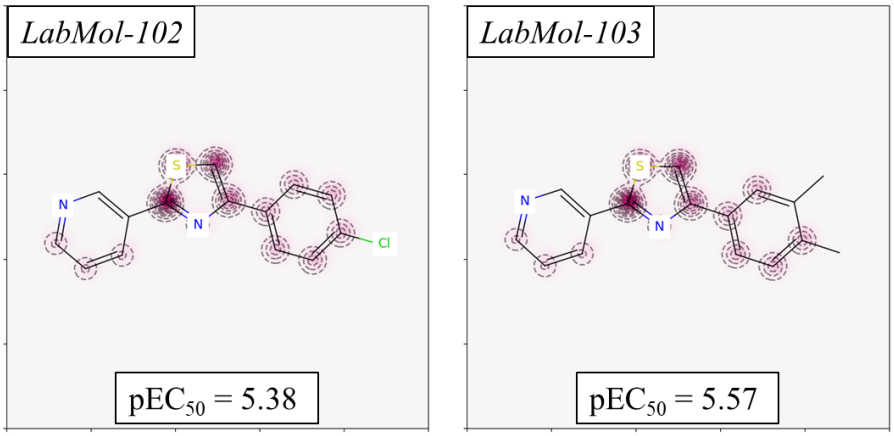


Fig. S11. Predicted probability maps calculated for the thiazoles. The fragments with negative contribution to biological activity are highlighted in purple. The pEC_{50} corresponds to the *in vitro* potency of the compounds after the experiment against intracellular *T. cruzi*, expressed in the negative logarithmic scale.

5. RESULTADOS

5.2. Identificação de novos potenciais inibidores da enzima piruvato quinase (PK) de *Leishmania* spp.

Nesse estudo, foi realizada uma triagem virtual de novos potenciais inibidores da enzima PK de *Leishmania*, utilizando diferentes estratégias baseadas no ligante (LBDD), como a aplicação de modelos *shape-based* e de QSAR, além de estratégia baseada na estrutura (SBDD), através da geração e aplicação do modelo baseado em *docking*. Conforme explicado na seção 4.2, foi utilizado um conjunto de dados contendo inibidores e não inibidores da enzima piruvato quinase (PK) de *L. mexicana*, depositado na base de dados *PubChem Bioassay* (PubChem Identifier: AID 1721; URL: *https://pubchem.ncbi.nlm.nih.gov/bioassay/1721*), o qual foi preparado conforme as boas práticas de CADD e utilizado na geração e validação dos modelos. Os resultados obtidos na validação e aplicação desses modelos serão discutidos a seguir.

5.2.1. Validação dos modelos *shape-based* e do modelo baseado em *docking*

Cinco modelos baseados na forma e volume moleculares (*shape-based*) e um modelo baseado em *docking* foram gerados e validados nesse estudo. Na tabela 4 estão apresentados os resultados da validação desses modelos.

Tabela 4 - Validação dos modelos *shape-based* (I-V) e SBDD (*docking*) a partir do conjunto de dados de inibidores e não inibidores da enzima PK (PubChem AID: 1721), balanceados na proporção de 1:36.

Modelo	AUC	Top 1% hits				Top 5% hits				Top 10% hits			
		EF	BEDROC	SE	SP	EF	BEDROC	SE	SP	EF	BEDROC	SE	SP
I	0,52	3,23	0,48	0,50	0,50	1,94	0,25	0,50	0,50	1,29	0,20	0,50	0,50
II	0,79	3,38	0,96	0,51	0,53	2,30	0,77	0,54	0,57	2,43	0,74	0,58	0,64
III	0,70	2,75	0,45	0,50	0,76	1,28	0,27	0,50	0,76	1,47	0,24	0,50	0,73
IV	0,80	5,17	0,52	0,53	0,52	4,33	0,48	0,54	0,57	3,61	0,50	0,58	0,62
V	**0,92**	**3,73**	**1,00**	**0,66**	**0,61**	**3,47**	**0,96**	**0,74**	**0,74**	**3,47**	**0,92**	**0,80**	**0,79**
Docking	**0,86**	**0,89**	**0,14**	**0,58**	**0,51**	**0,89**	**0,22**	**0,69**	**0,56**	**0,90**	**0,34**	**0,75**	**0,62**

AUC: área sob a curva ROC; EF: Fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; SE: sensibilidade; SP: especificidade; *Top x% hits*: métricas calculadas para os primeiros *x%* compostos da lista ordenada pelo *score TanimotoCombo* (modelo *shape-based*) ou pelo *score* de *docking* (modelo basedo em

docking); O melhor modelo *shape-based* e o modelo baseado em *docking*, usados posteriormente na triagem virtual, estão destacados em negrito.

Conforme observado na tabela 4, dentre os modelos *shape-based* (modelos I ao V), o modelo V foi aquele que apresentou o melhor desempenho nos estudos de validação, demonstrando a sua capacidade superior em classificar e ordenar corretamente os compostos inibidores e não inibidores. Esse modelo apresentou AUC de 0,92, em contraste ao pior modelo (modelo I), que apresentou AUC de apenas 0,52. Além disso, apresentou BEDROC sempre acima de 0,9, além de SE e SP sempre acima de 0,6, chegando a valores de 0,8 (*top 10% hits*). O modelo V está representado na figura 20.

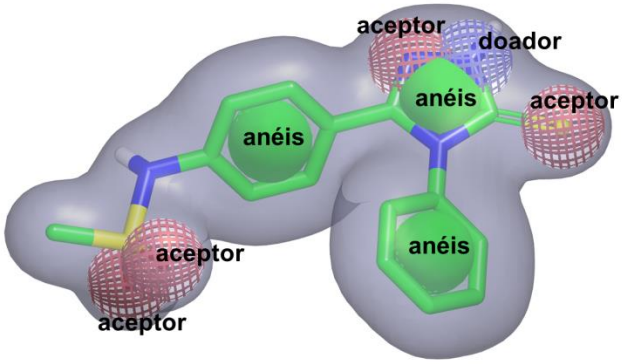


Figura 20 - Melhor modelo *shape-based* (modelo V) selecionado após a validação com inibidores e não inibidores de PK. Esferas vermelhas correspondem à característica de aceptor de ligação de hidrogênio; Esferas azuis correspondem à característica de doador de ligação de hidrogênio; Esferas verdes correspondem à presença de anéis. O contorno em cor cinza representa o *shape* (forma e volume) da molécula.

O modelo baseado em *docking* apresentou um bom valor de AUC (0,86) e, apesar de não apresentar EF e BEDROC comparáveis ao modelo V, apresentou valores aceitáveis de SE e SP (0,75 e 0,62) a partir de uma lista contendo os 10% primeiros compostos ranqueados (*top 10% hits*). O modelo V e o modelo baseado em *docking* foram utilizados posteriormente como filtros na triagem virtual.

5.2.2. Modelos de QSAR

O conjunto de dados de inibidores e não-inibidores de PK previamente preparado, padronizado e balanceado (1:1) foi utilizado para a geração de modelos de QSAR utilizando diferentes *fingerprints* moleculares e o método de aprendizado de máquina RF. Esses modelos foram combinados em um modelo consenso, usado posteriormente na triagem virtual de potenciais inibidores da PK de *Leishmania*. Na tabela 5, estão apresentados os resultados da avaliação da performance dos modelos de QSAR, obtidos após validação cruzada externa *5-fold*.

Tabela 5 - Parâmetros estatísticos dos modelos de QSAR após validação cruzada externa *5-fold*.

Modelo	CCR	*k*	SE	SP	Cobertura
MACCS-RF	0,85	0,71	0,86	0,84	0,65*
Atom Pair-RF	0,86	0,71	0,86	0,85	0,61*
Morgan-RF	**0,87**	**0,74**	**0,85**	**0,89**	**0,61***
FeatMorgan-RF	0,86	0,71	0,84	0,87	0,61*
Avalon-RF	0,86	0,71	0,85	0,87	0,65*
Consenso	**0,87**	**0,74**	**0,85**	**0,88**	**1,00**
Consenso Rigoroso	0,77	0,54	0,95	0,59	0,40**

RF: *Random Forest*; CCR: taxa de classificação correta; *k*: kappa; SE: sensibilidade; SP: especificidade; *Cobertura do modelo ao considerar o DA; **O consenso rigoroso levou em consideração apenas as predições de compostos que estavam dentro do DA de todos os cinco modelos; O modelo consenso foi obtido a partir da média aritmética das predições dos modelos individuais.

De acordo com a tabela 5, pode-se observar que todos os modelos individuais apresentaram desempenho muito próximo. Todos os modelos demonstraram boa preditividade com CCR, SE e SP acima de 0,8. Além disso, demonstraram uma boa concordância entre as predições e os valores experimentais de atividade biológica, com valores de *k* acima de 0,7. O modelo consenso apresentou uma pequena melhora nos parâmetros, enquanto o modelo rigoroso, que só considera predições de compostos que estão dentro do DA de todos os modelos, apresentou naturalmente a menor cobertura. O modelo consenso foi utilizado posteriormente na triagem virtual de novos potenciais inibidores da enzima PK.

Na avaliação da robustez, através do método de *Y-randomization*, foi observado que a correlação entre a estrutura e atividade não ocorreu ao acaso, uma vez que todos

os modelos apresentaram após as permutações, em média, CCR, SE e SP próximos a 0,5, além de valores de *k* próximos a zero.

O modelo Morgan-RF, que apresentou melhor CCR e *k* dentre os modelos individuais, foi utilizado para interpretação através da análise de mapas de probabilidade predita (MPP) desenvolvido por Riniker e Landrum, que foi aplicada aos compostos do conjunto de dados (RINIKER; LANDRUM, 2013b). Esse método consiste em estimar a contribuição de cada átomo presente na molécula para a atividade, seguindo-se os seguintes passos: i) remoção de um átomo da molécula (atribuição de valor 0 aos *bins* correspondentes a ele no vetor do *fingerprint* molecular) ; ii) predição da atividade da nova molécula (após retirada do átomo); iii) cálculo da diferença do valor de probabilidade predita entre a nova molécula e a molécula anterior (antes da retirada do átomo). Esses passos são repetidos para todos os átomos, permitindo então se estimar quais fragmentos (átomos) da molécula tem contribuição positiva, neutra ou negativa para a atividade. Ao final, um mapa de cores é gerado, destacando as regiões das moléculas e suas respectivas contribuições (RINIKER; LANDRUM, 2013b). A vantagem desse método consiste na sua aplicação na interpretação local de modelos que utilizam métodos de aprendizado de máquina de difícil interpretação, como no caso do RF (POLISHCHUK, 2017). Os MPP foram calculados através do *script* aberto disponibilizado pelos autores do método em questão, em seu artigo original (RINIKER; LANDRUM, 2013b). O *script*, que integra diversos pacotes abertos como RDKit, *scikit-learn* e *matplotlib*, foi executado em Python 2.7 (https://www.python.org).

Nas figuras 21 e 22, estão representados os MPP, os quais destacam os fragmentos com contribuição positiva mais frequentemente encontrados nos inibidores (compostos ativos) e os fragmentos com contribuição negativa encontrados com maior frequência nos não inibidores de PK (compostos inativos). Conforme observado na figura 21, os seguintes fragmentos foram encontrados com maior frequência no conjunto de compostos ativos: piridina-2-carboximidamida; 1,2,4-triazol-3-tiol; imidazol-2-tiol; fenildiazeno; 1,2,4-oxadiazol; imidazo[1,2-a]piridina; 1,3-tiazol; 1,3-benzotiazol. Os fragmentos mais frequentes no conjunto de compostos inativos (Figura 22) foram: piperazina; morfolina; sulfonilbenzeno; sulfonilpiperidina; N-etilacetamida. A informação obtida da análise dos PPM poderá ser útil em futuros estudos de otimização dos *hits* encontrados nesse trabalho ou no planejamento de novos inibidores de PK de *Leishmania*.

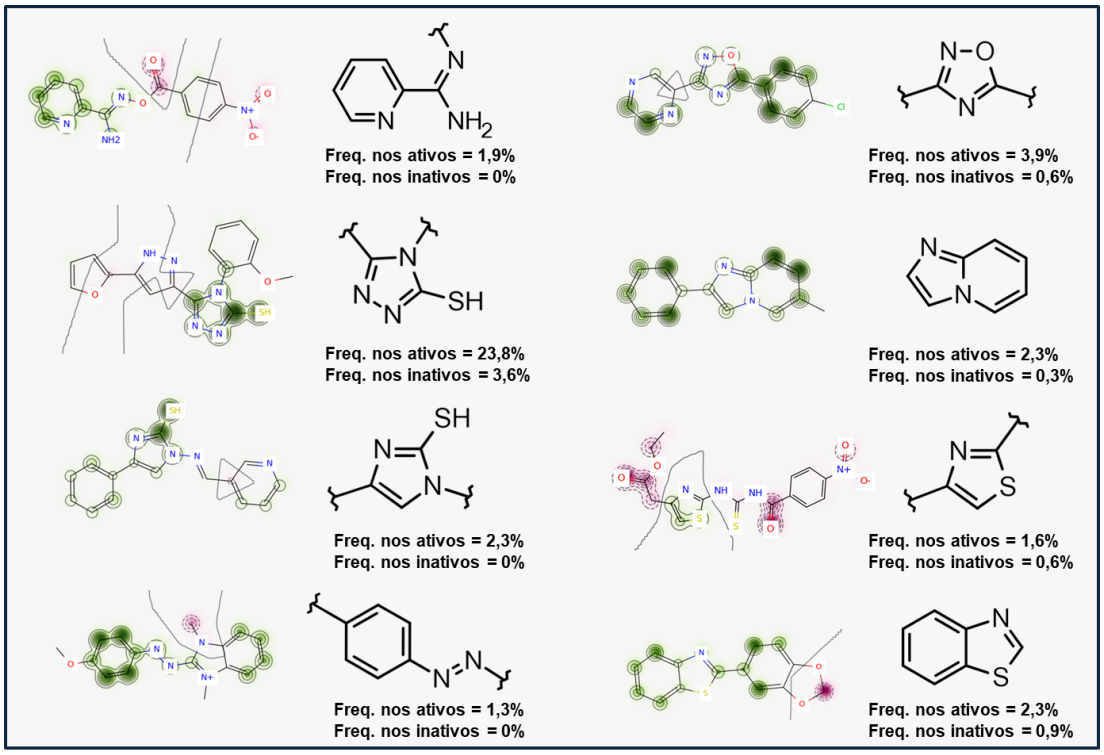


Figura 21 – Fragmentos mais frequentes encontrados no conjunto de inibidores da PK (compostos ativos), sua respectiva frequência e exemplos de MPP contendo esses fragmentos. Os fragmentos em verde indicam contribuição positiva para a atividade, enquanto fragmentos em rosa indicam contribuição negativa. As partes não coloridas indicam contribuição neutra.

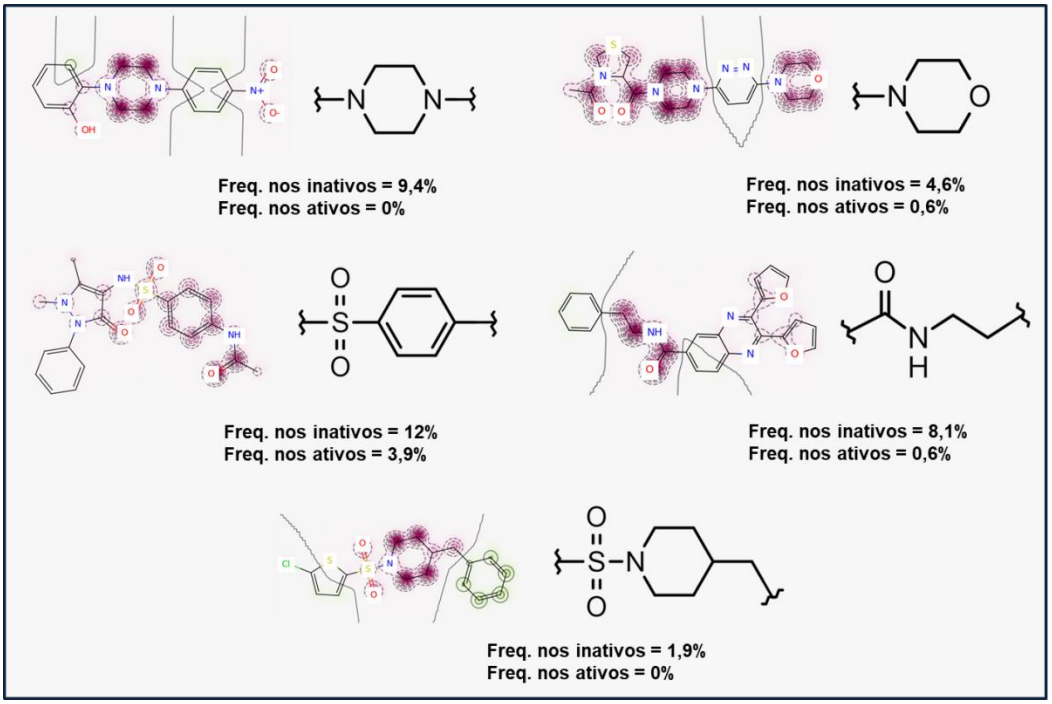


Figura 22 - Fragmentos mais frequentes encontrados no conjunto de não inibidores da PK (compostos inativos), sua respectiva frequência e exemplos de

MPP contendo esses fragmentos. Os fragmentos em verde indicam contribuição positiva para a atividade, enquanto fragmentos em rosa indicam contribuição negativa. As partes não coloridas indicam contribuição neutra.

5.2.3. Triagem virtual

A base de dados comercial ChemBridge, contendo 1.063.926 compostos, foi submetida à triagem virtual de novos potenciais inibidores da PK de *Leishmania*, utilizando os modelos *shape-based*, modelo baseado em *docking* e de QSAR, previamente validados. O protocolo da triagem virtual está apresentado em detalhes na seção 4.2.6. Na figura 23, estão representados os resultados da triagem virtual.

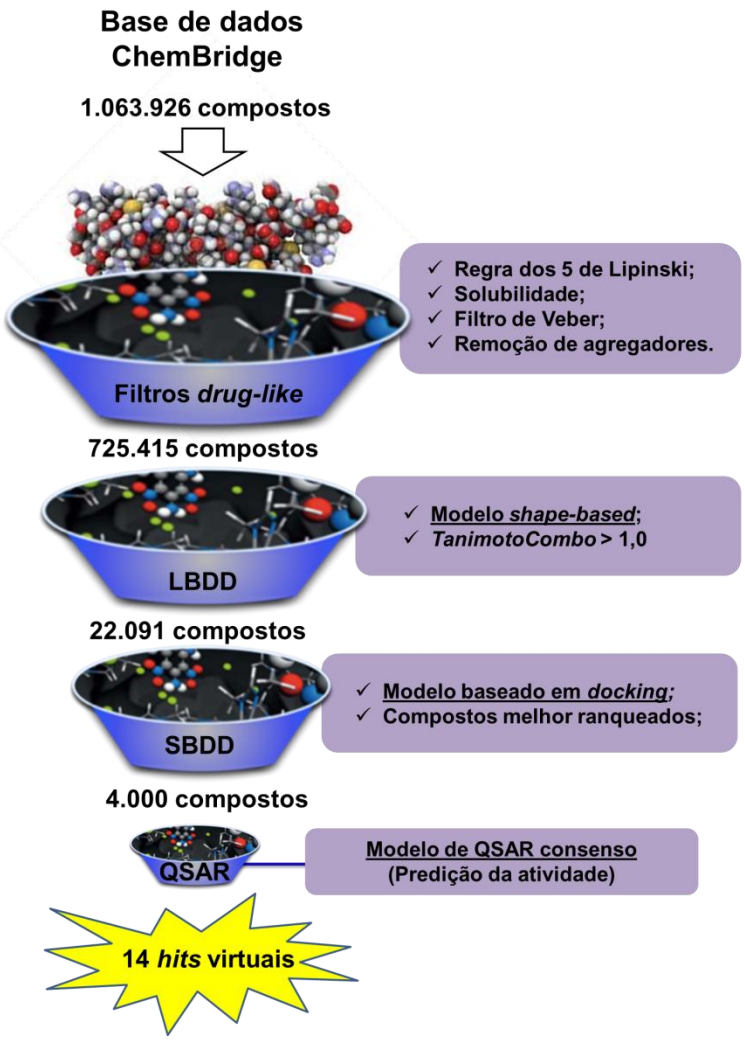


Figura 23 - Resultados da triagem virtual de potenciais inibidores da enzima PK de *Leishmania* spp.

Conforme observado na figura 23, o primeiro filtro utilizado na triagem virtual (filtro de propriedades *drug-like*) permitiu a seleção de 725.415 compostos a partir da base de dados ChemBridge, os quais foram submetidos, em seguida, ao melhor modelo *shape-based*. Nessa etapa, 22.091 compostos com *score TanimotoCombo* acima de 1,0, após sobreposição ao modelo *shape-based*, foram selecionados. Esses compostos foram então submetidos ao *docking* no sítio de ATP da enzima PK. Após o *docking*, os 4.000 compostos melhor ranqueados pela função de *score Chemgauss4* foram selecionados e submetidos à última etapa da triagem virtual. Nessa última etapa, os 4.000 compostos tiveram sua atividade predita pelo modelo de QSAR consenso, que classificou os compostos em inibidores ou não inibidores da PK. Foram classificados como ativos todos os compostos preditos como inibidores em pelo menos 80% dos modelos que compunham o consenso. Além disso, esses compostos deveriam estar dentro do DA de pelo menos 50% desses modelos. Ao final dessa etapa, 14 *hits* virtuais foram selecionados e adquiridos para serem testados. A tabela 6 contém as estruturas dos *hits* virtuais obtidos nesse estudo. A validação experimental desses potenciais inibidores de PK em diferentes espécies de *Leishmania* está sendo realizada em colaboração com o Dr. André Gustavo Tempone Cardoso (Instituto Adolfo Lutz - SP) e Dr. Jair Lage de Siqueira Neto (*Skaggs School of Pharmacy and Pharmaceutical Sciences* - UCSD, USA).

Tabela 6 - Estruturas dos 14 potenciais inibidores da enzima PK de *Leishmania* selecionados após triagem virtual através de modelos *shape-based*, baseado em *docking* e QSAR.

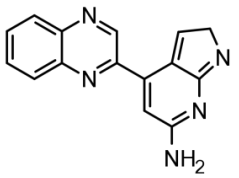
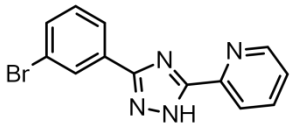
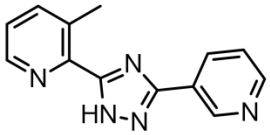
Estrutura	***Docking score* (*Chemgauss4*)**	**QSAR consenso *score* [b]**	**QSAR DA *score* [c]**
	-7,58	1,0	0,6
	-6,17	1,0	1,0
	-7,07	1,0	1,0

Tabela 6 (cont.) - Estruturas dos 14 potenciais inibidores da enzima PK de *Leishmania* selecionados após triagem virtual através de modelos *shape-based*, baseado em *docking* e QSAR.

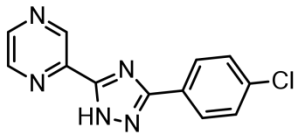
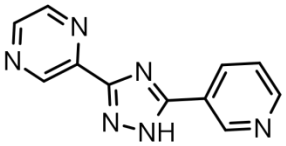
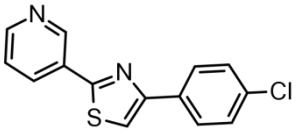
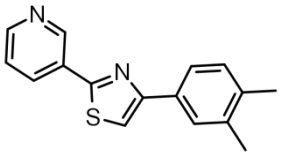
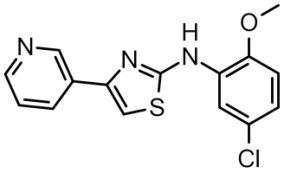
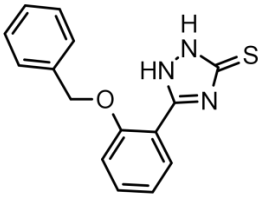
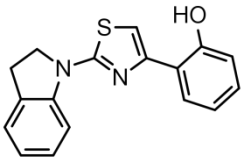
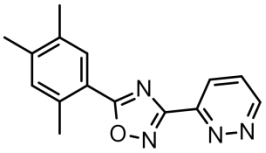
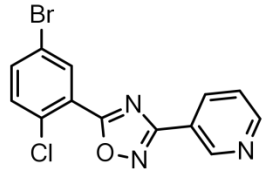
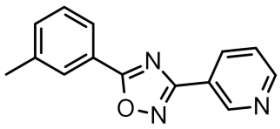
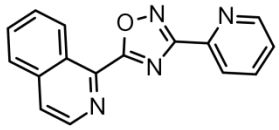
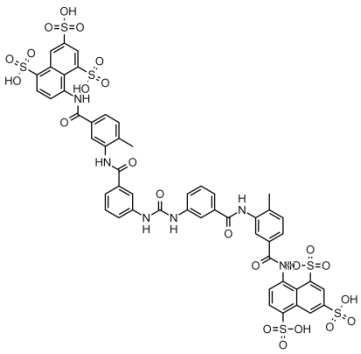
Estrutura	*Docking score* (*Chemgauss4*)	QSAR consenso *score* [b]	QSAR DA *score* [c]
	-6,93	1,0	1,0
	-7,32	1,0	1,0
	-7,65	1,0	1,0
	-7,80	1,0	1,0
	-7,59	1,0	0,6
	-9,28	1,0	0,8
	-6,51	1,0	0,8
	-7,85	1,0	1,0

Tabela 6 (cont.) - Estruturas dos 14 potenciais inibidores da enzima PK de *Leishmania* selecionados após triagem virtual através de modelos *shape-based*, baseado em *docking* e QSAR.

Estrutura	*Docking score* (*Chemgauss4*)	QSAR consenso *score* [b]	QSAR DA *score* [c]
	-7,75	1,0	1,0
	-6,96	1,0	1,0
	-7,65	1,0	1,0
 *Redocking* [a] Suramina (inibidor co-cristalizado)	-4,50	-	-

[a] O *score* de *docking* obtido para o inibidor de PK co-cristalizado (suramina), após *redocking*, foi de -4,50 ; [b] Proporção dos modelos de QSAR que classificaram o composto como inibidor de PK (composto ativo); [c] Proporção dos modelos de QSAR em que o composto estava dentro DA.

5.3. Identificação de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania* spp.

Nesse estudo, foi realizada a triagem virtual de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania*, utilizando diferentes modelos *shape-based* baseados em inibidores conhecidos das duas enzimas. Além disso, foi aplicado na triagem um modelo consenso de QSAR, originado da combinação de cinco modelos de

QSAR individuais e usado para predição da atividade dos compostos contra amastigotas de *L. infantum*. A descrição da obtenção e preparo dos conjuntos de dados, bem como do processo de geração e validação desses modelos, está disponível na seção 4.3. Todas as etapas do estudo foram executadas conforme as boas práticas de CADD. A seguir, serão discutidos os resultados obtidos na validação e posterior aplicação desses modelos na triagem virtual.

5.3.1. Validação dos modelos *shape-based*

O melhor modelo *shape-based* baseado em inibidores de PK (modelo V), previamente discutido no item 6.2.1, foi selecionado para esse estudo. Além disso, foram gerados e validados cinco modelos *shape-based* baseados em inibidores conhecidos da enzima CYP51 de tripanossomatídeos (inibidores VNF, VFV, VNI, VNT e fluconazol). Na tabela 7, estão apresentados os resultados da validação desses modelos.

Tabela 7 - Validação dos modelos *shape-based* a partir do conjunto de dados de inibidores da enzima CYP51 e *decoys*, balanceados na proporção de 1:36.

Modelo	AUC	Top 1% hits				Top 5% hits				Top 10% hits			
		EF	BEDROC	SE	SP	EF	BEDROC	SE	SP	EF	BEDROC	SE	SP
VNF	0,75	15,77	0,39	0,50	0,53	6,76	0,34	0,51	0,62	4,57	0,38	0,53	0,65
VFV	**0,92**	**23,65**	**0,62**	**0,53**	**0,62**	**11,23**	**0,57**	**0,65**	**0,72**	**7,72**	**0,64**	**0,75**	**0,79**
VNI	0,67	6,16	0,21	0,52	0,52	2,54	0,14	0,56	0,53	1,98	0,18	0,57	0,54
VNT	0,64	1,89	0,05	0,53	0,50	1,55	0,08	0,56	0,51	1,61	0,13	0,58	0,52
fluconazol	0,76	16,67	0,48	0,50	0,57	8,73	0,43	0,50	0,66	5,72	0,48	0,51	0,70

AUC: área sob a curva ROC; EF: Fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; SE: sensibilidade; SP: especificidade; *Top x% hits*: métricas calculadas para os primeiros *x%* compostos da lista ordenada pelo *score TanimotoCombo*; O melhor modelo *shape-based*, usado posteriormente na triagem virtual, está destacado em negrito.

Conforme demonstrado na tabela 7, o modelo gerado a partir do ligante VFV apresentou desempenho superior aos demais modelos, com AUC igual a 0,92. Além disso, considerando os primeiros 1, 5 e 10% compostos ordenados, esse modelo apresentou bons valores de BEDROC (0,57-0,64), indicando a sua maior capacidade em ordenar compostos ativos no topo da lista, antes dos inativos. Os valores superiores de EF, também indicaram que o modelo VFV foi o mais eficiente em ordenar os

compostos ativos e inativos corretamente, em comparação a uma lista gerada aleatoriamente. Esse modelo também demonstrou boa SE e SP, alcançando valores acima de 0,7 no *top 10%*. O melhor modelo *shape-based*, descrito acima, está representado na figura 24.

As características e os pesos atribuídos ao modelo VFV foram baseados nas características do sítio ativo da CYP51 de tripanossomatídeos e nas interações necessárias para a sua inibição, descritas por Lepesheva e colaboradores (LEPESHEVA et al., 2010, 2015). A presença de grupos hidrofóbicos e a atribuição de peso 2 a eles, se justifica pelo fato de a enzima apresentar uma cavidade hidrofóbica projetada para o seu interior, além de possuir um canal de acesso ao substrato também capaz de acomodar ligantes lipofílicos. A atribuição de peso 3 à característica aceptor de ligação de hidrogênio, no N-3 do anel imidazólico, se justifica pela importante interação com o ferro do grupo heme (LEPESHEVA et al., 2010, 2015).

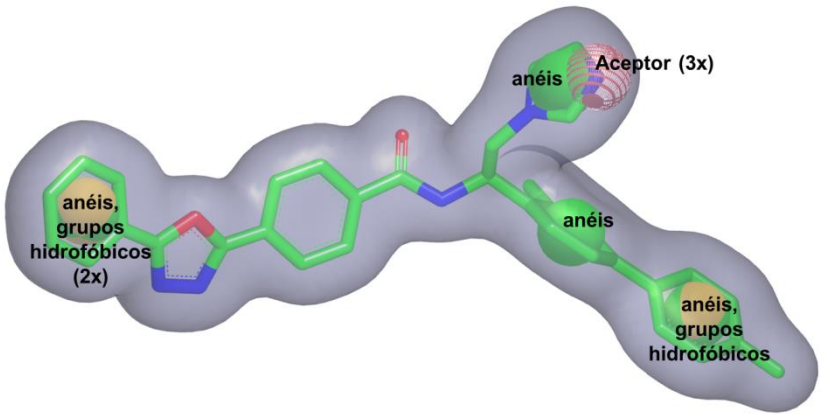


Figura 24 - Melhor modelo *shape-based* selecionado após a validação com inibidores de CYP51 e *decoys*. Esferas vermelhas correspondem à característica de aceptor de ligação de hidrogênio; Esferas verdes correspondem à presença de anéis; Esferas amarelas correspondem à presença de grupos hidrofóbicos; Foram atribuídos pesos 2 e 3 às características hidrofóbicas e aceptor de ligação de hidrogênio, respectivamente. O contorno em cor cinza representa o *shape* (forma e volume) da molécula.

5.3.2. Modelos de QSAR para predição da atividade contra amastigotas de *L. infantum*

O conjunto de dados contendo compostos testados contra amastigotas de *L. infantum* (ChEMBL2093840) previamente preparado, padronizado e balanceado (1:1) foi utilizado para a geração de modelos de QSAR utilizando quatro diferentes *fingerprints* moleculares, descritores CDK e o método de aprendizado de máquina RF. Esses modelos foram combinados em um modelo consenso, usado posteriormente na triagem virtual para predição de potenciais compostos ativos contra a forma amastigota de *L. infantum*. Na tabela 8, estão apresentados os resultados da avaliação do desempenho e preditividade dos modelos de QSAR, obtidos após validação cruzada externa *5-fold*.

Tabela 8 - Parâmetros estatísticos dos modelos de QSAR após validação cruzada externa *5-fold*.

Modelo	CCR	*k*	SE	SP	Cobertura
Atom Pair-RF	0,69	0,37	0,69	0,69	0,68*
Morgan-RF	0,67	0,34	0,64	0,69	0,66*
FeatMorgan-RF	**0,71**	**0,43**	**0,74**	**0,69**	**0,67***
Avalon-RF	0,67	0,34	0,67	0,70	0,69*
CDK-RF	0,68	0,36	0,68	0,68	0,74*
Consenso	**0,72**	**0,43**	**0,74**	**0,70**	**1,00**
Consenso Rigoroso	0,78	0,56	0,79	0,77	0,42**

RF: *Random Forest*; CCR: taxa de classificação correta; *k*: kappa; SE: sensibilidade; SP: especificidade; *Cobertura do modelo ao considerar o DA; **O consenso rigoroso levou em consideração apenas as predições de compostos que estavam dentro do DA de todos os cinco modelos; O modelo consenso foi obtido a partir da média aritmética das predições dos modelos individuais.

Conforme apresentado na tabela 8, o modelo FeatMorgan-RF apresentou o melhor desempenho dentre os modelos individuais, com CCR e SE acima de 0,7 e SP muito próximo a 0,7 (SP = 0,69). Além disso, esse modelo apresentou *k* igual a 0,43, o melhor dentre os modelos individuais. Os demais modelos também apresentaram desempenho bom ou aceitável, com valores de CCR, SE e SP sempre muito próximos a 0,7. O modelo consenso apresentou bom desempenho, com CCR, SE e SP acima de 0,7 e *k* igual a 0,43. Este modelo foi usado posteriormente na triagem virtual.

Na avaliação da robustez, através do método de *Y-randomization*, foi observado que a correlação entre a estrutura e atividade não ocorreu ao acaso, uma vez que todos os modelos apresentaram após as permutações, em média, CCR, SE e SP próximos a 0,5, além de valores de *k* próximos a zero.

O modelo FeatMorgan-RF, que apresentou o melhor desempenho dentre os modelos individuais, foi utilizado para interpretação através da análise de mapas de probabilidade predita (MPP), realizada para os compostos do conjunto de dados (RINIKER; LANDRUM, 2013b). Uma breve descrição do cálculo de MPP está disponível no item 6.2.2.

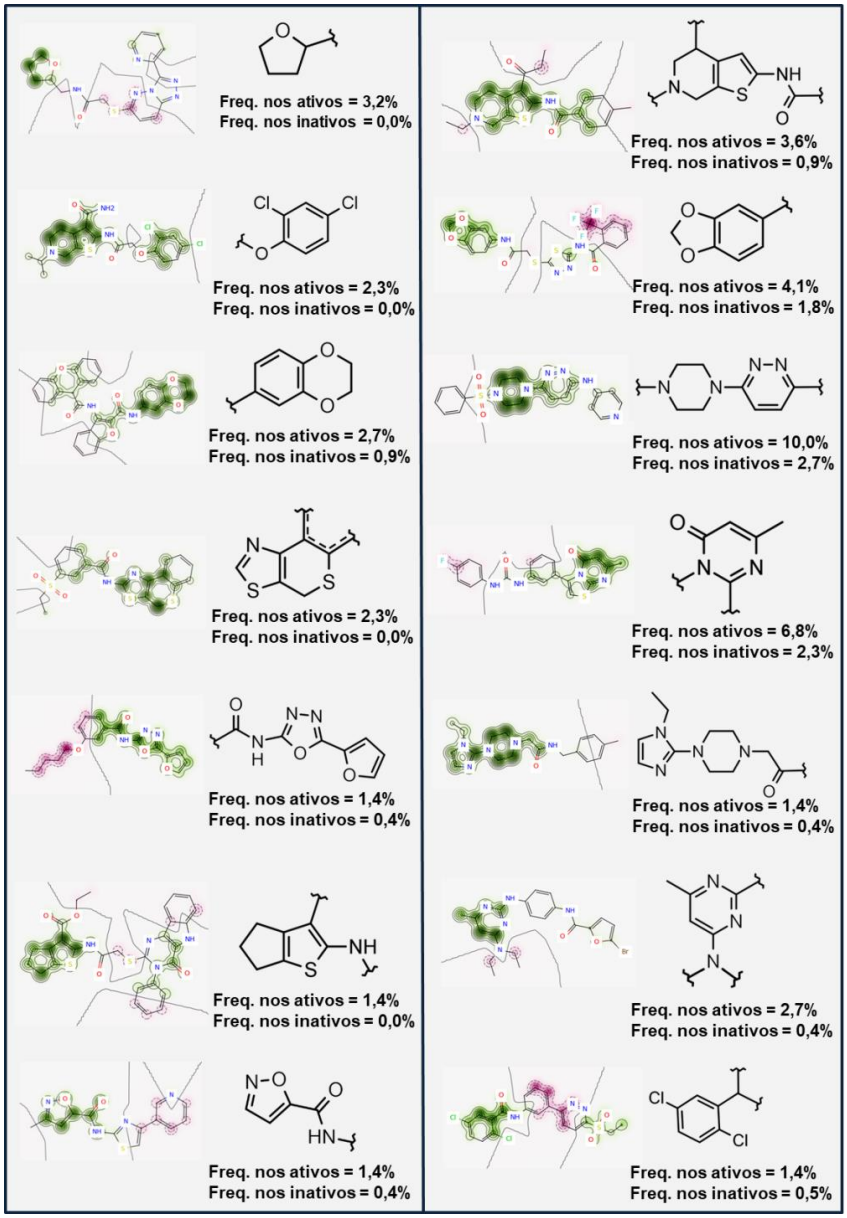


Figura 25- Fragmentos mais frequentes encontrados no conjunto de compostos ativos contra *L. infantum*, sua respectiva frequência e exemplos de MPP contendo esses

fragmentos. Os fragmentos em verde indicam contribuição positiva para a atividade, enquanto fragmentos em rosa indicam contribuição negativa. As partes não coloridas indicam contribuição neutra.

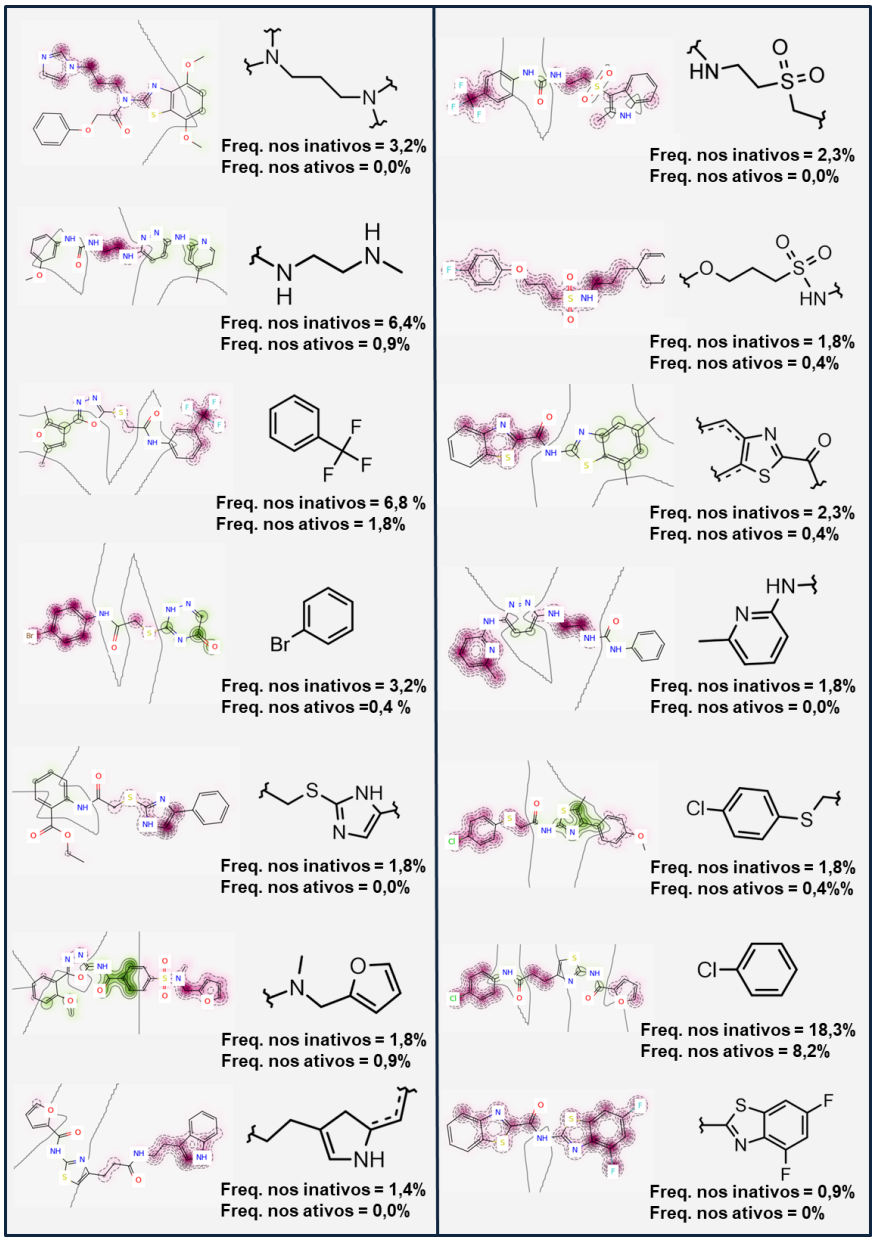


Figura 26- Fragmentos mais frequentes encontrados no conjunto de compostos inativos contra *L. infantum*, sua respectiva frequência e exemplos de MPP contendo esses fragmentos. Os fragmentos em verde indicam contribuição positiva para a atividade, enquanto fragmentos em rosa indicam contribuição negativa. As partes não coloridas indicam contribuição neutra.

Nas figuras 25 e 26, estão representados os MPP, os quais destacam os fragmentos com contribuição positiva e negativa encontrados com maior frequência nos compostos ativos e inativos contra *L. infantum*, respectivamente.

5.3.3. Triagem virtual

A base de dados comercial ChemBridge, contendo 1.063.926 compostos, foi submetida à triagem virtual *dual target* de novos potenciais inibidores das enzimas PK e CYP51 de *Leishmania*, utilizando modelos *shape-based* (baseados em inibidores conhecidos de PK e CPY51) e modelos de QSAR para predição da atividade contra amastigotas de *L. infantum*, previamente validados. O protocolo da triagem virtual está apresentado em detalhes na seção 4.3.4. Na figura 27, estão representados os resultados da triagem virtual.

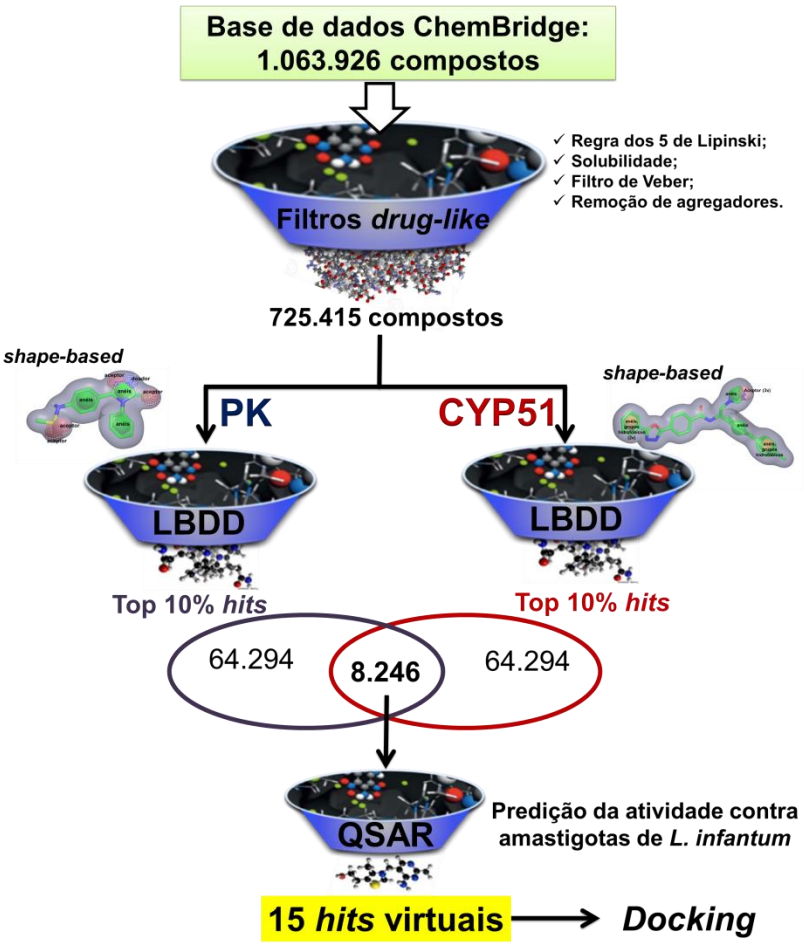


Figura 27 - Resultados da triagem virtual *dual-target* de potenciais inibidores das enzimas PK e CYP51 e com potencial de serem ativos contra as formas amastigotas de *Leishmania infantum*.

Conforme observado na Figura 27, o primeiro filtro utilizado na triagem virtual (filtro de propriedades *drug-like*) permitiu a seleção de 725.415 compostos a partir da base de dados ChemBridge, os quais foram submetidos, em seguida, aos melhores modelos *shape-based* correspondentes às enzimas PK e CYP51. Nessa etapa, foram selecionados os 10% melhores (top 10%) compostos de cada modelo (72.540 compostos), ranqueados de acordo com o *score TanimotoCombo,* formando duas listas que foram comparadas entre si, permitindo a identificação de 8.246 compostos em comum, ou seja, compostos com potencial de inibirem, concomitantemente, as enzimas PK e CYP51. Na próxima etapa, os 8.246 compostos foram submetidos ao modelo de QSAR consenso para predição da atividade contra amastigotas de *L. infantum*, levando à identificação de 15 *hits* virtuais. Os critérios para seleção dos *hits* virtuais foram: i) os compostos deveriam ser preditos como ativos em pelo menos 80% dos modelos de QSAR que compunham o consenso; ii) os compostos deveriam estar dentro do DA de mais que 50% dos modelos de QSAR. Ao final, os 15 *hits* virtuais foram adquiridos e submetidos à validação experimental. Além disso, esses compostos foram submetidos a estudos de *docking* nas duas enzimas. A tabela 9 contém as estruturas dos *hits* virtuais obtidos nesse estudo. A validação experimental desses 15 potenciais inibidores de PK e CYP51, preditos como ativos contra amastigotas de *L. infantum*, está sendo realizada em colaboração com o Prof. Dr. Jair Lage de Siqueira Neto (*Skaggs School of Pharmacy and Pharmaceutical Sciences* - UCSD, USA).

Tabela 9- Estruturas dos 15 potenciais inibidores das enzimas PK e CYP51 e preditos como ativos contra amastigotas de *L. infantum* selecionados após triagem virtual através de modelos *shape-based* e QSAR.

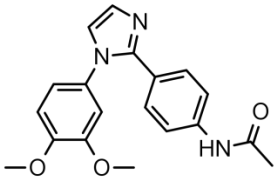
Estrutura	***Docking score* (PK) [a]**	***Docking score* (CYP51) [b]**	**QSAR consenso *score* [c]**	**QSAR DA *score* [d]**
	-5,73	-10,95	1,0	0,8

Tabela 9 (cont.) - Estruturas dos 15 potenciais inibidores das enzimas PK e CYP51 e preditos como ativos contra amastigotas de *L. infantum* selecionados após triagem virtual através de modelos *shape-based* e QSAR.

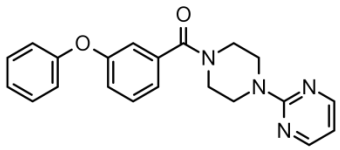
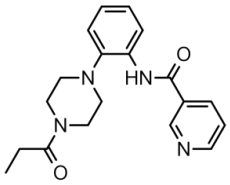
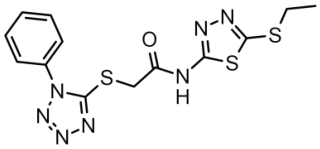
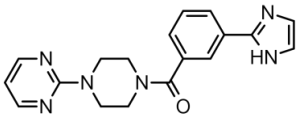
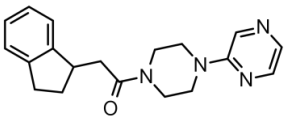
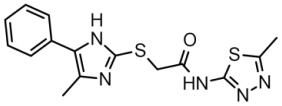
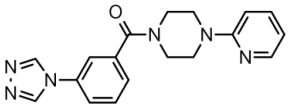
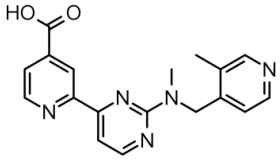
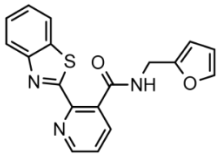
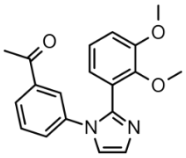
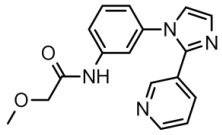
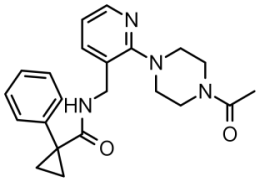
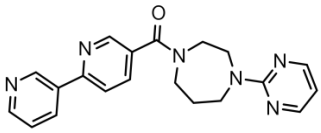
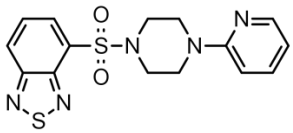
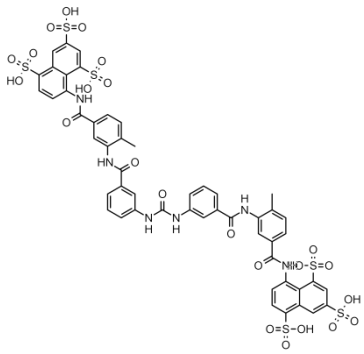
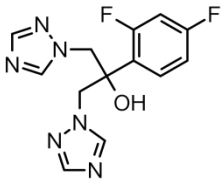
Estrutura	*Docking score* (PK) [a]	*Docking score* (CYP51) [b]	QSAR consenso *score* [c]	QSAR DA *score* [d]
	-5,42	-10,23	1,0	0,8
	-7,36	-12,37	1,0	0,8
	-5,60	-11,88	1,0	0,8
	-5,75	-11,57	1,0	0,8
	-7,46	-13,94	1,0	0,8
	-5,83	-13,23	1,0	0,8
	-6,09	-13,82	1,0	0,8
	-8,07	-14,95	1,0	0,6
	-4,86	-10,72	1,0	0,6
	-4,71	-15,41	1,0	0,6

Tabela 9 (cont.) - Estruturas dos 15 potenciais inibidores das enzimas PK e CYP51 e preditos como ativos contra amastigotas de *L. infantum* selecionados após triagem virtual através de modelos *shape-based* e QSAR.

Estrutura	*Docking score* (PK) [a]	*Docking score* (CYP51) [b]	QSAR consenso *score* [c]	QSAR DA *score* [d]
	-6,80	-14,01	1,0	0,6
	-4,95	-15,05	1,0	0,6
	-6,16	-14,12	1,0	0,6
	-5,17	-15,27	1,0	0,6
 *Redocking* da suramina [a] (inibidor de PK co-cristalizado)	-4,50	-	-	-
 *Redocking* do fluconazol [b] (inibidor de CYP51 co-cristalizado)	-	-14,40	-	-

[a] O *score* de *docking* obtido para o inibidor de PK co-cristalizado (suramina), após *redocking*, foi de -4,50; [b] O *score* de *docking* obtido para o inibidor de CYP51 co-cristalizado (fluconazol), após *redocking*, foi de -14,40; [c] Proporção dos modelos de QSAR que classificaram o composto como ativo contra amastigotas de *L. infantum*; [d] Proporção dos modelos de QSAR em que o composto estava dentro DA.

6. DISCUSSÃO

No primeiro estudo apresentado nesta tese, foi realizada uma triagem virtual baseada em QSAR, seguida pela validação experimental *in vitro* de novos *hits* potentes e seletivos contra *T. cruzi* intracelular.

De acordo com Katsuno e colaboradores (2015), apoiados nas definições feitas em conjunto por autoridades do GHIT (*Global Health Innovative Technology Fund)* e do DNDi, o critério para a seleção de *hits* para doença de Chagas se baseia na obtenção de compostos com potência micromolar *in vitro* (< 10 µM) contra *T. cruzi* intracelular e seletividade maior que 10 (KATSUNO et al., 2015). Considerando esse critério, nesse trabalho, 13 novos *hits* contra *T. cruzi* intracelular foram identificados. Outro critério levantado pelos autores é a necessidade dos *hits* passarem por filtros *drug-like* básicos, como por exemplo, conformidade com a regra dos 5 de Lipinski (KATSUNO et al., 2015). Essa recomendação reforça a importância do filtro *drug-like* utilizado como primeira etapa da triagem virtual desse estudo.

Levando em consideração as futuras etapas desse trabalho, relacionadas à otimização *hit-to-lead*, vários critérios relacionados às propriedades físico-químicas, toxicológicas, farmacocinéticas e de metabolismo assumem grande importância. Nesse sentido, foi realizada uma avaliação *in silico* multi-paramétrica dos compostos testados *in vitro*, a fim de se analisar equilíbrio entre potência, seletividade e diversas propriedades ADMET. O uso de ferramentas *in silico* para predição dessas propriedades tem crescido nos últimos anos, uma vez que permite, em fases preliminares, a priorização de compostos com menores chances de atrito em fases mais avançadas do planejamento de fármacos, além de fornecer informações importantes que podem guiar o processo de otimização dos compostos (BRAGA et al., 2015; SANDERS et al., 2017).

Dentre os *hits* obtidos nesse estudo, destacou-se o *LabMol-134* (EC_{50} = 0.3 µM; SI = 136.8; logP = 1.82), que apresentou um bom equilíbrio entre potência, seletividade, propriedades físico-químicas e ADMET. Outros *hits* como o *LabMol-104*, *LabMol-125, LabMol-128* e *LabMol-126* também apresentaram resultados promissores. Em futuros estudos, o *LabMol-134*, juntamente com os *hits* mais promissores, serão submetidos a ensaios *in vivo* e avaliação experimental de propriedades ADMET, podendo ser otimizados para o desenvolvimento de novos candidatos a protótipos de fármacos contra a doença de Chagas.

No segundo estudo demonstrado nesta tese, foi realizada a triagem virtual de novos potenciais inibidores da enzima PK de *Leishmania*. Essa enzima, situada fora dos glicossomos, participa da via glicolítica do parasito e apresenta papel importante na produção de ATP e piruvato (HARRIS; MITCHELL; MORRIS, 2014). Evidências indicam que a PK é expressa tanto na forma promastigota, quanto amastigota (BROTHERTON et al., 2010). Além disso, estudos anteriores identificaram inibidores da PK que também apresentaram atividade *in vitro* contra parasitos do gênero *Leishmania* (NOWICKI et al., 2008).

Seguindo as recomendações propostas pelo GHIT e DNDi, que reforçam a importância da identificação de compostos ativos contra DTNs e que ao mesmo tempo apresentem características *drug-like* (KATSUNO et al., 2015), o primeiro filtro da triagem virtual foi baseado na regra dos 5 de Lipinski, no filtro de Veber e na solubilidade dos compostos (LIPINSKI et al., 2001; VEBER et al., 2002).

A aplicação de modelos *shape-based*, durante a triagem, se apoiou na vantagem desses modelos em não considerarem apenas a sobreposição de grupos específicos (doadores e aceptores de ligação de hidrogênio, anéis, grupos hidrofóbicos), observada em outros métodos, mas a inclusão da sobreposição da forma e volume moleculares, representadas pela função *ShapeTanimoto* (HAWKINS; SKILLMAN; NICHOLLS, 2007). A combinação dessas duas sobreposições (sobreposição de grupos e da forma e volume moleculares) já demonstrou ser eficiente na identificação de compostos ativos, com resultados comparáveis a métodos SBDD (MCGAUGHEY et al., 2007). Esta estratégia permitiu a seleção de compostos com boa sobreposição a inibidores já conhecidos da enzima PK de *Leishmania*, tanto em termos de forma (*shape*) quanto de características moleculares específicas.

Na etapa de *docking*, foram selecionados os compostos melhor classificados pela função de *score Chemgauss4*, que se baseia no reconhecimento de ligações de hidrogênio entre ligante e proteína, incluindo sua geometria, além de interações metálicas, efeitos de dessolvatação e a complementaridade entre volume do ligante e o volume do sítio ativo (MCGANN, 2012). Dessa forma, foram selecionados compostos com potencial de apresentarem boa afinidade ao sítio de ATP da PK de *Leishmania*.

O modelo consenso de QSAR, aplicado como filtro final, priorizou os compostos com maiores chances de inibirem a PK de *Leishmania*. Além disso, foram gerados os mapas de probabilidade predita (MPP), que foram aplicados para identificação de fragmentos moleculares com contribuição positiva ou negativa para a

inibição da enzima PK (RINIKER; LANDRUM, 2013b). Os MPP são um exemplo de interpretação local de modelos de QSAR (POLISHCHUK, 2017).

Segundo Vitaku e colaboradores (2014), dentre os fármacos aprovados pelo FDA, considerando apenas moléculas pequenas e únicas, 59% apresentam átomos de nitrogênio em heterociclos. Além disso, os anéis tiazol e 1,2,4-triazol estão na lista dos 25 heterociclos contendo nitrogênio mais frequentemente observados em fármacos aprovados (VITAKU; SMITH; NJARDARSON, 2014). Considerando os 14 potenciais inibidores de PK identificados nesse estudo, oito deles apresentam pelo menos um desses heterociclos, reforçando a capacidade dos modelos utilizados em identificar compostos com características de fármacos. É importante destacar que os fragmentos contendo esses dois heterociclos foram identificados como favoráveis à inibição de PK, durante as análises dos MPP. Os *hits* virtuais selecionados nesse estudo estão em fase de validação experimental *in vitro* contra *L. infantum*, realizada em colaboração com o Dr. André G. Tempone Cardoso (Instituto Adolfo Lutz - SP).

No terceiro estudo mostrado nesta tese, foi realizada a triagem virtual de novos potenciais inibidores *dual-target* das enzimas PK e CYP51 de *Leishmania*, utilizando modelos *shape-based* baseados em inibidores conhecidos das duas enzimas. Além disso, um modelo consenso de QSAR foi aplicado na triagem, usado para predição da atividade dos compostos contra amastigotas de *L. infantum*.

Recentemente, tem sido observado o crescente interesse no planejamento de fármacos cuja atividade se baseia na interação, concomitante, com dois ou mais alvos biológicos. Essa abordagem tem sido proposta como alternativa às tradicionais estratégias de planejamento direcionadas a apenas um alvo molecular (LAURIA et al., 2016). Diversas estratégias *in silico* podem ser utilizadas no planejamento de ligantes que atuem em mais de um alvo biológico, como por exemplo, o planejamento de fármacos baseado em fragmentos e a triagem virtual em paralelo, utilizando diferentes alvos (BOTTEGONI et al., 2012; ZHOU et al., 2016). No estudo apresentado nesta tese, foi realizada uma triagem virtual *dual-target* em paralelo, utilizando dois modelos *shape-based* diferentes gerados a partir de inibidores conhecidos das enzimas PK e CYP51 de *Leishmania*. Os compostos melhor ranqueados (top 10%) para cada enzima foram comparados. Aqueles presentes concomitantemente na lista de *hits* das duas enzimas foram então selecionados.

Ao longo dos anos, tem-se observado uma crescente aplicação de ensaios fenotípicos (baseados em células inteiras) para identificação de compostos bioativos,

aplicando-se uma diversidade de métodos e tecnologias (CHATELAIN; IOSET, 2018). Esses avanços têm contribuído para a geração de um crescente volume de dados de ensaios, os quais são depositados em bases de dados como o ChEMBL e PubChem BioAssay (GAULTON et al., 2017; WANG et al., 2017; ZHENG; THORNE; MCKEW, 2013). Esses dados podem ser úteis para geração de modelos de QSAR preditivos que podem ser usados na identificação de novos potenciais compostos bioativos. Dessa forma, como filtro final da triagem virtual executada nesse estudo, foi utilizado um modelo de QSAR consenso para predição da atividade dos compostos contra amastigotas de *L. infantum*, construído a partir de dados de ensaio fenotípico contra o parasito, depositado na base de dados ChEMBL. Além disso, foram gerados mapas de probabilidade predita (MPP), utilizados na identificação de fragmentos moleculares com contribuição positiva ou negativa para atividade contra *L. infantum*. Esses mapas poderão ser usados em futuros estudos de otimização dos compostos (RINIKER; LANDRUM, 2013b)

Ao final da triagem virtual, foram selecionados compostos que reuniam, concomitantemente, características similares àquelas de inibidores já conhecidos de PK e CYP51, usados na geração dos modelos *shape-based*. Os estudos de *docking*, realizados após a triagem virtual, evidenciaram o potencial desses compostos em apresentar afinidade pelas duas enzimas. De modo geral, os *hits* virtuais identificados apresentaram aceptores de ligação de hidrogênio, capazes de realizar interações com o grupo heme da enzima CYP51, além de grupos hidrofóbicos, capazes de realizar interações importantes, permitindo a acomodação dos ligantes na cavidade hidrofóbica da enzima (LEPESHEVA et al., 2010, 2015). Além disso, todos os compostos apresentaram anéis capazes de estabelecer importantes interações *π-stacking* com os aminoácidos His54 e Tyr59, localizados na entrada do sítio de ligação de ATP/ADP da enzima PK (MORGAN et al., 2011a). A presença de uma carbonila, na extremidade da molécula ou funcionando como *linker* entre anéis, também foi percebida em praticamente todos os *hits* virtuais, mimetizando a carbonila presente em inibidores conhecidos de PK, permitindo interações importantes no sítio ativo (MORGAN et al., 2011b). A validação experimental dos potenciais inibidores *dual-target*, preditos como ativos contra amastigotas de *L. infantum*, está sendo realizada em colaboração com o Dr. Jair Lage de Siqueira Neto (*Skaggs School of Pharmacy and Pharmaceutical Sciences* - UCSD, USA).

7. CONCLUSÕES

Nesse trabalho, foram utilizadas diferentes estratégias para a identificação de novos compostos ativos contra tripanossomatídeos baseando-se na integração de métodos de quimioinformática, modelagem molecular e avaliação experimental. Foram realizados três estudos: i) triagem virtual baseada em QSAR para identificação de novos compostos ativos e seletivos contra *T. cruzi* intracelular, seguida de validação experimental; ii) triagem virtual integrando métodos de LBDD, SBDD e QSAR para identificação de novos potenciais inibidores da enzima PK de *Leishmania* spp.; iii) triagem virtual *dual-target* baseada em métodos de LBDD e QSAR para identificação de novos potenciais inibidores das enzimas PK e CYP51, também preditos como potenciais compostos ativos contra amastigotas de *L. infantum*.

No primeiro trabalho, modelos de QSAR foram gerados e validados, seguindo as boas práticas de validação sugeridas na literatura. Esses modelos foram aplicados com sucesso na triagem virtual de novos compostos potencialmente ativos e seletivos contra *T. cruzi* intracelular. Após a validação experimental *in vitro*, foram identificados 13 compostos com atividade promissora e seletividade contra o parasito, sendo sete deles ativos em concentrações submicromolares. Dada a importância dos múltiplos parâmetros ADMET e físico-químicos no desenvolvimento de novos candidatos a protótipos de fármacos, foi realizada, para todos os compostos testados, uma análise *in silico* desses múltiplos parâmetros. Ao final, foram identificados compostos bastante promissores, dentre os quais se destacou um composto com condições de prosseguir em estudos *in vivo*. A validação experimental dos parâmetros ADMET e ensaios *in vivo* deverão ser realizados em futuros estudos.

No segundo trabalho, foram gerados e validados modelos baseados na forma e volume moleculares (*shape-based*), modelos baseados em *docking* e modelos de QSAR para dados de inibidores da enzima PK de *Leishmania*. Foram obtidos modelos com capacidade de serem usados na triagem virtual de novos potenciais inibidores da PK de *Leishmania*. Após a triagem virtual, 14 novos inibidores potenciais dessa enzima foram identificados e adquiridos. Esses compostos se encontram em fase de validação experimental, cujos resultados serão publicados em breve, juntamente aos resultados computacionais.

No último trabalho, foram utilizados modelos *shape-based* (baseados em inibidores das enzimas PK e CYP51) e de QSAR (baseados em dados de ensaio fenotípico contra *L. infantum*), gerados e validados de acordo com as recomendações para estudos de CADD, os quais foram aplicados em uma triagem virtual *dual-target* que permitiu a identificação de 15 potenciais inibidores das enzimas PK e CYP51, também preditos como ativos contra amastigotas de *L. infantum*. Esses compostos estão sendo avaliados experimentalmente.

Portanto, a presente tese de doutorado apresentou três exemplos da aplicação integrada de métodos computacionais e experimentais, para acelerar e otimizar a identificação de compostos bioativos contra doenças tropicais negligenciadas. Além disso, foi demonstrada a aplicação de diferentes abordagens e estratégias do CADD como a triagem virtual, SBDD, LBDD e modelagem por QSAR. Nesta última, foi possível observar a capacidade da modelagem por QSAR de, além de predizer a atividade de novos compostos, também ser aplicada à interpretação e sugestão de fragmentos importantes para a atividade biológica, através dos mapas de probabilidade predita, que podem ser úteis em futuros estudos de otimização dos compostos identificados.

Os compostos mais promissores serão futuramente avaliados em experimentos *in vitro* e *in vivo* para investigação do mecanismo de ação, parâmetros farmacocinéticos e de toxicidade, e eficácia *in vivo*, para prosseguirem no processo de otimização do ligante ao protótipo, e poderão contribuir na descoberta de novos fármacos contra essas doenças tão importantes, porém, negligenciadas.

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. e0004349, 2016.

ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. **The American Statistician**, v. 46, n. 3, p. 175–185, 1992.

ALVAR, J.; YACTAYO, S.; BERN, C. Leishmaniasis and poverty. **Trends in parasitology**, v. 22, n. 12, p. 552–557, 2006.

ANDERSON, S. Graphical representation of molecules and substructure-search queries in MACCS. **Journal of Molecular Graphics**, v. 2, n. 3, p. 83–90, 1984.

ANDRIANI, G. et al. Antitrypanosomal Lead Discovery: Identification of a Ligand-Efficient Inhibitor of Trypanosoma cruzi CYP51 and Parasite Growth. **Journal of Medicinal Chemistry**, v. 56, n. 6, p. 2556–2567, 2013.

AVORN, J. The $2.6 billion pill-methodologic and policy considerations. **The New England journal of medicine**, v. 372, n. 20, p. 1877–1879, 2015.

BAJORATH, J. Computational chemistry in pharmaceutical research: at the crossroads. **Journal of computer-aided molecular design**, v. 26, n. 1, p. 11–12, 2012.

BANKS, J. L. et al. **Integrated Modeling Program, Applied Chemical Theory (IMPACT)Journal of Computational Chemistry**, 2005.

BARROS-ALVAREZ, X. et al. Glycosomal targets for anti-trypanosomatid drug discovery. **Current medicinal chemistry**, v. 21, n. 15, p. 1679–706, 2014.

BEISKEN, S. et al. KNIME-CDK: Workflow-driven cheminformatics. **BMC bioinformatics**, v. 14, p. 257, 2013.

BENTO, A. P. et al. The ChEMBL bioactivity database: an update. **Nucleic Acids Research**, v. 42, n. D1, p. 1083–1090, 2014.

BERMUDEZ, J. et al. Current drug therapy and pharmaceutical challenges for Chagas disease. **Acta Tropica**, v. 156, p. 1–16, 2016.

BERN, C. Chagas' Disease. **New England Journal of Medicine**, v. 373, n. 5, p. 456–466, 2015.

BERTHOLD, M. R. et al. KNIME: The Konstanz Information Miner. **SIGKDD Explorations**, v. 11, n. 1, p. 26–31, 2009.

BOELAERT, M.; SUNDAR, S. Leishmaniasis. In: FARRAR, J. et al. (Eds.). . **Manson's Tropical Diseases**. 23. ed. Philadelphia: Elsevier Saunders, 2014. p. 631–651.

BOGITSH, B. J.; CARTER, C. E.; OELTMANN, T. N. Blood and Tissue Protozoa I: Hemoflagellates. In: BOGITSH, B. J.; CARTER, C. E.; OELTMANN, T. N. (Eds.). . **Human Parasitology**. 4. ed. Oxford: Academic Press: Elsevier, 2013. p. 85–113.

BOTTEGONI, G. et al. The role of fragment-based and computational methods in polypharmacology. **Drug discovery today**, v. 17, n. 1–2, p. 23–34, jan. 2012.

BRAGA, R. C. et al. Virtual Screening Strategies in Medicinal Chemistry: The state of the art and current challenges. **Current topics in medicinal chemistry**, v. 14, n. 16, p. 1899–1912, 2014.

BRAGA, R. C. et al. Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity. **Molecular Informatics**, v. 34, n. 10, p. 698–701, 2015.

BRAGA, R. C.; ANDRADE, C. H. Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good are They? **Current topics in medicinal chemistry**, v. 13, n. 9, p. 1127–1138, 2013.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BRINGAUD, F.; RIVIÈRE, L.; COUSTOU, V. Energy metabolism of trypanosomatids: adaptation to available carbon sources. **Molecular and biochemical parasitology**, v. 149, n. 1, p. 1–9, 2006.

BROTHERTON, M. et al. Analysis of Stage-Specific Expression of Basic Proteins in Leishmania infantum. **Journal of Proteome Research**, v. 9, n. 8, p. 3842–3853, 2010.

BUCKNER, F. S. Sterol 14-Demethylase Inhibitors for Trypanosoma cruzi Infections. In: MAJUMDER, H. K. (Ed.). . **Drug Targets in Kinetoplastid Parasites**. New York:

Springer Science+Business Media, 2008. p. 61–80.

BUCKNER, F. S. et al. Pharmacological Characterization, Structural Studies, and In Vivo Activities of Anti-Chagas Disease Lead Compounds Derived from Tipifarnib. **Antimicrobial Agents and Chemotherapy**, v. 56, n. 9, p. 4914–4921, 2012.

CALVET, C. M. et al. 4-Aminopyridyl-Based CYP51 Inhibitors as Anti- Trypanosoma cruzi Drug Leads with Improved Pharmacokinetic Profile and in Vivo Potency. **Journal of Medicinal Chemistry**, v. 57, n. 16, p. 6989–7005, 2014.

CARHART, R. E.; SMITH, D. H.; VENKATARAGHAVAN, R. Atom pairs as molecular features in structure-activity studies: definition and applications. **Journal of Chemical Information and Computer Science**, v. 25, n. 2, p. 64–73, 1985.

CASTILLO, E. et al. The kinetoplastid chemotherapy revisited: current drugs, recent advances and future perspectives. **Current medicinal chemistry**, v. 17, n. 33, p. 4027–51, 2010.

CASTRO, J. E. Z. Enfermedad de Chagas y otras tripanosomiasis. In: FLORES, M. A. B. (Ed.). . **Parasitología Médica**. 3. ed. México: McGraw-Hill, 2011. p. 81–94.

CDC. **DPDx - Laboratory Identification of Parasitic Diseases of Public Health Concern - Leishmaniasis**. Disponível em: <http://www.cdc.gov/dpdx/leishmaniasis/gallery.html#promastigotes>. Acesso em: 28 jul. 2016.

CDC. **Leishmaniasis Biology**. Disponível em: <http://www.cdc.gov/parasites/leishmaniasis/biology.html>. Acesso em: 2 ago. 2016.

CDC. **Chagas Disease Detailed FAQs**. Disponível em: <https://www.cdc.gov/parasites/chagas/gen_info/detailed.html>. Acesso em: 10 jan. 2018a.

CDC. **American Trypanosomiasis - Parasite Biology**. Disponível em: <https://www.cdc.gov/dpdx/trypanosomiasisamerican/index.html>. Acesso em: 14 jan. 2018b.

CHATELAIN, E.; IOSET, J.-R. Phenotypic screening approaches for Chagas disease drug discovery. **Expert Opinion on Drug Discovery**, v. 13, n. 2, p. 141–153, 2018.

CHAWLA, B.; MADHUBALA, R. Drug targets in Leishmania. **Journal of parasitic diseases : official organ of the Indian Society for Parasitology**, v. 34, n. 1, p. 1–13, 2010.

CHERKASOV, A. et al. QSAR modeling: where have you been? Where are you going to? **Journal of medicinal chemistry**, v. 57, n. 12, p. 4977–5010, 2014.

CHOI, J. Y.; PODUST, L. M.; ROUSH, W. R. Drug Strategies Targeting CYP51 in Neglected Tropical Diseases. **Chemical Reviews**, v. 114, n. 22, p. 11242–11271, 2014.

CLAYTON, J. Chagas disease: pushing through the pipeline. **Nature**, v. 465, n. 7301, p. S12–S15, 2010.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

COHEN, N. C. **Guidebook on molecular modeling in drug design**. San Diego: Academic Press, 1996.

CONSONNI, V.; TODESCHINI, R. Molecular descriptors. In: PUZYN, T.; LESZCZYNSKI, J.; CRONIN, M. T. D. (Eds.). . **Recent advances in QSAR studies - methods and applications**. 1. ed. New York: Springer Science+Business Media B.V., 2010. p. 29–102.

CRONIN, M. T. D. Quantitative structure-activity relationships (QSARs)-applications and methodology. In: PUZYN, T.; LESZCZYNSKI, J.; CRONIN, M. T. D. (Eds.). . **Recent advances in QSAR studies - methods and applications**. 1. ed. New York: Springer Science+Business Media B.V., 2010. p. 3–11.

CROWTHER, G. J. et al. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. **PLoS neglected tropical diseases**, v. 4, n. 8, p. e804, 2010.

CRUZ, A. K. et al. Current treatment and drug discovery against Leishmania spp. and Plasmodium spp.: a review. **Current drug targets**, v. 10, n. 3, p. 178–92, 2009.

DAVID, C. V; CRAFT, N. Cutaneous and mucocutaneous leishmaniasis. **Dermatologic therapy**, v. 22, n. 6, p. 491–502, 2012.

DEARDEN, J. C.; CRONIN, M. T. D.; KAISER, K. L. E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). **SAR and QSAR in environmental research**, v. 20, n. 3–4, p. 241–266, 2009.

DILL, J. D. et al. **Search and Retrieval Using an Automated Molecular Access System.**182nd National Meeting of the American Chemical Society. **Anais**...New York: 1981

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. 151–85, 2003.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78, 2012.

DUFFY, B. C. et al. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. **Bioorganic & medicinal chemistry**, v. 20, n. 18, p. 5324–5342, 2012.

DURANT, J. L. et al. Reoptimization of MDL keys for use in drug discovery. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 6, p. 1273–1280, 2002.

ELEBRING, T.; GILL, A.; PLOWRIGHT, A. T. What is the most important approach in current drug discovery: doing the right things or doing things right? **Drug discovery today**, v. 17, n. 21–22, p. 1166–9, 2012.

EMPEREUR-MOT, C. et al. Predictiveness curves in virtual screening. **Journal of Cheminformatics**, v. 7, n. 1, p. 1–17, 2015.

FDA. **FDA approves first U.S. treatment for Chagas disease**. Disponível em: <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm573942.htm>. Acesso em: 12 jan. 2018.

FIELD, M. C. et al. Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. **Nature reviews. Microbiology**, v. 15, n. 4, p. 217–231, 2017.

FOURCHES, D.; MURATOV, E. N.; TROPSHA, A. Trust, But Verify II: A Practical Guide to Chemogenomics Data Curation. **Journal of Chemical Information and Modeling**, p. acs.jcim.6b00129, 2016.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of chemical information and modeling**, v. 50, n. 7, p. 1189–204, 2010.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Curation of chemogenomics data. **Nature Chemical Biology**, v. 11, n. 8, p. 535–535, 2015.

FRANCO-PAREDES, C. American Trypanosomiasis: Chagas Disease. In: FARRAR, J. et al. (Eds.). . **Manson's Tropical Diseases**. 23. ed. Philadelphia: Elsevier, 2014. p. 622–630.

FREITAS-JUNIOR, L. H. et al. Visceral leishmaniasis treatment: What do we have, what do we need and how to deliver it? **International journal for parasitology: Drugs and drug resistance**, v. 2, p. 11–19, 2012.

GASHAW, I. et al. What makes a good drug target? **Drug discovery today**, v. 16, n. 23–24, p. 1037–43, 2011.

GASPAR, L. et al. Current and Future Chemotherapy for Chagas Disease. **Current Medicinal Chemistry**, v. 22, n. 37, p. 4293–4312, 2015.

GASTEIGER, J. Chemoinformatics: Achievements and Challenges, a Personal View. **Molecules**, v. 21, n. 2, p. 151, 2016.

GAULTON, A. et al. The ChEMBL database in 2017. **Nucleic acids research**, v. 45, n. D1, p. D945–D954, 2017.

GEDECK, P.; ROHDE, B.; BARTELS, C. QSAR − How Good Is It in Practice? Comparison of Descriptor Sets on an Unbiased Cross Section of Corporate Data Sets. **Journal of Chemical Information and Modeling**, v. 46, n. 5, p. 1924–1936, 2006.

GRANT, J. A.; GALLARDO, M. A.; PICKUP, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. **Journal of computational chemistry**, v. 17, n. 14, p. 1653–1666, 15 nov. 1996.

GUALDR, M. et al. Function of Glycosomes in the Metabolism of Trypanosomatid Parasites and the Promise of Glycosomal Proteins as Drug Targets. In: JÄGER, T.; KOCH, O.; FLOHÉ, L. (Eds.). . **Trypanosomatid Diseases: Molecular Routes to Drug Discovery**. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2013.

p. 121–151.

GUNATILLEKE, S. S. et al. Diverse Inhibitor Chemotypes Targeting Trypanosoma cruzi CYP51. **PLoS Neglected Tropical Diseases**, v. 6, n. 7, p. e1736, 2012.

GUNN, A.; PITT, S. J. Parasitic protozoa, fungi and plants: Phylum Kinetoplastida. In: GUNN, A.; PITT, S. J. (Eds.). . **Parasitology: An Integrated Approach**. 1. ed. Chichester: John Wiley & Sons, Ltd, 2012. p. 62–81.

HARGROVE, T. Y. et al. CYP51 structures and structure-based development of novel, pathogen-specific inhibitory scaffolds. **International Journal for Parasitology: Drugs and Drug Resistance**, v. 2, p. 178–186, 2012.

HARRIS, M. T.; MITCHELL, W. G.; MORRIS, J. C. Targeting protozoan parasite metabolism: glycolytic enzymes in the therapeutic crosshairs. **Current medicinal chemistry**, v. 21, n. 15, p. 1668–1678, 2014.

HAWKINS, P. C. D. et al. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. **Journal of chemical information and modeling**, v. 50, n. 4, p. 572–584, 2010.

HAWKINS, P. C. D.; SKILLMAN, A G.; NICHOLLS, A. Comparison of shape-matching and docking as virtual screening tools. **Journal of medicinal chemistry**, v. 50, n. 1, p. 74–82, 2007.

HECKER, E. A. et al. Use of catalyst pharmacophore models for screening of large combinatorial libraries. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 5, p. 1204–1211, 2002.

HERWALDT, B. L. Leishmaniasis. **Lancet**, v. 354, n. 9185, p. 1191–9, 1999.

HOUWELING, T. A. J. et al. Socioeconomic Inequalities in Neglected Tropical Diseases: A Systematic Review. **PLOS Neglected Tropical Diseases**, v. 10, n. 5, p. e0004546, 2016.

HUANG, N.; SHOICHET, B. K.; IRWIN, J. J. Benchmarking Sets for Molecular Docking. **Journal of Medicinal Chemistry**, v. 49, n. 23, p. 6789–6801, 2006.

HUGHES, J. P. et al. Principles of early drug discovery. **British journal of pharmacology**, v. 162, n. 6, p. 1239–1249, 2011.

IMMING, P. Medicinal Chemistry: Definitions and Objectives, Drug Activity Phases, Drug Classification Systems. In: WERMUTH, C. G. (Ed.). . **The Practice of Medicinal Chemistry**. 3. ed. [s.l.] Academic Press, 2011. p. 63–72.

IRWIN, J. J.; SHOICHET, B. K. ZINC--a free database of commercially available compounds for virtual screening. **Journal of chemical information and modeling**, v. 45, n. 1, p. 177–82, 2005.

IVENS, A. C. et al. The genome of the kinetoplastid parasite, Leishmania major. **Science (New York, N.Y.)**, v. 309, n. 5733, p. 436–42, 2005.

JACOBSSON, M. et al. Improving Structure-Based Virtual Screening by Multivariate Analysis of Scoring Data. **Journal of Medicinal Chemistry**, v. 46, n. 26, p. 5781–5789, 2003.

JAIN, K.; JAIN, N. K. Novel therapeutic strategies for treatment of visceral leishmaniasis. **Drug discovery today**, v. 18, n. 23–24, p. 1272–1281, 2013.

JORGENSEN, W. L. The many roles of computation in drug discovery. **Science (New York, N.Y.)**, v. 303, n. 5665, p. 1813–1818, 2004.

KAPETANOVIC, I. M. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. **Chemico-biological interactions**, v. 171, n. 2, p. 165–176, 2008.

KATSUNO, K. et al. Hit and lead criteria in drug discovery for infectious diseases of the developing world. **Nature Reviews Drug Discovery**, v. 14, n. 11, p. 751–758, 2015.

KAYE, P.; SCOTT, P. Leishmaniasis: complexity at the host-pathogen interface. **Nature reviews. Microbiology**, v. 9, n. 8, p. 604–15, 2011.

KIRCHHOFF, L. V. Trypanosoma Species (AmericanTrypanosomiasis, Chagas' Disease): Biology of Trypanosomes. In: MANDELL, G. L.; BENETT, J. E.; DOLIN, R. (Eds.). . **Principles and Practice of Infectious Diseases**. 7. ed. Philadelphia: Churchill Livingstone: Elsevier, 2010. p. 3481–3488.

KIRCHHOFF, L. V. American Trypanosomiasis (Chagas Disease). In: GUERRANT, R. L.; WALKER, D. H.; WELLER, P. F. (Eds.). . **Tropical Infectious Diseases: Principles, Pathogens and Practice**. 3. ed. New York: Saunders:Elsevier, 2011. p. 689–695.

KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. **Drug Discovery Today**, v. 11, n. July, p. 580–594, 2006.

KLOTZ, S. A. et al. Kissing Bugs in the United States: Risk for Vector-Borne Disease in Humans. **Environmental Health Insights**, v. 8s2, p. EHI.S16003, 2014.

KOBETS, T.; GREKOV, I.; LIPOLDOVA, M. Leishmaniasis: prevention, parasite detection and treatment. **Current medicinal chemistry**, v. 19, n. 10, p. 1443–74, 2012.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. **Artificial Intelligence Review**, v. 26, n. 3, p. 159–190, 2007.

KRAUS, J. M. et al. Second Generation Analogues of the Cancer Drug Clinical Candidate Tipifarnib for Anti-Chagas Disease Drug Discovery. **Journal of Medicinal Chemistry**, v. 53, n. 10, p. 3887–3898, 2010.

LANDRUM, G. **RDKit Documentation.Release 2017.09.1**, 2017.

LAURIA, A. et al. Drugs Polypharmacology by In Silico Methods: New Opportunities in Drug Discovery. **Current pharmaceutical design**, v. 22, n. 21, p. 3073–3081, 2016.

LAVECCHIA, A. Machine-learning approaches in drug discovery: methods and applications. **Drug Discovery Today**, v. 20, n. 3, p. 318–331, 2014.

LEPESHEVA, G. I. et al. CYP51: A major drug target in the cytochrome P450 superfamily. **Lipids**, v. 43, n. 12, p. 1117–25, 2008.

LEPESHEVA, G. I. et al. Structural insights into inhibition of sterol 14α-demethylase in the human pathogen Trypanosoma cruzi. **Journal of Biological Chemistry**, v. 285, n. 33, p. 25582–25590, 2010.

LEPESHEVA, G. I. et al. VFV as a new effective CYP51 structure-derived drug candidate for chagas disease and visceral leishmaniasis. **Journal of Infectious**

Diseases, v. 212, n. 9, p. 1439–1448, 2015.

LEPESHEVA, G. I.; VILLALTA, F.; WATERMAN, M. R. Targeting Trypanosoma cruzi sterol 14α-demethylase (CYP51). **Advances in parasitology**, v. 75, p. 65–87, 2011.

LEPESHEVA, G. I.; WATERMAN, M. R. Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. **Biochimica et Biophysica Acta - General Subjects**, v. 1770, n. 3, p. 467–477, 2007.

LEPESHEVA, G. I.; WATERMAN, M. R. Sterol 14alpha-demethylase (CYP51) as a therapeutic target for human trypanosomiasis and leishmaniasis. **Current topics in medicinal chemistry**, v. 11, n. 16, p. 2060–2071, 2011.

LIPINSKI, C. A et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced drug delivery reviews**, v. 46, n. 1–3, p. 3–26, 2001.

LOMBARDINO, J. G.; LOWE, J. A. The role of the medicinal chemist in drug discovery-then and now. **Nature reviews. Drug discovery**, v. 3, n. 10, p. 853–862, 2004.

MACHADO, F. S.; TANOWITZ, H. B. Chagas Disease. In: JONG, E. C.; STEVENS, D. L. (Eds.). . **Netter's Infectious Diseases**. 1. ed. Philadelphia: [s.n.]. p. 522–526.

MANTOOTH, M.; ZEIBIG, E. The Hemoflagellates. In: GOCKEL-BLESSING, E. A. (Ed.). . **Clinical Parasitology: A Practical Approach**. 2. ed. Saint Louis: Elsevier: Saunders, 2013. p. 104–128.

MCCALL, L.-I. et al. Targeting Ergosterol Biosynthesis in Leishmania donovani: Essentiality of Sterol 14alpha-demethylase. **PLOS Neglected Tropical Diseases**, v. 9, n. 3, p. e0003588, 2015.

MCGANN, M. FRED and HYBRID docking performance on standardized datasets. **Journal of computer-aided molecular design**, v. 26, n. 8, p. 897–906, 2012.

MCGAUGHEY, G. B. et al. Comparison of topological, shape, and docking methods in virtual screening. **Journal of chemical information and modeling**, v. 47, n. 4, p. 1504–19, 2007.

MERRITT, C. et al. Kinases as Druggable Targets in Trypanosomatid Protozoan Parasites. **Chemical Reviews**, v. 114, n. 22, p. 11280–11304, 2014.

MICHALIK, M. S. M. Gênero Leishmania. In: NEVES, D. P. (Ed.). . **Parasitologia humana**. 11. ed. São Paulo: Editora Atheneu, 2005. p. 41–46.

MICHELS, P. A M. et al. Metabolic functions of glycosomes in trypanosomatids. **Biochimica et biophysica acta**, v. 1763, n. 12, p. 1463–77, 2006.

MIGNANI, S. et al. Compound high-quality criteria: A new vision to guide the development of drugs, current situation. **Drug Discovery Today**, v. 21, n. 4, p. 573–584, 2016.

MOLINA, I. et al. Randomized Trial of Posaconazole and Benznidazole for Chronic Chagas' Disease. **New England Journal of Medicine**, v. 370, n. 20, p. 1899–1908, 2014.

MORAES, C. B. et al. Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against Trypanosoma cruzi: implications for Chagas disease drug discovery and development. **Scientific reports**, v. 4, p. 4703, 2014.

MORGAN, H. P. et al. Allosteric mechanism of pyruvate kinase from Leishmania mexicana uses a rock and lock model. **The Journal of biological chemistry**, v. 285, n. 17, p. 12892–8, 2010.

MORGAN, H. P. et al. The trypanocidal drug suramin and other trypan blue mimetics are inhibitors of pyruvate kinases and bind to the adenosine site. **The Journal of biological chemistry**, v. 286, n. 36, p. 31232–31240, 2011a.

MORGAN, H. P. et al. The trypanocidal drug suramin and other trypan blue mimetics are inhibitors of pyruvate kinases and bind to the adenosine site.Supporting information. **The Journal of biological chemistry**, v. 286, n. 36, p. 31232–40, 9 set. 2011b.

MORGAN, H. P. et al. A new family of covalent inhibitors block nucleotide binding to the active site of pyruvate kinase. **The Biochemical journal**, v. 448, n. 1, p. 67–72, 2012.

MORGAN, S. et al. The cost of drug development: a systematic review. **Health policy (Amsterdam, Netherlands)**, v. 100, n. 1, p. 4–17, 2011c.

MS. **Leishmaniose Tegumentar Americana: Situação Epidemiológica-Dados**. Disponível em: <http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/723-secretaria-svs/vigilancia-de-a-a-z/leishmaniose-tegumentar-americana-lta/11328-situacao-epidemiologica-dados>. Acesso em: 18 jul. 2016a.

MS. **Leishmaniose Tegumentar Americana: descrição da doença**. Disponível em: <http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/723-secretaria-svs/vigilancia-de-a-a-z/leishmaniose-tegumentar-americana-lta/11324-descricao-da-doenca>. Acesso em: 19 jul. 2016b.

MS. **Leishmaniose Visceral: Situação Epidemiológica-Dados**. Disponível em: <http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/726-secretaria-svs/vigilancia-de-a-a-z/leishmaniose-visceral-lv/11334-situacao-epidemiologica-dados>. Acesso em: 18 jul. 2016c.

MS. **Ministério da Saúde. Doença de Chagas: Situação Epidemiológica**. Disponível em: <http://portalms.saude.gov.br/saude-de-a-z/doenca-de-chagas/situacao-epidemiologica#>. Acesso em: 10 jan. 2018a.

MS. **Ministério da Saúde. Doença de Chagas Aguda - Casos Confirmados Notificados no Sistema de Informação de Agravos de Notificação - Brasil: Casos confirmados segundo Região/UF de notificação - Período: 2007-2016.** Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/chagasbr.def>. Acesso em: 20 abr. 2018b.

MURRAY, H. W. et al. Advances in leishmaniasis. **Lancet**, v. 366, n. 9496, p. 1561–77, 2005.

MUSGROVE, P.; HOTEZ, P. J. Turning neglected tropical diseases into forgotten maladies. **Health affairs (Project Hope)**, v. 28, n. 6, p. 1691–706, 2009.

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.

NEVES, B. J. **Reposicionamento de fármacos e planejamento de novos compotos ativos contra Schistosoma mansoni.**, 2015.

NICOLAOU, K. C. Advancing the drug discovery and development process. **Angewandte Chemie (International ed. in English)**, v. 53, n. 35, p. 9128–9140, 2014.

NOWICKI, M. W. et al. Design, synthesis and trypanocidal activity of lead compounds based on inhibitors of parasite glycolysis. **Bioorganic & Medicinal Chemistry**, v. 16, n. 9, p. 5050–5061, 2008.

NUSSBAUM, K. et al. Trypanosomatid parasites causing neglected diseases. **Current medicinal chemistry**, v. 17, n. 15, p. 1594–617, 2010.

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

NWAKA, S.; RIDLEY, R. G. Virtual drug discovery and development for neglected diseases through public-private partnerships. **Nature reviews. Drug discovery**, v. 2, n. 11, p. 919–928, 2003.

OECD. **OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationships models**, 2004.

OECD. **Guidance document on the validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] models**, 2007.

OKWOR, I.; UZONNA, J. Social and economic burden of human leishmaniasis. **American Journal of Tropical Medicine and Hygiene**, v. 94, n. 3, p. 489–493, 2016.

OPPERDOES, F. R.; COOMBS, G. H. Metabolism of Leishmania: proven and predicted. **Trends in parasitology**, v. 23, n. 4, p. 149–58, 2007.

ORTEGA-BARRÍA, E. Trypanosoma Species (Trypanosomiasis). In: LONG, S. S.; PICKERING, L. K.; PROBER, C. G. (Eds.). . **Principles and Practice of Pediatric Infectious Diseases**. 3. ed. London: Churchill Livingstone: Elsevier, 2008. p. 1290–1296.

PARISE, M. E. et al. What Do We Know About Chagas Disease in the United States? **The American Journal of Tropical Medicine and Hygiene**, v. 95, n. 6, p. 1225–1227, 2016.

PAUCAR, R.; MORENO-VIGURI, E.; PÉREZ-SILANES, S. Challenges in Chagas

Disease Drug Discovery: A Review. **Current medicinal chemistry**, v. 23, n. 28, p. 3154–3170, 2016.

PAUL, S. M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. **Nature reviews. Drug discovery**, v. 9, n. 3, p. 203–14, 2010.

PAVLI, A.; MALTEZOU, H. C. Leishmaniasis, an emerging infection in travelers. **International journal of infectious diseases**, v. 14, n. 12, p. e1032–e1039, 2010.

PDB. **2014 RCSB Protein Data Bank Annual Report**, 2014.

PEACOCK, C. S. et al. Comparative genomic analysis of three Leishmania species that cause diverse human disease. **Nature genetics**, v. 39, n. 7, p. 839–47, 2007.

PEDRIQUE, B. et al. The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. **The Lancet Global Health**, v. 1, n. 6, p. e371–e379, 2013.

POLISHCHUK, P. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. **Journal of Chemical Information and Modeling**, v. 57, n. 11, p. 2618–2639, 2017.

POLLASTRI, M. P. Finding new collaboration models for enabling neglected tropical disease drug discovery. **PLoS neglected tropical diseases**, v. 8, n. 7, p. e2866, 2014.

RAJASEKARAN, R.; CHEN, Y.-P. P. Potential therapeutic targets and the role of technology in developing novel antileishmanial drugs. **Drug discovery today**, v. 20, n. 8, p. 958–968, 2015.

RENSLO, A. R.; MCKERROW, J. H. Drug discovery and development for neglected parasitic diseases. **Nature chemical biology**, v. 2, n. 12, p. 701–10, 2006.

RIGDEN, D. J. et al. The structure of pyruvate kinase from Leishmania mexicana reveals details of the allosteric transition and unusual effector specificity. **Journal of molecular biology**, v. 291, n. 3, p. 615–35, 1999.

RINIKER, S.; LANDRUM, G. A. Open-source platform to benchmark fingerprints for ligand-based virtual screening. **Journal of cheminformatics**, v. 5, n. 1, p. 26, 2013a.

RINIKER, S.; LANDRUM, G. A. Similarity maps - A visualization strategy for molecular fingerprints and machine-learning methods. **Journal of Cheminformatics**, v.

5, n. 9, p. 1–7, 2013b.

ROBERTS, L. S.; JAVONY, J. J. Kinetoplasta: Trypanosomes and Their Kin. In: ROBERTS, L. S.; JAVONY, J. J. (Eds.). . **Foundations of Parasitology**. 8. ed. New York: McGraw-Hill, 2009. p. 61–88.

ROGERS, D.; HAHN, M. Extended-connectivity fingerprints. **Journal of chemical information and modeling**, v. 50, n. 5, p. 742–754, 2010.

RØTTINGEN, J.-A. et al. Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory? **Lancet (London, England)**, v. 382, n. 9900, p. 1286–1307, 2013.

SANDERS, J. M. et al. Informing the Selection of Screening Hit Series with in Silico Absorption, Distribution, Metabolism, Excretion, and Toxicity Profiles. **Journal of Medicinal Chemistry**, v. 60, n. 16, p. 6771–6780, 2017.

SHELLEY, J. C. et al. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. **Journal of computer-aided molecular design**, v. 21, n. 12, p. 681–691, 2007.

SHOICHET, B. K. Virtual screening of chemical libraries. **Nature**, v. 432, n. December, p. 862–865, 2004.

SINGH, V. P. et al. Short report: Estimation of under-reporting of visceral leishmaniasis cases in Bihar, India. **American Journal of Tropical Medicine and Hygiene**, v. 82, n. 1, p. 9–11, 2010.

SOEIRO, M. D. N. C. et al. In Vitro and In Vivo Studies of the Antiparasitic Activity of Sterol 14 -Demethylase (CYP51) Inhibitor VNI against Drug-Resistant Strains of Trypanosoma cruzi. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 9, p. 4151–4163, 2013.

SONEGO, P.; KOCSOR, A.; PONGOR, S. ROC analysis: Applications to the classification of biological sequences and 3D structures. **Briefings in Bioinformatics**, v. 9, n. 3, p. 198–209, 2008.

SURYADEVARA, P. K. et al. Structurally Simple Inhibitors of Lanosterol 14α-Demethylase Are Efficacious In a Rodent Model of Acute Chagas Disease. **Journal of**

Medicinal Chemistry, v. 52, n. 12, p. 3703–3715, 2009.

SURYADEVARA, P. K. et al. Dialkylimidazole inhibitors of Trypanosoma cruzi sterol 14α-demethylase as anti-Chagas disease agents. **Bioorganic & Medicinal Chemistry Letters**, v. 23, n. 23, p. 6492–6499, 2013.

SVETNIK, V. et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. **Journal of chemical information and computer sciences**, v. 43, n. 6, p. 1947–58, 2003.

TADA, H. et al. An improved colorimetric assay for interleukin 2. **Journal of immunological methods**, v. 93, n. 2, p. 157–165, nov. 1986.

TALELE, T. T.; KHEDKAR, S. A; RIGBY, A. C. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. **Current topics in medicinal chemistry**, v. 10, n. 1, p. 127–141, 2010.

TEIXEIRA, D. E. et al. Interactive Multimedia to Teach the Life Cycle of Trypanosoma cruzi, the Causative Agent of Chagas Disease. **PLoS Neglected Tropical Diseases**, v. 6, n. 8, p. e1749, 28 ago. 2012.

TODESCHINI, R.; CONSONNI, V. **Handbook of molecular descriptors.** 11. ed. Weinheim: WILEY-VCH Verlag GmbH, 2000.

TODESCHINI, R.; CONSONNI, V. **Molecular descriptors for chemoinformatics.** 2. ed. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, 2010.

TRUCHON, J. F.; BAYLY, C. I. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. **Journal of Chemical Information and Modeling**, v. 47, p. 488–508, 2007.

TULLOCH, L. B. et al. Sulphate removal induces a major conformational change in Leishmania mexicana pyruvate kinase in the crystalline state. **Journal of molecular biology**, v. 383, n. 3, p. 615–26, 2008.

VANNIER-SANTOS, M. A; MARTINY, A.; DE SOUZA, W. Cell biology of

Leishmania spp.: invading and evading. **Current pharmaceutical design**, v. 8, n. 4, p. 297–318, 2002.

VEBER, D. F. et al. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. **Journal of Medicinal Chemistry**, v. 45, n. 12, p. 2615–2623, 2002.

VERLINDE, C. L. et al. Glycolysis as a target for the design of new anti-trypanosome drugs. **Drug resistance updates**, v. 4, n. 1, p. 50–65, 2001.

VERMA, J.; KHEDKAR, V. M.; COUTINHO, E. C. 3D-QSAR in drug design - a review. **Current topics in medicinal chemistry**, v. 10, n. 1, p. 95–115, 2010.

VIEIRA, D. F. et al. Binding Mode and Potency of N -Indolyloxopyridinyl-4-aminopropanyl-Based Inhibitors Targeting Trypanosoma cruzi CYP51. **Journal of Medicinal Chemistry**, v. 57, n. 23, p. 10162–10175, 2014.

VITAKU, E.; SMITH, D. T.; NJARDARSON, J. T. Analysis of the Structural Diversity, Substitution Patterns, and Frequency of Nitrogen Heterocycles among U.S. FDA Approved Pharmaceuticals. **Journal of Medicinal Chemistry**, v. 57, n. 24, p. 10257–10274, 2014.

WANG, Y. et al. PubChem BioAssay: 2014 update. **Nucleic acids research**, v. 42, n. Database issue, p. D1075-82, 2014.

WANG, Y. et al. PubChem BioAssay: 2017 update. **Nucleic acids research**, v. 45, n. D1, p. D955–D963, 2017.

WEI, B. Q. et al. A model binding site for testing scoring functions in molecular docking: This work is dedicated to the memory of Andy Morton (1964-1997). **Journal of Molecular Biology**, v. 322, n. 2, p. 339–355, 2002.

WELLING, M. **A first encounter with machine learning.** 1. ed. Irvine, CA: University of California, 2011.

WERMUTH, C. et al. Glossary of terms used in medicinal chemistry (IUPAC recommendations 1998). **Pure and applied chemistry**, v. 71, n. 10, p. 1919–1981, 1998.

WHO. **Control of the leishmaniases.World Health Organization technical report**

series, 2010.

WHO. **Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected diseases**, 2015.

WHO. **Clinical forms of leishmaniasis**. Disponível em: <http://www.who.int/leishmaniasis/clinical_forms_leishmaniases/en/index2.html>. Acesso em: 20 jul. 2016.

WHO. **Integrating neglected tropical diseases into global health and development: fourth WHO report on neglected tropical diseases**, 2017.

WHO. **Chagas disease (American trypanosomiasis) Fact Sheet**. Disponível em: <http://www.who.int/mediacentre/factsheets/fs340/en/>. Acesso em: 12 jan. 2018.

WILLIGHAGEN, E. L. et al. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. **Journal of Cheminformatics**, v. 9, n. 1, p. 1–19, 2017.

WILLSON, M. et al. Synthesis and activity of inhibitors highly specific for the glycolytic enzymes from Trypanosoma brucei. **Molecular and biochemical parasitology**, v. 59, n. 2, p. 201–10, 1993.

YARDLEY, V.; CROFT, S. L. A comparison of the activities of three amphotericin B lipid formulations against experimental visceral and cutaneous leishmaniasis. **International journal of antimicrobial agents**, v. 13, n. 4, p. 243–8, fev. 2000.

ZHENG, W.; THORNE, N.; MCKEW, J. C. Phenotypic screens as a renewed approach for drug discovery. **Drug Discovery Today**, v. 18, n. 21–22, p. 1067–1073, 2013.

ZHOU, J. et al. Progress in the Rational Design for Polypharmacology Drug. **Current pharmaceutical design**, v. 22, n. 21, p. 3182–3189, 2016.

ANEXOS

Anexo 1 - Artigos publicados através de colaborações durante o doutorado

- QSAR-Driven Discovery of Novel Chemical Scaffolds Active against *Schistosoma mansoni*. **Journal of Chemical Information and Modeling**, (56) 7, p.1357-1372, 2016. (Primeiro Autor).

  DOI: 10.1021/acs.jcim.6b00055

- Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening. **Journal of Medicinal Chemistry,** (59) 15, p.7075-7088, 2016.

  DOI: 10.1021/acs.jmedchem.5b02038

- Computer-aided discovery of two novel chalcone-like compounds active and selective against *Leishmania infantum*. **Bioorganic & Medicinal Chemistry Letters**, (27) 11, p. 2459-2464, 2017.

  DOI: 10.1016/j.bmcl.2017.04.010

- QSAR-Driven Design and Discovery of Novel Compounds with Antiplasmodial and Transmission Blocking Activities. **Frontiers in Pharmacology**, (9), p. 1-13, 2018.

  DOI: 10.3389/fphar.2018.00146

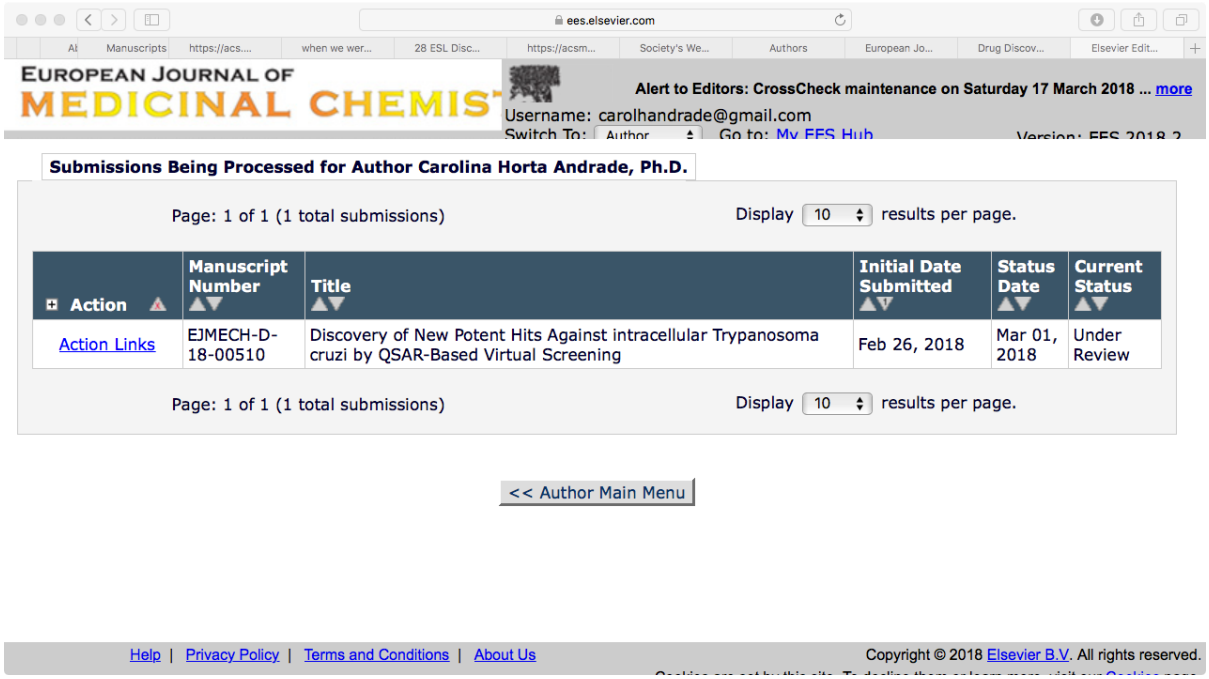
- In Silico Chemogenomics Drug Repositioning Strategies for Neglected Tropical Diseases. **Current Medicinal Chemistry,** (25), 2018.

  DOI: 10.2174/0929867325666180309114824

Anexo 2 - Artigo submetido e comprovante de submissão

Artigo:

Discovery of New Potent Hits Against intracellular *Trypanosoma cruzi* by QSAR-Based Virtual Screening

Autores: Cleber C. Melo-Filho, Rodolpho C. Braga, Eugene N. Muratov, Caio Haddad Franco, Carolina B. Moraes, Lucio H. Freitas-Junior & Carolina Horta Andrade

(Artigo submetido à revista *European Journal of Medicinal Chemistry – under review*)