UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUTOS NATURAIS E SINTÉTICOS BIOATIVOS

CHONNY ALEXANDER HERRERA ACEVEDO

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE SESQUITERPENOS LACTONIZADOS ISOLADOS DA FAMÍLIA ASTERACEAE COM POTENCIAL ATIVIDADE LEISHMANICIDA E TRIPANOCIDA

João Pessoa
2017

CHONNY ALEXANDER HERRERA ACEVEDO

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE SESQUITERPENOS LACTONIZADOS ISOLADOS DA FAMÍLIA ASTERACEAE COM POTENCIAL ATIVIDADE LEISHMANICIDA E TRIPANOCIDA

Dissertação de Mestrado do Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos da UFPB, como requisito para obtenção do título de Mestre em Produtos Naturais e Sintéticos Bioativos, na área de concentração: Farmacoquímica.
Área de concentração: FARMACOQUÍMICA

Orientador: Prof. Dr. Marcus Tullius Scotti

Co-orientador: Prof. Dr. Mateus Feitosa Alves

João Pessoa

2017

Catalogação na publicação
Seção de Catalogação e Classificação

A174e Acevedo, Chonny Alexander Herrera.
Estudos quimiotaxonômicos e triagem virtual de sesquiterpenos lactonizados isolados da família asteraceae com potencial atividade leishmanicida e tripanocida / Chonny Alexander Herrera Acevedo. - João Pessoa, 2017.
270 f. : il.

Orientador: Dr. Marcus Tullius Scotti.
Co-orientador: Dr. Mateus Feitosa Alves.
Dissertação (Mestrado) - UFPB/CCS.

1. Farmacoquímica. 2. Descritores moleculares. 3. Docking molecular. I. Scotti, Marcus Tullius. II. Alves, Mateus Feitosa. III. Título.

UFPB/BC

CHONNY ALEXANDER HERRERA ACEVEDO

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE SESQUITERPENOS LACTONIZADOS ISOLADOS DA FAMÍLIA ASTERACEAE COM POTENCIAL ATIVIDADE LEISHMANICIDA E TRIPANOCIDA

Aprovada em: ___ / ___ / _____

COMISSÃO EXAMINADORA

Prof. Dr. Marcus Tullius Scotti
PhD em Química Orgânica
Universidade Federal da Paraíba
Orientador

Prof. Dr. Mateus Feitosa Alves
PhD em Produtos Naturais e sintéticos Bioativos
Universidade Federal da Paraíba
Co-orientador

Prof. Dr. Francisco Jaime Bezerra Mendonça Júnior
PhD em Ciências Biológicas
Universidade Federal da Paraíba
Avaliador Interno

Prof. Dr. Gerd Bruno da Rocha
PhD em Química
Universidade Federal da Paraíba
Avaliador externo

Prof. Dr. Josean Fechine Tavares
PhD em Produtos Naturais e sintéticos Bioativos
Universidade Federal da Paraíba
Avaliador Interno
Orientador

Prof. Dr. Karen Cacilda Weber
PhD em Química (Físico-Química).
Universidade Federal da Paraíba
Avaliadora externa

A mis padres y Hermanos, por todo su amor y grandes sacrificios
Ustedes son mi mayor orgullo, mi felicidad.

Aos meus pais e irmãos, por todo o seu amor e os grandes sacrifícios
Vocês são o meu maior orgulho, minha felicidade.

AGRADECIMENTOS

Ao CNPq e ao Programa Estudante Convenio pós-graduação (PEC-PG).

À Universidade Federal da Paraíba e o seu Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos.

Ao coordenador do PgPNSB, Prof. Dr. Josean Fechine Tavares, pela ajuda constante durante estes dois anos.

Aos professores do PgPNSB, por seus importantes ensinamentos para minha vida professional e pessoal.

À minha turma de mestrado e os alunos do programa por compartilhar comigo nas disciplinas.

Ao Laboratório de Quimioinformática: Élida, Gabriela, Jéssika, Luana, Mayara Maia, Mayara Barbalho, Vanessa, Alixandre, Renata e Renan.

Ao meu orientador Prof. Dr. Marcus Tullius Scotti pela oportunidade de trabalhar ao seu lado, permitindo-me viver uma maravilhosa experiência académica.

Ao meu Co-orientador Prof. Dr. Mateus Feitosa Alves por acompanhar sempre este processo e seus ensinamentos.

À Profa. Dra. Luciana Scotti por seu grande exemplo como pesquisadora e como pessoa.

Ao Prof. Dr. Ericsson Coy Barrera na UMNG, pela sua confiança e amizade.

Aos membros da banca: Prof. Dr. Francisco, Prof. Dr. Gerd, Prof. Dr. Josean, Profa. Dra. Karen, por ler o presente trabalho, assim como pelos seus importantes comentários e sugestões.

Ao povo Brasileiro por permitir-me realizar meus estudos de mestrado neste País, desejo o melhor para todos vocês.

AGRADECIMENTOS

A Dios y la virgen, por la fuerza para enfrentar cada día y las múltiples dificultades.

A mis papitos: Martha y Jorge, nada ha sido fácil, pero sin duda, la satisfacción de todo lo que hemos conseguido gracias al trabajo y la disciplina, es mi mayor felicidad.

A Nury, Nery y Kenyi, mis hermanos del alma. Estoy muy orgulloso de todo lo que son y lo que me han permitido ser a su lado.

A mi Familia, mis tíos: Gladys, Hernando, Hugo, Humberto, Nubia, mis primos Juancho, Sofia, Nani, Valeria y Daniela. A Edgar Alfonso Herrera, por su ayuda incondicional cuando lo necesité.

Al Departamento de Química de la Universidad Nacional de Colombia y sus profesores, por tantas enseñanzas y experiencias que me han brindado, siempre será un orgullo ser egresado de la Universidad pública colombiana.

A todos mis amigos en Colombia, por la compañía constante en tantos momentos difíciles, por escucharme y ayudarme cuando lo necesité

A la señora Consuelo Bernal por acompañarme siempre en este sueño llamado Brasil.

À família Scotti por me receber desde o primeiro dia que cheguei em João Pessoa e converter-se na minha família no Brasil, toda minha admiração e respeito por vocês.

Ao Professor Marcus, pela grande amizade y compartilhar além da parte académica, uma das razoes principal da vida, o Futebol. Obrigado por deixar-me ver que o Napoli é um sinónimo do meu jeito de ver este lindo esporte.

Ao Professor Mateus, um grande amigo. Obrigado pela ajuda e colaboração nestes dois anos.

À minhas colegas e amigas do grupo de pesquisa, agradeço muito tudo o que vocês fizeram por mim, sempre vou me lembrar de vocês.

A Bianka Xavier por ser uma pessoa tão maravilhosa comigo, e sempre ajudar-me em tudo o que eu precisei. Meus melhores desejos para você e Tássio. Obrigado.

A K, ¡hoy por fin!

ACEVEDO, Chonny Alexander Herrera. **Estudos Quimiotaxonômicos e Triagem Virtual de Sesquiterpenos Lactonizados Isolados da Família Asteraceae com Potencial Atividade Leishmanicida e Tripanocida, 2017**
Dissertação (Mestrado em Produtos Naturais e Sintéticos Bioativos – Farmacoquímica)
Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, 2017.

RESUMO

A leishmaniose e a tripanossomíase americana são as doenças negligenciadas que afetam em maior percentagem à população do Brasil e da Colômbia. Algumas das principais dificuldades para o controle e eliminação destas doenças são a limitação e ineficácia dos atuais tratamentos usados, principalmente nas fases crónicas. Os produtos naturais são potenciais fontes de novos e seletivos agentes para o tratamento de doenças tropicais causados por protozoários e outros parasitos. A família Asteraceae tem permitido o isolamento de uma grande variedade de substâncias que exibem diversas atividades biológicas entre estes, sesquiterpenos lactonizados que ainda não têm sido estudados com profundidade, mas apresentam um grande potencial de atividade antiprotozoária. Os métodos computacionais, têm emergido, como uma importante ferramenta na área da química medicinal, melhorando a eficiência de múltiplos processos para o desenvolvimento de novos tratamentos, reduzindo custos e recursos. Portanto, este estudo tem como objetivo, utilizar, criar e combinar diversas metodologias de triagem virtual de banco de dados de produtos naturais através de docking molecular e aprendizado de máquina utilizando descritores moleculares para propor sesquiterpenos lactonizados com potencial atividade leishmanicida e tripanocida multialvo. No capítulo 2, apresenta-se uma revisão da importância biológica dos sesquiterpenos lactonizados na luta contra as doenças negligenciadas incluindo algumas investigações desenvolvidas com ferramentas in silico na busca de novas moléculas com atividade leishmanicida e antichagásica. O capitulo 3, apresenta uma moderna e inovadora ferramenta web para o gerenciamento do banco de dados de metabolitos secundários, chamada SISTEMATX a qual foi desenvolvida no laboratório de quimioinformática da Universidade federal da Paraíba. Os últimos dois capítulos, mostram estudos computacionais dirigidos a busca de estruturas com potencial atividade antichagásica e leishmanicida multialvo, usando um banco de dados próprio (SISTEMATX) constituído por 1,306 SLs. Os modelos classificatórios *Random forest* construídos para cada uma das formas parasitarias de *T. cruzi e L. donovani* alcançaram um poder preditivo acima de 75,4% de acurácia. Mediante a triagem virtual baseada na estrutura do ligante, o número de estruturas classificadas como ativas foram: 34, 17 e 420 para amastigotas, tripomastigotas e epimastigotas de *T. cruzi*, assim como, 741 e 11 para amastigotas e promastigotas de *L. donovani*. Estudos de triagem virtual baseado na estrutura do receptor foram realizados mediante docking molecular do mesmo banco de moléculas usando 8 enzimas alvos do *T. cruzi* e 4 proteínas de *L. donovani*, incluindo um modelo de homologia de Redutase de pteridina 1 (*Ld*PTR1). Finalmente, mediante uma análise de consenso das duas técnicas, buscou-se normalizar os valores de probabilidades, para verificar compostos potencialmente ativos contra estas duas doenças negligenciadas e estabelecer o possível mecanismo de ação destes.

Palavras-chave: Asteraceae, descritores moleculares, sesquiterpenos lactonizados, antiprotozoários, triagem virtual, Random forest, *docking* molecular.

ACEVEDO, Chonny Alexander Herrera. **Chemotaxonomic studies and virtual screening of sesquiterpene lactones isolates of the Asteraceae family with leishmanicidal and trypanocidal potential activity., 2017**
Master's final dissertation (Postgraduate Program in Natural and Synthetic Bioactive Products-Pharmacochemistry)
Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, 2017.

ABSTRACT

Leishmaniasis and American trypanosomiasis are the neglected diseases that affect a large percentage of the Brazil and Colombia population. Some of the difficulties for the control and elimination of these diseases are the limitation and inefficacy of the current treatments mainly in the chronic phases. Natural products are potential sources of new and selective agents for the treatment of tropical diseases caused by protozoa and other parasites. Asteraceae family has allowed the isolation of a great variety of substances that exhibit diverses biological activities, among them, sesquiterpenes lactones, which have not been studied yet deepth, but present a great potential of antiprotozoal activity. Computational methods have emerged, as an important tool in the field of medicinal chemistry, improving the effectiveness of processes to develop of new treatments, reducing costs and resources. Therefore, this study aims use, create and combine several methodologies of virtual screening of a natural products database through molecular docking, and machine learning algorithms using molecular descriptors, in order to find sesquiterpenes lactones with potential trypanocidal and leishmanicidal activity. In Chapter 2, is presented a review of the biological importance of sesquiterpenes lactones in the fight against neglected diseases including some investigations performed with *in silico* tools in the search of new molecules with leishmanicidal and antichagasic activity. Chapter 3 presents a modern and innovative web tool for database managing of secondary metabolites, called SISTEMATX, which was developed in the cheminformatics laboratory of the Federal University of Paraíba. The last two chapters, present computational studies aimed at finding structures with potential antichagasic and leishmanicidal multitarget activity, using an in-house databank (SISTEMATX) consisting of 1,306 sesquiterpenes lactones. Random forest classificatory models built for each one of the parasitic forms of *T. cruzi* and *L. donovani* reached a predictive power above of 75,4% of accuracy. By ligand-based virtual screening, the number of structures classified as active were: 34, 17 and 420 for amastigotes, trypomastigotes and epimastigotes of *T. cruzi*, as well as, 741 and 11 for amastigotes and promastigotes of *L. donovani*. Likewise, structure-based virtual screening was performed by molecular docking of the same set of molecules using eight target enzymes of *T. cruzi* and four proteins of *L. donovani*, including a homology model of Pteridine reductase 1 (*Ld*PTR1). Finally, by means of a consensus analysis of the two techniques, it was sought to normalize the probability values, to verify potentially active compounds against these two neglected diseases and to establish their possible mechanism of action.

Keywords: Asteraceae, molecular descriptors, sesquiterpene lactones, antiprotozoal, virtual Screening, random forest, molecular docking.

LISTA DE ABREVIATURAS, SIGLAS E FÓRMULAS

2D - Two-dimensional (Bidimensional)

3D - Three-dimensional (Tridimensional)

APD - Applicability domain (Dominio de aplicabilidade)

API - Application Programming Interface (Interface de programação de aplicativos)

AUC - Area Under Curve (Área sob curva)

CADD - Computer-Aided Drug Design (Desenho de drogas assistido por computador)

CSS - Cascading Style Sheets

CV - Cross-validation (Validação cruzada)

DAO - Data Access Object (Objeto de acesso a dados)

DNP - Dictionary of Natural Products (Dicionário de produtos naturais)

DOM - Document Object Model (Modelo de Documento Objeto)

EC_{50} - Effective concentration 50% (Concentração Efetiva 50%)

FN - False Negative (Falso negativo)

FP - False Positive (Falso positivo)

GPS - Global Positioning System (Sistema de posicionamento global)

H-bond - Hydrogen bonding (Ligação de hidrogênio)

HIV – Human Immunodeficiency Virus (Vírus da Imunodeficiência Humana)

HMBC -Heteronuclear Multiple Bond Correlation (Correlação de ligação múltipla heteronuclear)

HTML - HyperText Markup Language (Linguagem de Marcação de Hipertexto)

HTS - High-throughput screening (Triagem de Alto Desempenho)

IC_{50} - Inhibition Concentration 50% (Concentração Inibitória 50%)

ID – Identifier (Identificador)

JSP - JavaServer Pages

LD_{50} - Letal Dose 50% (Dose letal 50%)

MCC - Matthews's correlation coefficient (Coeficiente de Correlação de Matthews)

MDA – Malondialdehyde (Malondialdeído)

MDL - Molecular Design Limited

MIF – Molecular Interaction Field (Campo de Interação Molecular)

MPK3 - Mitogen-activated protein kinase 3 (Proteíno-quinase ativada por mitógeno 3)

NADPH - Nicotinamide adenine dinucleotide phosphate (Fosfato de dinucleótido de nicotinamida e adenina)

NMT - N-Myristoyltransferase (N-miristoiltransferase)

NOX - Oxidation number (Número de oxidação)

NTD - Neglected Tropical Disease (Doença tropical negligenciada)

ODC - Ornithine Decarboxylase (Ornitina decarboxylase)

PAINS - Pan Assay Interference Compounds

PDB - Protein Data Bank (Banco de Dados de Proteínas)

PLS - Partial Least Squares (Mínimos Quadrados Parciais)

PTR1 - Pteridine Reductase 1

QSAR - Quantitative-Structure Activity Relationship (Relação Quantitativa Estrutura-Atividade)

RF - Random Forest (Florestas Aleatórias)

RMN – Ressonância Magnética Nuclear

RMSD - Root Mean Square Deviation (Raiz do Desvio Quadrático Médio)

ROC - Receiver Operating Characteristic (Característica de Operação do Receptor)

ROS - Reactive Oxygen Species (Espécies Reactivas de Oxigénio)

SAR – Structure-Activity Relationship (Relação Estrutura-Atividade)

SEM - Scanning Electron Microscopy (Microscópio Eletrônico de Varredura)

SL - Sesquiterpene Lactone (Sesquiterpeno lactonizado)

SMILES - Simplified Molecular Input Line Entry System

SVM – Support Vector Machine (Máquina de Vetores de Suporte)

TGR - Thioredoxin Glutathione Reductase

TN - True Negative (Verdadeiro Negativo)

TP - True Positive (Verdadeiro Positivo)

VL - Visceral Leishmaniasis (Leishmaniose Visceral)

VS - Virtual Screening (Triagem Virtual)

WHIM - Weighted Holistic Invariant Molecular

LISTA DE FIGURAS

CAPÍTULO I

CAPÍTULO II

CAPÍTULO III

CAPÍTULO IV

CAPÍTULO V

LISTA DE TABELAS

CAPÍTULO III

CAPÍTULO IV

CAPÍTULO V

SUMÁRIO

CAPÍTULO I

1. INTRODUÇÃO

Historicamente, os produtos naturais têm contribuído amplamente para o desenvolvimento da medicina moderna, e ainda continuam desempenhando um importante papel no descobrimento de fármacos (SEN, 2017). Mais do 50% das novas pequenas moléculas aprovadas pela FDA (do inglês, *Food and Drug Administration*) entre 1981 a 2014, são diretamente ou indiretamente relacionadas aos produtos naturais (compostos de origem sintética inspirados em estruturas de produtos naturais). (NEWMAN, 2016). O panorama é ainda mais promissório tendo em conta que mais do 95% da biodiversidade mundial, ainda não foi avaliada para nenhum tipo de atividade biológica (BERNARDINI, 2017).

Uma das famílias de plantas mais estudadas neste campo é Asteraceae também denominada Compositae, a qual é considerada como a maior família de plantas floridas, formada por mais de 23.000 espécies que estão presentes em todos os continentes, principalmente no novo mundo (BOHM & STUESSY, 2001). Diversos tipos de metabolitos secundários têm sido isolados de espécies de Asteraceae sendo os mais comuns: monoterpenos, sesquiterpenos, sesquiterpenos lactonizados, poliacetilenos, flavonóides, benzofuranos e benzopiranos, cumarinas, diterpenóides e triterpenóides (EMERENCIANO, 2006)

Os sesquiterpenos lactonizados são metabolitos secundários formados pela condensação de três unidades de isopreno, seguido pela ciclização e transformação oxidativa para formar uma lactona fusionada cis- ou trans-, sendo este anel γ-lactona, a principal característica deste grupo de compostos, usualmente com a presença de um grupo α-metileno (IVANESCU, 2015). Na família Asteraceae mais de 5.000 compostos têm sido relatados, sendo usados como marcadores quimiotaxonômicos desta família, apresentando uma ampla gama de atividades protetoras nas plantas, inclusive atuando como substâncias antimicrobianas e antimicrobianas ou inibindo o crescimento de plantas competidoras (PADILLA-GONZÁLEZ, 2016).

Estruturalmente é possível estabelecer quatro classes representativas de sesquiterpenos lactonizados (Figura 1):

- Germacranolideos que possuem um anel de 10 membros.
- Eudesmanolideos os quais têm dois anéis de seis átomos de carbono fusionados.
- Guaianolideos e pseudoguaianolídeos, se caracterizam por ter um anel de sete carbonos, um anel de cinco membros e um grupo metila.

Em todas as classes um anel lactona de cinco membros com uma carbonila na posição alfa se encontra fusionado ao esqueleto (YOSHIOKA, 1973; SEAMAN, 1982).

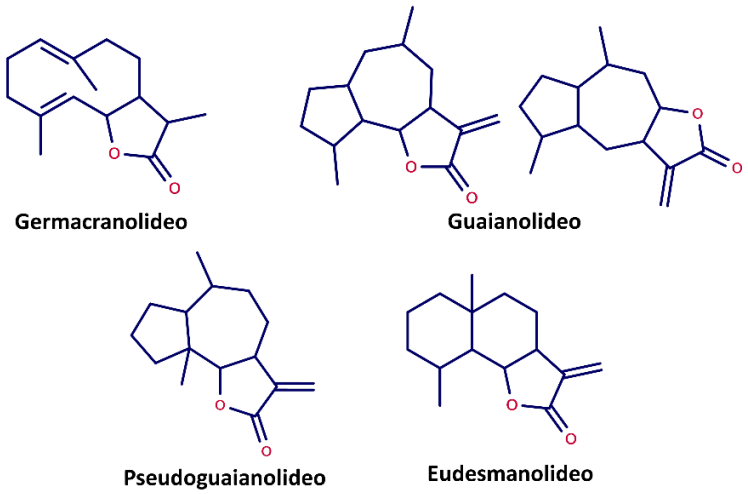


Figura 1: Principais esqueletos dos sesquiterpenos lactonizados.

Devido a sua grande diversidade estrutural, juntamente com as potenciais atividades biológicas contra várias doenças, os sesquiterpenos lactonizados têm aumentado o interesse entre químicos medicinais de todo o mundo (CHATURVEDI & DWIVEDI, 2017). Estruturas com atividade biológica definidas já têm sido utilizadas com sucesso em seres humanos, sendo a descoberta de artemisinina um dos exemplos mais destacados (Prêmio Nobel em medicina e fisiologia 2015) (TU, 2011). Este fármaco, tem a ação mais rápida contra a malária causada pelo *Plasmodium falciparum* sendo atualmente a base das terapias combinadas que compõem o tratamento de primeira linha contra a malária causada por *P. falciparum* (OMS, 2006). Casos como o anterior, permitiram abrir um novo panorama com relação à pesquisa de novas terapêuticas usando sesquiterpenos lactonizados que, além da atividade antimalárica, têm presentado potencial antitumoral, prevenção de neurodegeneração, atividade analgésica e sedativa.

As estratégicas clássicas para a descoberta e desenvolvimento de novas drogas requerem de muito tempo assim como de processos que consumem uma grande quantidade de recursos. A partir dessas dificuldades, as ferramentas computacionais emergem como uma importante alternativa para melhorar a efetividade e eficiência desses processos, favorecendo entre outros, a redução de custos, o uso de animais e o impacto no meio ambiente (KAPETANOVIC, 2008).

Um dos principais objetivos do desenvolvimento de drogas assistido por computador (CADD, do inglês *Computer-Aided Drug Design*) é a redução de um grande número de

compostos em menores conjuntos de compostos ativos que podem ser testados experimentalmente (IMAM, 2017), o que está diretamente relacionado com o aumento exponencial na quantidade de novas estruturas, sendo fundamental o uso de bases de dados que permitam a pesquisa rápida e classificação das moléculas assim como informações úteis relacionadas com estas.

Os métodos CADD podem ser classificados em dois tipos, os baseados na estrutura do ligantes que usam uma abordagem indireta para facilitar o desenvolvimento de compostos farmacologicamente ativos ao estudar moléculas que interagem com o alvo biológico de interesse, e os baseados na estrutura do receptor, que usam diretamente o conhecimento da estrutura tridimensional (3D) da molécula alvo para identificar ou otimizar os candidatos a medicamentos (ACHARIA, 2011).

Nesse contexto, o presente trabalho apresenta estudos computacionais baseados na estrutura do ligante como também baseado na estrutura do receptor, com o objetivo de identificar sesquiterpenos lactonizados com potencial atividade antichagásica e leishmanicida, a partir de um grupo de 1.306 moléculas que estão cadastradas no SistematX, uma base de dados de produtos naturais desenvolvida no laboratório de quimioinformática da Universidade Federal da Paraíba.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 DESENVOLVIMENTO DE DROGAS ASSISTIDO POR COMPUTADOR

A descoberta de novos medicamentos é um processo lento, custoso e dispendioso. O processo desde a ideia até chegar ao mercado consta de sete passos básicos: a seleção da doença, seleção do alvo, identificação do composto *lead (*composto ou serie de compostos que satisfazem os critérios mínimos predefinidos para otimização aprofundada da estrutura e atividade), otimização do *lead*, ensaios pré-clínicos, ensaios clínicos e otimização farmacogenômica, sendo necessária várias repetições dos últimos cinco passos. Assim, um medicamento disponível no mercado tem sido o resultado de um entre 100,000 compostos investigados com um custo médio para seu desenvolvimento acima dos 800 milhões de dólares (HOQUE, 2017).

Nos últimos anos, o uso de metodologias de desenvolvimento de drogas associado ao uso dos computadores tem aumentado exponencialmente, a partir da grande evolução da informática assim como as múltiplas vantagens que esta classe de técnicas fornece, pois, alguns dos passos mencionados anteriormente podem ser substituídos por técnicas *in silico*, gerando uma significativa redução no custo de pesquisa assim como do tempo de desenvolvimento da droga (HOQUE, 2017; DAS, 2017).

Os métodos de desenvolvimento de drogas assistidos por computador (CADD, do inglês *Computer-Aided Drug Design*) podem ser classificados em dois grandes grupos: CADD baseados na estrutura do ligante e CADD baseados na estrutura do receptor (Figura 2). Os estudos computacionais baseados na estrutura do ligante, usam a informação sobre compostos com valores de atividade conhecidos para a construção de modelos de relação quantitativa entre a estrutura química e a atividade biológica (QSAR, do inglês *Quantitative Structure-Activity Relationship)* preditivos ou realizando pesquisas de similaridade química, no entanto os estudos baseados na estrutura do receptor, utilizam informação estrutural do alvo (ABDOMALEKI, 2017).

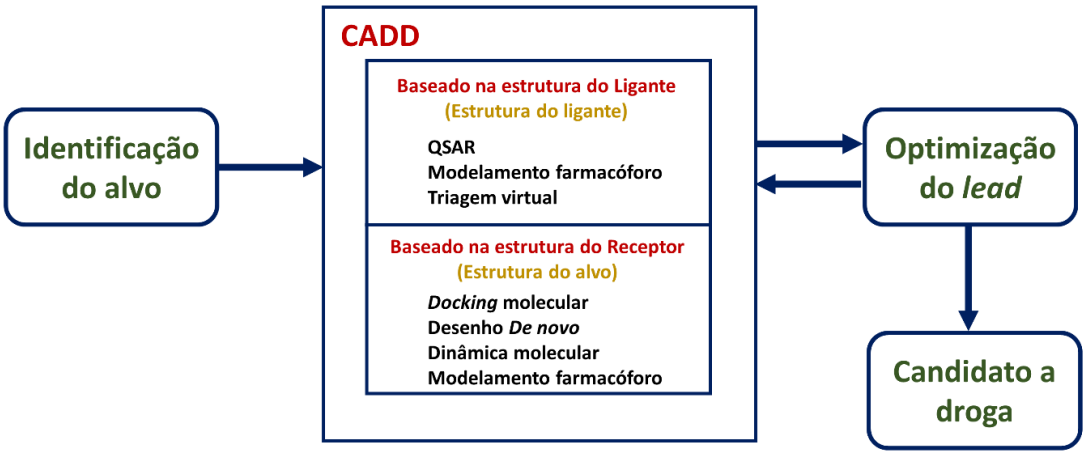


Figura 2: Esquema geral do desenvolvimento de drogas assistido por computador. Adaptado de ABDOMALEKI, 2017

2.1.1 CADD baseado na estrutura do ligante

A habilidade de determinar a estrutura dos compostos assim como sua atividade biológica têm permitido estabelecer uma relação entre estes (SAR, do inglês *Structure-Activity Relationship*), que se baseia em observar quais mudanças na estrutura do composto podem resultar em uma mudança em sua atividade biológica (THOMAS, 2011).

Os estudos de Hammett que relacionavam de maneira linear o logaritmo da constante de ionização do ácido benzoico meta ou para-substituído com o logaritmo da constante de ionização do ácido benzoico (HAMMETT, 1937), serviram de base para entender as diferenças das estruturas com relação a diversas propriedades/atividades biológicas. Na década dos 60s, começaram a ser publicados estudos que propuseram que "reações" biológicas poderiam ser tratadas como as reações químicas, por técnicas de físico-química orgânica, surgindo assim as metodologias para estudos de QSAR que correlacionam a dependência das atividades biológicas com outras propriedades físico-químicas ("*Hansch analysis*", abordagem extratermodinâmica), ou de acordo com a ausência ou presença de características estruturais (abordagem *Free-Wilson*) (GRAMATICA, 2008; KUBINYI, 2008).

Assim, os estudos QSAR partem da base que a estrutura de uma molécula ou uma conformação molecular deve possuir características responsáveis pelo seu comportamento físico, químico e biológico, definido com base em descritores numéricos. Igualmente, estes

estudos fundamentam-se no princípio que moléculas com similaridade estrutural devem apresentar características físicas, químicas e propriedades biológicas semelhantes (TODESCHINI, 1994).

2.1.2 Descritores Moleculares

As pesquisas em QSAR apresentam uma grande dificuldade devido à complexidade de muitos dos sistemas estudados. Por isso é fundamental realizar uma boa descrição do problema assim como a adequada seleção do método utilizado para extrair a informação, para garantir a qualidade dos resultados obtidos (TODESCHINI, 1994).

Os estudos QSAR começam com a representação das estruturas mediante o uso alguns descritores moleculares (RANDIĆ, 1993), o qual é definido como o resultado final de um procedimento lógico e matemático que transforma informação química codificada dentro de uma representação simbólica de uma molécula em um número útil ou o resultado de algum experimento padronizado. (TODESCHINI, 2008)

Os descritores moleculares estão classificados de acordo às várias dimensões utilizadas nas análises de QSAR, sendo os mais comuns, aqueles que são derivados da fórmula química (sem dimensão); os derivados usando a representação de vários sub-fragmentos moleculares estruturais (1D – uma dimensão); os obtidos a partir da representação teórica do gráfico de moléculas. Este grupo também inclui vários aspectos estruturais e / ou físico-químicos (2D – duas dimensões). Finalmente encontram-se essas variáveis independentes que codificam diferentes dimensões espaciais e geométricas formações de compostos e são determinadas a partir da representação tridimensional de estrutura (3D) (ROY, 2014).

Os descritores Volsurf (*VOLume and SURFace*) estão classificados neste último grupo, estes extraem a informação presente em mapas 3D de campos moleculares denominados MIFs (do inglês *Molecular Interactions Fields*) gerando uns poucos descritores numéricos quantitativos que são fáceis de interpretar (CRUCIANI, 2000a).

Os MIFs são campos de interação gerados pelo campo de força GRID, o qual usa um potencial baseado na energia total da interação entre a molécula a ser avaliada e uma sonda (CRUCIANI, 2000b). Desta maneira pode ser explorado o espaço das propriedades físico-

químicas das moléculas a partir das energias de interação entre estas e uma sonda virtual química (LORENZO, 2016).

Na figura 3, está representado um MIF, o qual é visto como um cubo (3D) com várias divisões. Cada vértice do cubo é a representação de um ponto a partir do qual as interações entre a molécula estudada e a sonda química respectiva são determinadas. Finalmente, uma das principais vantagens do uso de descritores Volsurf é que estes são dificilmente influenciados pela conformação da molécula alvo (CRUCIANI, 2000b).

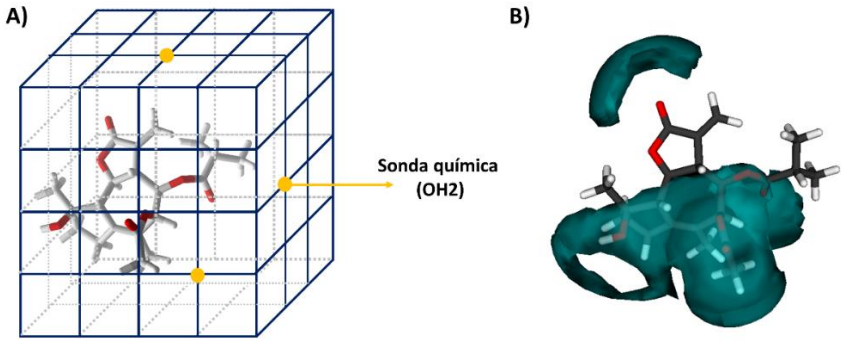


Figura 3: A) Campo de Interação Molecular (MIF) entre a sonda e uma molécula alvo. B) Derivação dos descritores Volsurf, campo favorável é representado em azul.

2.1.3 Triagem virtual baseada na estrutura do ligante

A técnica experimental mais comum na descoberta de drogas é a denominada *High-throughput screening* (HTS). Esta metodologia usa a robótica, o processamento de dados assim como múltiplos detectores cada vez mais sensíveis, para testar bibliotecas compostas por milhares (até milhões) de moléculas permitindo uma fácil identificação de compostos ativos (LIEBERT, 2008; CARAUS 2015).

No entanto, o custo das bibliotecas de compostos assim como dos equipamentos que permitam o sucesso da técnica dificulta em muitos casos a aplicação desta para a busca de novas moléculas ativas contra várias doenças. A partir disso, novas técnicas têm sido desenvolvidas nas últimas décadas, entre as mais destacadas aparece a triagem virtual baseada na estrutura do ligante (SUBRAMANIAM, 2008).

A triagem virtual baseada na estrutura do ligante e o HTS, têm essencialmente o mesmo objetivo, a identificação de novas moléculas ativas. No entanto, a filosofia das técnicas é

totalmente oposta, sendo que o HTS considera testar experimentalmente a maior quantidade de compostos da maneira mais eficiente, e a triagem virtual tenta racionalizar a seleção composta para reduzir o número de candidatos para avaliação experimental tanto quanto possível (RIPPHAUSEN, 2011).

Entre as principais vantagens da triagem virtual em relação ao HTS, encontra-se a possibilidade de pesquisar partes extremamente amplas do espaço químico, devido ao fato de que os compostos nesta metodologia não precisam existir fisicamente o que permite ter uma grande quantidade de moléculas e derivados, já que isto permite a proposição de novas estruturas. Sob este conceito, múltiplas bases de dados foram desenvolvidas por diferentes grupos de pesquisa (*in-house*), permitindo a identificação de moléculas promissórias, *hits* para diferentes doenças (KOEPEN, 2011). Um *hit* é definido como uma molécula que produz atividade reprodutível acima de um limite definido em um ensaio biológico e cuja identidade estrutural foi estabelecida (BUCKLE, 2013)

A partir da importância e grandes vantagens que apresentam este tipo de técnicas, diversas aproximações de triagem virtual baseada na estrutura do ligante têm sido desenvolvidas. Na figura 4, são observadas as principais metodologias de usadas nesta área.

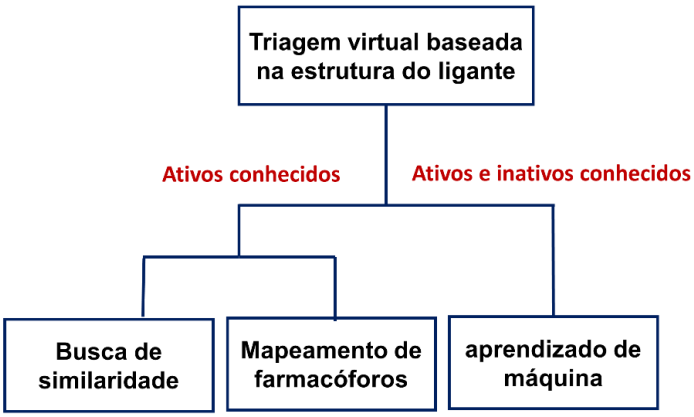


Figura 4: Classificação das metodologias de Triagem virtual baseada na estrutura do ligante. Adaptado de GILLET, 2013. (acesso em 2017-11-11)

Por uma parte, são observadas metodologias que se baseiam no uso de estruturas que apresentam atividade biológica. A aplicação destas técnicas no desenvolvimento de novas drogas está fundamentada no princípio da vizinhança, segundo o qual, moléculas estruturalmente similares tendem a ter propriedades similares (TEIXEIRA, 2014). Assim por exemplo, no mapeamento de farmacóforos, o objetivo é a identificação daqueles que contenham

o maior número de características comuns a todas as moléculas ativas, e que possam ser apresentadas por cada molécula em uma conformação de baixa energia (LEACH, 2007). Para o desenvolvimento destes procedimentos são usados principalmente os denominados métodos de *fingerprint* (ECKERT, 2007)

A sua vez, os métodos de aprendizado de máquina que utilizam tanto a informação de moléculas ativas como inativas, são os mais usados em triagem virtual, alcançando grandes sucessos. Em geral, estas técnicas utilizam bancos de dados de compostos com valores de atividade biológica (sejam ativos ou inativos, dependendo de um critério numérico de corte) e com o uso de diferentes algoritmos são criados modelos matemáticos a partir de descritores os quais são obtidos a partir das estruturas (2D ou 3D) dos ligantes para serem empregados na triagem virtual (BARROS, 2017).

Entre as principais abordagens usadas encontram-se as árvores de decisão (*Decision Trees*), floresta aleatória (*Random Forest),* mapas auto organizáveis (*Self-Organizing Map*s -SOMs), redes neurais artificiais e as máquinas de vetores de suporte (DONALEK, 2011).

2.1.4. Aprendizado de máquina

As técnicas mais comuns de aprendizado de máquina podem ser classificadas em dois grupos: não supervisionadas e supervisionadas. Os algoritmos não supervisionados são aqueles nos quais somente os dados de padrões de entrada são utilizados na construção do modelo (BRAGA, 2000). Assim, estes métodos permitem o agrupamento de dados de entrada baseado exclusivamente nas suas propriedades estatísticas. Entre os algoritmos mais comuns neste grupo, podemos citar os denominados mapas auto organizáveis (DONALEK, 2011; GERTRUDES, 2012).

Um mapa auto organizável, é um gráfico de similaridade e um diagrama de agrupamento, categorizado como uma rede neural artificial não supervisionada. É uma ferramenta que permite a visualização de dados de alta dimensionalidade. O algoritmo faz um mapeamento ordenado de uma distribuição de alta dimensionalidade num grid de baixa dimensionalidade (figura 5a). O algoritmo SOM calcula os modelos para que eles descrevam otimamente o domínio de observações (discretas ou continuamente distribuídas). Assim, é capaz de converter relações

estatísticas complexas, sem linhas em elementos de dados de alta dimensão em relações geométricas simples em uma tela de baixa dimensão (KOHONEN, 1998).

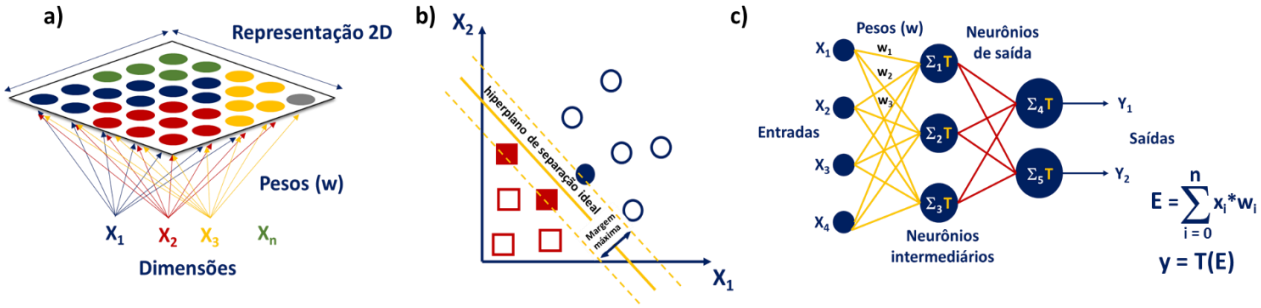


Figura 5: a) Representação de um mapa auto organizável (SOM). b) uma máquina de vetor de suporte (SVM) e c) uma rede neural artificial. T= função de ativação.

Em tanto, nas técnicas supervisionadas os dados do treino incluem tanto o valor de entrada como o resultado desejado, gerando métodos usualmente mais rápidos e precisos, para isso é fundamental a construção de um modelo de treino, validação e teste adequado. Neste grupo são encontradas, máquinas de vetores de suporte, árvores de decisão e redes neurais bayesianas (DONALEK, 2011; GERTRUDES, 2012).

As máquinas de vetores de suporte (SVM, do inglês *Support Vector Machines*) são um conjunto de métodos relacionados para o aprendizado supervisionado, aplicável tanto a problemas de classificação como de regressão. Um classificador SVM cria um hiperplano de margem máxima que se encontra em um espaço de entrada transformado e divide as classes de exemplo (figura 5b), ao mesmo tempo em que maximiza a distância para os exemplos de limpeza limpa mais próximos (SCHÖLKOPF, 2005).

As redes neurais artificiais são algoritmos que possuem a capacidade de aprender por exemplos e fazer interpolações e extrapolações do que aprendem. Ao conjunto de processos bem-definidos para adaptar os parâmetros de uma rede neural para que esta possa aprender uma determinada função, denomina-se algoritmo de aprendizado (BRAGA, 2000).

Assim, a fase fundamental para a utilização de uma rede neural na solução de uma tarefa é a fase de aprendizagem, a qual é o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando, sendo o tipo específico de aprendizagem realizada pela maneira particular como ocorrem os ajustes realizados nos parâmetros (BRAGA, 2000).

Na figura 5c representa-se a arquitetura de uma rede neural artificial. A partir dos valores das entradas, pesos arbitrários (w) são atribuídos para ajustar os neurônios no processo de treino com a finalidade que os valores dos neurônios de saída coincidam com os resultados observados. O valor de saída de um neurônio por Y que é uma função (função de ativação) que depende de E. O valor de E é obtido pela somatória do produto entre os dados de entrada e os respetivos pesos (CABELLO, 2005).

2.1.5 Random forest (Floresta aleatória)

Random forest (RF) é um algoritmo supervisionado que se baseia na combinação de árvores de predição, de modo que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e a mesma distribuição para todas as árvores na floresta (Figura 6). O erro para cada floresta converge até um limite, à medida que o número de árvores se torna maior. Assim, o erro da floresta depende da força das árvores individuais na floresta e da correlação entre eles (BREIMAN, 2001).

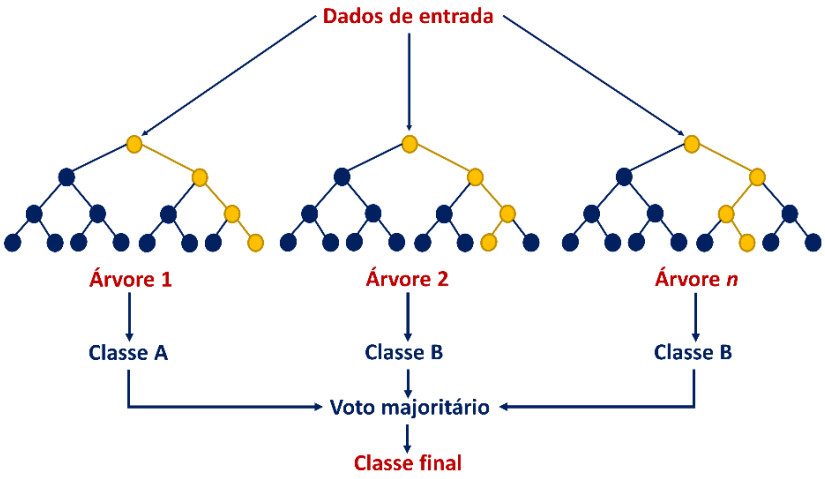


Figura 6: Representacao simplificada de um modelo *random forest*

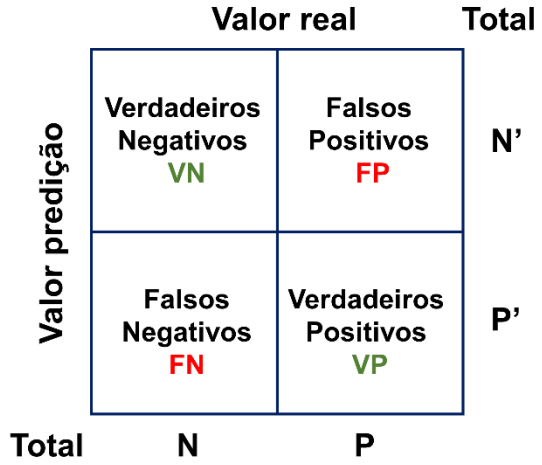
Entre as principais características do random forest, encontra-se que este algoritmo é relativamente simples, facilmente paralelizado e fornece estimativas úteis de erro, robustez, correlação e importância variável, sendo relativamente robusto para *outliers* (observações que estão numericamente distante do resto dos dados) e ruído. Assim mesmo, sendo comparado

com outros algoritmos de aprendizado de máquina, como *Adaboost*, *random forest* pode alcançar o mesmo nível de precisão sendo às vezes até melhor (BREIMAN, 2001).

O algoritmo Random Forest, tem sido projetado como uma solução muito eficaz para estudos de QSAR, sendo que esta metodologia parece ser útil porque cada floresta representa o consenso de um modelo não linear que é derivado de um grande número de modelos (árvores). Além disso, *random forest* apresenta múltiplas vantagens para o uso em procedimentos QSAR: a) Os modelos *Random forest* são bastante resistentes à superposição, b) *random forest* não exige processos bastante complicados e demorados de seleção de variáveis e c) os compostos com vários mecanismos de ação poderiam ser estudados dentro do mesmo (único) conjunto de treinamento (POLISHCHUK, MURATOV, 2009)

2.1.6 Avaliação de classificadores

Para a verificação ou medição do rendimento dos modelos classificatórios é comumente usada a matriz de confusão. Esta matriz permite a visualização mediante uma tabela de contingência a distribuição dos erros cometidos por um classificador (CORSO, 2009). Considerando um modelo classificador de duas classes: positivos (P) e negativos (N) as escolhas são estruturadas para predizer a ocorrência ou não de um evento simples (Figura 7). De esta maneira, se uma instância positiva e corretamente predita, denomina-se verdadeiro positivo (VP), se esta é classificada de maneira incorreta, conta-se como falso negativo (FN). Assim mesmo, se um elemento negativo é corretamente predito, classifica-se como verdadeiro negativo (VN), enquanto o erro na classificação é denominado falso positivo (FP) (MONARD, 2003).



$$\text{Taxa de VP (TVP)} = \frac{VP}{VP + FN}$$

$$\text{Taxa de FP (TFP)} = 1 - \text{TVN}$$

$$\text{Taxa de VN (TVN)} = \frac{VN}{VN + FP}$$

$$\text{Taxa de FN (TFN)} = 1 - \text{TVP}$$

Figura 7: Matriz de confusão para duas classes.

A partir da matriz de confusão outras medidas são derivadas as quais permitem avaliar a qualidade do modelo desenvolvido (MONARD, 2003):

Acurácia: $\text{Acc} = \frac{V_p + V_N}{P+N}$ Sensibilidade $\text{Sen} = \frac{V_p}{V_P + F_n}$

Especificidade: $\text{Spe} = \frac{V_n}{V_n + F_p}$ Eficiência: $\text{Efi} = \frac{\text{Sen}+\text{Spe}}{2}$

Valor preditivo negativo: $\text{Vpn} = \frac{V_n}{V_n + F_n}$ Valor preditivo positivo: $\text{Vpp} = \frac{V_p}{V_p + F_p}$

A curva ROC (do inglês *Receiver operating characteristics)*, é uma técnica para visualizar, organizar e selecionar classificadores com base no seu desempenho (FAWCETT, 2006), sendo atualmente um dos principais e mais comuns avaliadores da qualidade de modelos classificadores. O gráfico bidimensional é realizado a partir dos valores de sensibilidade e especificidade (Figura 8).

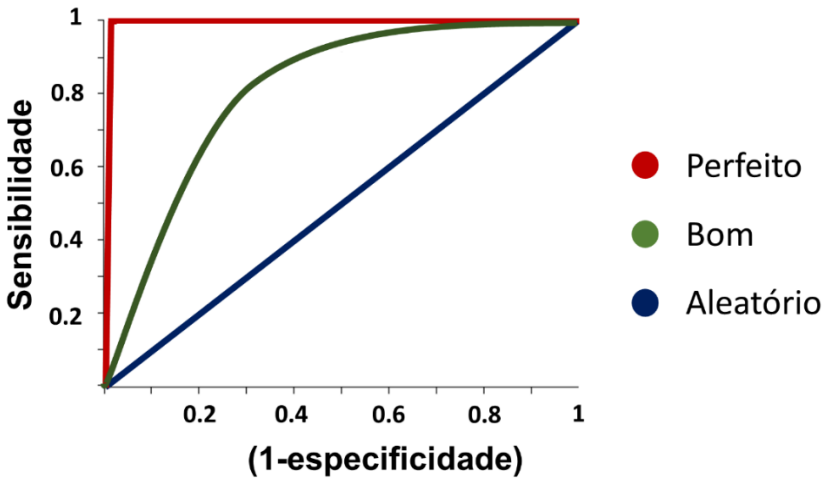


Figura 8: Curva ROC (*Receiver operating characteristics*)

Um gráfico ROC representa a taxa de verdadeiros positivos (Sensibilidade) no eixo Y versus a taxa de falso positivos (1-especificidade) localizada no eixo X. A partir deste é calculada a área sob a curva, a qual pode ter valores entre zero (0) e um (1) (HANLEY, 1982).

Assim, quando uma variável de interesse não pode ser distinguida entre os dois grupos (aleatória), a área sob a curva ROC será igual a 0,5 (figura 8, azul), enquanto um caso oposto onde exista uma separação perfeita dos valores de dois grupos sem sobreposição de distribuições, a área sob a curva ROC será igual a 1 (Figura 8, vermelho). Por tanto aqueles modelos classificadores considerados bons, serão aqueles com uma baixa taxa de falso positivos com um alto nível de acertos das instâncias (Figura 8, verde).

Outro parâmetro que avalia a qualidade dos modelos classificadores e o coeficiente de correlação de Matthews (MCC, do inglês *Matthews Correlation Coefficient*) que relaciona todas as classificações observações e preditas obtidas na matriz de confusão, por médio da equação:

$$\mathrm{MCC} = \frac{\mathrm{VP\ x\ VN} - \mathrm{FP\ x\ FN}}{\sqrt{(\mathrm{VP}+\mathrm{FP})(\mathrm{VP}+\mathrm{FN})(\mathrm{VN}+\mathrm{FP})(\mathrm{VN}+\mathrm{FN})}}$$

Os resultados do MCC podem variar entre (-1) e 1. Um valor MCC igual a um (1) indica uma correlação perfeita, zero (0) uma previsão aleatória, enquanto que (-1) é um indicativo de desacordo total entre predição e observação (MATTHEWS, 1975).

Assim, para um modelo classificatório feito com 100 dados que contém 50 estruturas classificadas como ativas (P) e 50 classificadas como negativas (N) sendo encontrado um

número de 10 falsos positivos e 15 falsos negativos, o valor do MCC calculado para o modelo a partir dos valores de VP (35), VN (40), FP (10) e FN (15) (Figura 7) é igual a 0,502.

2.1.7 CADD baseado na estrutura do receptor

A união do ligante ao receptor é fundamental nas reações enzimáticas e por tanto na inibição destas. Um estudo detalhado das interações entre as moléculas e as proteínas alvo é importante na estratégia do desenho racional de drogas (RUICK, 2016). Assim, a segunda classe de técnicas de CADD, engloba aquelas que são baseadas na estrutura do receptor, as quais têm emergido como métodos confiáveis e baratos para a identificação de candidatos a *leads* (LYNE, 2002).

Estas são abordagens computacionais, utilizam a estrutura 3D do alvo biológico, obtida a partir de raios-X, ressonância magnética nuclear (RMN) ou modelagem computacional, para encaixar uma coleção de compostos químicos no local de ligação, num procedimento comumente conhecido como *docking* molecular (LI & SHAH, 2017).

Inicialmente, é importante definir o reconhecimento molecular como aquele processo de formação de um complexo entre macromoléculas biológicas que interagem entre si com uma ou várias moléculas pequenas através de interações não covalentes. Este processo está caraterizado pela especificidade, que é explicada como a preferência de um ligante em respeito aos outros, e a afinidade, que determina que uma alta concentração de ligantes de interação fraca não pode substituir o efeito de uma baixa concentração do ligante específico interagindo com alta afinidade (DU, 2016).

Diversos fatores estão envolvidos nos processos de reconhecimento molecular, destacam-se as interações entálpicas e entrópicas, flexibilidade do receptor e do ligante, interações com moléculas de água, o efeito do ambiente proteico entre outros. Devido a essa complexidade, no *docking* molecular proteína - ligante, algumas simplificações são feitas, desde que a especificidade e a afinidade na predição sejam mantidas na medida do possível (DU, 2016).

Cada programa de *docking* faz uso de um ou mais algoritmos de busca específicos, que são os métodos usados para prever as possíveis conformações de um complexo binário (DIAZ & de AZEVEDO, 2008). Igualmente, os softwares especializados nesta área, calculam uma

pontuação no *docking* molecular a qual descreve de maneira aproximada, a soma dos diferentes tipos de interações presentes no complexo formado (ligações de hidrogênio, interações eletrostáticas, interações de van der Waals), representando a potencialidade da união entre receptor-ligante (PAGADALA, 2017).

A figura 9, apresenta de maneira geral o procedimento de uma triagem virtual baseada na estrutura do receptor (*Docking* molecular) (CHENG, 2012). Inicialmente a preparação tanto da estrutura alvo como das estruturas a serem testadas é fundamental. No caso do receptor, remoção de moléculas de água, de outras subunidades da proteína, cofatores, etc, são os procedimentos comuns; enquanto que para os ligantes, geração das estruturas 3D assim como procedimentos de optimização da estrutura são realizados.

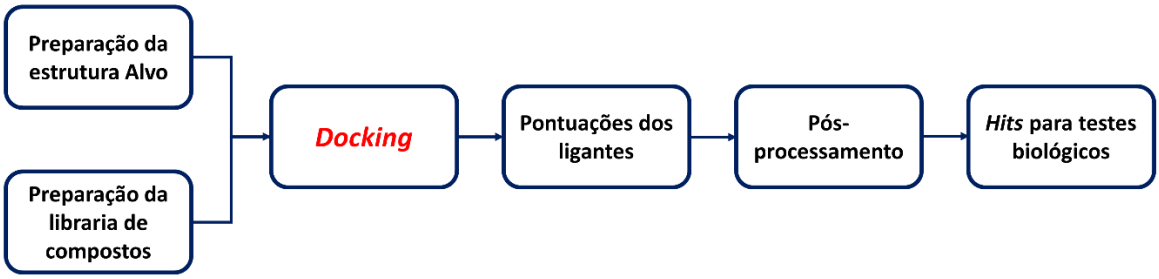


Figura 9: Fluxograma de uma triagem virtual baseada na estrutura do receptor (*Docking* molecular). Adaptado de CHENG, 2012.

Uma vez preparadas as moléculas o *Docking* é feito usando uma das múltiplas opções de software disponíveis. Os resultados podem ser analisados diretamente a partir das pontuações dadas pelo software, ou processos matemáticos adicionais podem ser realizados com o objetivo de diminuir a taxa de falsos positivos ou de refinar a seleção dos compostos com maior capacidade inibitória. Finalmente, as melhores estruturas são selecionadas para realizar testes biológicos permitindo por esta metodologia diminuir os custos económicos e operacionais na busca de *leads* (CHENG, 2012).

Nas últimas duas décadas, mais de 60 diferentes ferramentas de *docking* e programas têm sido desenvolvidos para fins académicos e comerciais, AutoDock FlexX, Surflex, GOLD, MOE-Dock, AutoDock Vina, são alguns dos mais destacados entre a grande variedade de opções disponíveis, no entanto os cálculos de *docking* com receptores flexíveis ainda apresenta um grande desafio para os programas atualmente disponíveis (PAGADALA, 2017).

O Molegro Virtual Docker é um ambiente integrado para estudar e prever como ligantes interagem com macromoléculas, o qual usa dois algoritmos para realizar a pontuação do complexo: MolDock Scoring Function e PLANTS scoring functions, sendo a energia de pontuação (E_{score}) dada pela soma dos termos energéticos de interação ligante-proteína (E_{inter}) e da energia interna do ligante (E_{intra}) (THOMSEN, 2006).

$$E_{score} = E_{inter} + E_{intra}$$

No entanto, funções de pontuação empírica pela natureza aditiva, geralmente exibem uma dependência do peso molecular, resultando numa tendencia para estruturas mais pesadas entrar no topo de uma lista classificatoria de posturas de moléculas (CARTA, 2007). Múltiplas soluções têm sido utilizadas para este problema como: diferentes métodos de normalização dos valores de pontuação respeito ao peso molecular dos compostos, assim como o uso de escores ponderados que otorgue pesos diferentes aos termos da função de pontuação usada (IBRAHIM, 2015; LUO,2017).

O software Molegro, apresenta o *Rerank score,* uma função de pontuação computacionalmente mais dispendiosa do que a função de pontuação utilizada durante a simulação do *docking*, mas geralmente é melhor do que a função de pontuação na determinação da melhor pose entre várias poses do mesmo ligante. O *Rerank Score* usa uma combinação ponderada dos termos aplicados pela pontuação MolDock combinada com alguns termos de adição, que inclui termos estéricos, como aproximações de Lennard-Jones energia estérica (LASKOWSKI, 2007).

3. OBJETIVOS

3.1 OBJETIVO GERAL

Utilizar, criar e combinar diversas metodologias de triagem virtual de banco de dados de produtos naturais através de docking molecular e aprendizado de máquina usando descritores moleculares para propor sesquiterpenos lactonizados com potencial atividade leishmanicida e tripanocida multialvo.

3.2 OBJETIVOS ESPECÍFICOS

Desenvolver um banco de dados de sesquiterpenos lactonizados cujas informações serão disponibilizadas na ferramenta web disponível no endereço para livre consulta, http://sistematx.ufpb.br

Estudar as interações moleculares em diversos meios para auxiliar nos estudos de triagem virtual e/ou relação estrutura atividade, através do campo de força GRID

Realizar uma triagem virtual nos bancos de dados de produtos naturais em busca de sesquiterpenos lactonizados com promissora atividade antiprotozoária, tanto utilizando modelos baseados nos ligantes como modelos utilizando informações da estrutura alvo (enzimas)

REFERENCIAS

ABDOLMALEKI, A.; B GHASEMI, J.; GHASEMI, F. Computer Aided Drug Design for Multi-Target Drug Design: SAR/QSAR, Molecular Docking and Pharmacophore Methods. **Current Drug Targets**, v. 18, n. 5, p. 556-575, 2017.

ACHARYA, C.; et al. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach**. Current computer-aided drug design**, v. 7, n. 1, p. 10-22, 2011.

BARROS, R.P.C. Triagem virtual de metabólitos secundários com potencial atividade antimicrobiana do gênero Solanum e estudo fitoquímico de Solanum capsicoides all. 2017.

BERNARDINI, S. et al. Natural products for human health: an historical overview of the drug discovery approaches. **Natural Product Research**, p. 1-25, 2017.

BOHM, B.A.; STUESSY, T.F. Flavonoids of the sunflower family (Asteraceae). Springer Vienna: 2001.

BRAGA, A. de P.; CARVALHO, A.P.L.F; LUDERMIR, T. B. Redes neurais artificiais: teoria e aplicações. Livros Técnicos e Científicos, 2000.

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

BUCKLE, D.R. et al. Glossary of terms used in medicinal chemistry. Part II (IUPAC Recommendations 2013). **Pure and Applied Chemistry**, v. 85, n. 8, p. 1725-1758, 2013.

CABELLO, J. T. et al. Redes neuronales artificiales en Medicina Intensiva. Ejemplo de aplicación con las variables del MPM II. **Medicina intensiva**, v. 29, n. 1, p. 13-20, 2005.

CARAUS, L., et al. Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions. **Briefings in bioinformatics**, v. 16, n. 6, p. 974-986, 2015.

CARTA, G.; KNOX, A.J.S.; LLOYD, D.G. Unbiasing scoring functions: a new normalization and rescoring strategy. **Journal of chemical information and modeling**, v. 47, n. 4, p. 1564-1571, 2007.

CHATURVEDI D.; DWIVEDI P.K., Recent Developments on the Antidiabetic Sesquiterpene Lactones and Their Semisynthetic Analogues, In: Discovery and Development of Antidiabetic Agents from Natural Products, Elsevier, 2017, p. 185-207.

CHENG, T.; et al. Structure-based virtual screening for drug discovery: a problem-centric review. **The AAPS journal**, v. 14, n. 1, p. 133-141, 2012.

CORSO, Cynthia Lorena. Aplicación de algoritmos de clasificación supervisada usando Weka. Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba, 2009.

CRUCIANI, G. et al. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. **Journal of Molecular Structure: THEOCHEM**, v. 503, n. 1, p. 17-30, 2000a.

CRUCIANI, G.; PASTOR, M.; GUBA, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. **European Journal of Pharmaceutical Sciences**, v. 11, p. S29-S39, 2000b.

DAS, P.S.; SAHA, P., A Review on Computer Aided Drug Design in Drug Discovery. **World Journal of Pharmacy and Pharmaceutical Sciences**, v. 6, n. 7, p. 279-291 2017.

DIAS, R.; DE AZEVEDO, Jr; WALTER, F. Molecular docking algorithms. **Current Drug Targets**, v. 9, n. 12, p. 1040-1047, 2008.

DONALEK, C. Supervised and Unsupervised learning. In: Astronomy Colloquia. USA. 2011.

DU, X.; et al. Insights into protein–ligand interactions: Mechanisms, models, and methods**. International journal of molecular sciences**, v. 17, n. 2, p. 144, 2016.

ECKERT, H.; BAJORATH, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. **Drug discovery today**, v. 12, n. 5, p. 225-233, 2007.

EMERENCIANO, V.P.; CABROL-BASS, D.; FERREIRA, M.J.; ALVARENGA, S.A.; BRANT, A.J.; SCOTTI, M.T.; BARBOSA, K.O. Chemical evolution in the Asteraceae. The oxidation-reduction mechanism and production of secondary metabolites. **Natural Product Communications**, p. 495-507, 2006.

FAWCETT, T. An introduction to ROC analysis. **Pattern recognition letters**, v. 27, n. 8, p. 861-874, 2006.

GERTRUDES, J. C. et al. Machine learning techniques and drug design. **Current medicinal chemistry**, v. 19, n. 25, p. 4289-4297, 2012.

GILLET, V. Ligand-Based and Structure-Based Virtual Screening. The University of Sheffield. 2003. Disponível em: https://www.ebi.ac.uk/sites/ebi.ac.uk/. Aceso em 22/10/2017.

GRAMATICA, P. A short history of QSAR evolution. 2008 Disponível em: <h http://qsarworld.com/TempFileupload/Shorthistoryofqsar.Pdf> aceso em 31/10/2017.

HAMMETT, L.P. The effect of structure upon the reactions of organic compounds. Benzene derivatives. **Journal of the American Chemical Society**, v. 59, n. 1, p. 96-103, 1937.

HANLEY, J.A.; MCNEIL, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. **Radiology**, v. 143, n. 1, p. 29-36, 1982.

HOQUE, I.; et al. An Approach of Computer-Aided Drug Design (CADD) Tools for In Silico Pharmaceutical Drug Design and Development. **International Journal of Advanced Research in Biological Sciences,** v. 4, n. 2, p. 60-71, 2017

IBRAHIM, T.M.; BAUER, M.R.; BOECKLER, F.M. Applying DEKOIS 2.0 in structure-based virtual screening to probe the impact of preparation procedures and score normalization. **Journal of cheminformatics**, v. 7, n. 1, p. 21, 2015.

IMAM, S. S.; GILANI, S. J. Computer Aided Drug Design: A Novel Loom To Drug Discovery. **Organic and Medicinal Chemistry**, v. 1, p. 1-6, 2017

IVANESCU, B.; MIRON, A.; CORCIOVA, A. Sesquiterpene lactones from Artemisia genus: biological activities and methods of analysis. **Journal of analytical methods in chemistry**, v. 2015, 2015.

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

KOEPPEN, H.; et al. Ligand-Based Virtual Screening. In: Virtual Screening: Principles, Challenges, and Practical Guidelines, p. 61-85, 2011.

KOHONEN, T. The self-organizing map. **Neurocomputing**, v. 21, n. 1, p. 1-6, 1998.

KUBINYI, H. Statistical methods. **QSAR: Hansch Analysis and Related Approaches**, p. 91-107, 2008.

LASKOWSKI, R. A. et al. Molegro virtual docker MVD-2007, user manual. Molegro Aps, Denmark, v. 10.

LEACH, A.R.; GILLET, V.J. Representation and Manipulation Of 3D Molecular Structures. An Introduction to Chemoinformatics, p. 27-52, 2007.

LI Q., SHAH S. Structure-Based Virtual Screening. In: Protein Bioinformatics. Methods in Molecular Biology, Humana Press, New York, v. 1558, 2017.

LIEBERT, M. Roundtable Discussion: High-Throughput Screening Challenges. **Genetic Engineering & Biotechnology News**, v. 28, n. 14, p. 26-27, 2008.

LYNE, P.D. Structure-based virtual screening: an overview. **Drug discovery today**, v. 7, n. 20, p. 1047-1055, 2002

LORENZO, V.P. Estudos in silico com alcalóides oriundos de produtos naturais. 2016.

LUO, Q.; et al. The scoring bias in reverse docking and the score normalization strategy to improve success rate of target fishing. **PloS one,** v. 12, n. 2, p. e0171433, 2017.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. **Biochimica et Biophysica Acta (BBA)-Protein Structure**, v. 405, n. 2, p. 442-451, 1975.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas Inteligentes-Fundamentos e Aplicações**, v. 1, n. 1, 2003.

NEWMAN, D.J.; CRAGG, G. M. Natural products as sources of new drugs from 1981 to 2014. **Journal of Natural Products**, v. 79, n. 3, p. 629-661, 2016.

ORGANIZAÇÃO MUNDIAL DA SAÚDE, Guidelines for the Treatment of Malaria. Geneva: World Health Organization. 2006.

PADILLA-GONZALEZ, G. F.; DOS SANTOS, F. A.; DA COSTA, F. B. Sesquiterpene Lactones: More Than Protective Plant Compounds with High Toxicity. **Critical Reviews in Plant Sciences**, v. 35, n. 1, p. 18-37, 2016.

PAGADALA, N.S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. **Biophysical reviews**, p. 1-12, 2017.

POLISHCHUK P.G.; MURATOV E.N.; ARTEMENKO A.G.; KOLUMBIN O.G.; MURATOV N.N.; KUZ'MIN V.E. Application of random forest approach to QSAR prediction of aquatic toxicity. **Journal of chemical information and modeling**, v. 49, n. 11, p. 2481-2488, 2009

RANDIĆ, M. Novel molecular descriptor for structure—property studies. **Chemical Physics Letters**, v. 211, n. 4-5, p. 478-483, 1993.

RIPPHAUSEN, Peter; NISIUS, Britta; BAJORATH, Jürgen. State-of-the-art in ligand-based virtual screening. **Drug discovery today**, v. 16, n. 9, p. 372-376, 2011.

ROY, K.; NARAYAN DAS, R. A review on principles, theory and practices of 2D-QSAR. **Current drug metabolism**, v. 15, n. 4, p. 346-379, 2014.

DE RUYCK, J.; et al. Molecular docking as a popular tool in drug design, an in silico travel. **Advances and applications in bioinformatics and chemistry: AABC**, v. 9, p. 1, 2016.

SCHÖLKOPF, B.; SMOLA, A. Support vector machines. Encyclopedia of Biostatistics, 2005.

SEAMAN, F.C. Sesquiterpene lactones as taxonomic characters in the Asteraceae. **The Botanical Review**, v.48, p. 121-594, 1982.

SEN, T.; SAMANTA, S. K., Medicinal plants, human health and biodiversity: a broad review. In: Biotechnological Applications of Biodiversity. Springer Berlin Heidelberg, 2014. p. 59-110.

SUBRAMANIAM, S.; MEHROTRA, M.; GUPTA, D. Virtual high throughput screening (vHTS)-A perspective. **Bioinformation,** v. 3, n. 1, p. 14, 2008.

TEIXEIRA, A. L.; FALCAO, A. O. Structural similarity based kriging for quantitative structure activity and property relationship modeling**. Journal of chemical information and modeling**, v. 54, n. 7, p. 1833-1849, 2014.

THOMAS, G. Medicinal chemistry: An introduction. John Wiley & Sons: 2011

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: a new technique for high-accuracy molecular docking. Journal of medicinal chemistry, v. 49, n. 11, p. 3315-3321, 2006.

TODESCHINI, R.; LASAGNI, M.; MARENGO, E. New molecular descriptors for 2D and 3D structures. Theory. **Journal of chemometrics**, v. 8, n. 4, p. 263-272, 1994.

TODESCHINI, R.; CONSONNI, V. Handbook of molecular descriptors. John Wiley & Sons, 2008.

TU, Y. The discovery of artemisinin (qinghaosu) and gifts from chinese medicine. **Nature Medicine**, v. 17, p. 1217-1220, 2011

YOSHIOKA, H.; MABRY, T.J.; TIMMERMANN, B.N. Sesquiterpene lactones: Chemistry, nmr and plant distribution. University of Tokyo Press: 1973.

.

CAPÍTULO II

Neste capítulo foi realizada uma revisão dos principais estudos *in vitro* e *in silico* realizados com sesquiterpenos lactonizados em busca de novos tratamentos contra quatro doenças tropicais negligenciadas: leishmaniose, doença de chagas, doença do sono e esquistossomose.

As doenças tropicais negligencias, são um grupo de doenças transmissíveis que afetam mais de um bilhão de pessoas das áreas tropicais e subtropicais de 149 países (OMS, 2017). Um dos principais problemas para o controle e eliminação deste grupo de doenças está relacionada com as quimioterapias usadas atualmente. Múltiplos efeitos colaterais, altos níveis de toxicidade e um aumento nos casos de resistência por parte do parasita têm sido observados em pacientes que sofrem destas doenças (ULIANA, 2017; SILVA, 2017).

A necessidade de novos tratamentos, mais eficazes e com menores níveis de toxicidade (ORTU,2017), permitem que as plantas, que historicamente têm sido uma fonte o tratamento de muitas doenças, emerjam como uma interessante alternativa para o desenvolvimento de novos fármacos contra esta classe de doenças. A família Asteraceae, que está composta por mais de 23,000 espécies ao redor do mundo principalmente no continente Americano é uma das mais estudadas neste campo devido a sua grande diversidade (BOHM & STUESSY, 2001).

Os sesquiterpenos lactonizados são uma classe de metabolitos secundários característicos desta família, estão compostos por um esqueleto de 15 átomos de carbonos contendo um anel lactona; possuem alta diversidade estrutural, sendo a maior parte destes (assim como as formas mais funcionais) de caráter cíclico. Têm sido usados principalmente na classificação taxonômica da família Asteraceae (SEAMAN 1982; CHADWICK, 2013).

Um dos principais casos de sucesso do desenvolvimento de novos medicamentos a partir de plantas é o descobrimento da Artemisinina também conhecido como qinghaosu, um sesquiterpeno lactonizado isolado de *Artemisia annua*, uma espécie frequentemente usada na medicina tradicional chinesa (PHILLIPS et al, 2017). Este composto reduz rapidamente os níveis de parasitemia possuindo toxicidade seletiva contra os parasitas de *Plasmodium* (MESHNICK, 1991).

Neste contexto, múltiplos estudos *in vitro* e *in vivo* usando sesquiterpenos lactonizados têm sido realizados na busca de moléculas com atividade promissória contra doenças negligenciadas.

Algumas espécies como *Pseudelephantopus spiralis*, *Tanacetum parthenium* e *Calea zacatechichi* C, possuem sesquiterpenos lactonizados com altos níveis de atividade contra as

diferentes formas parasitárias das espécies de *Leishmania* (GIRARDI, 2015; TIUMAN, 2005; WU, 2011). Na mesma direção, metodologias de fracionamento guiado por bioensaio têm permitido a identificação desta classe de metabolitos secundários com atividade *in vitro* contra *Trypanosoma cruzi* o parasita responsável da doença de chagas. Enidrina e uvedalina isolados da espécie *Smallanthus sonchifolius* assim como peruvina e psilostatina de *Ambrosia tenuifolia* possuem atividade contra epimastigotes de *T. cruzi* com valores de IC_{50} menores a 6,2 µM (FRANK, 2013; SULSEN, 2008).

Recentemente, as ferramentas quimioinformáticas têm emergido como uma importante alternativa no desenvolvimento de novos medicamentos, desde que resultados com sucesso podem ser alcançados com baixo custo e em pouco tempo. Dentro das principais metodologias usadas se encontram os modelos de relação estrutura atividade (QSAR pelas suas siglas em inglês) assim como estudos de *docking* molecular.

Para os sesquiterpenos lactonizados têm sido realizados diferentes estudos de classificação taxonômica, avaliações toxicológicas, *docking* molecular e modelos QSAR (DA COSTA, 2005; SCOTTI 2012). Aqui, são mencionadas algumas destas investigações, as quais têm permitido a identificação de novas moléculas com atividades promissórias contra doenças negligencias, aumentando o conhecimento respeito ao mecanismo de ação, a identificação de grupos farmacofóricos ou o incremento mesmo da atividade biológica destas.

REFERENCIAS

BOHM, B.A.; STUESSY, T.F. Flavonoids of the Sunflower Family (Asteraceae); Springer: Vienna, Austria, 2001.

CHADWICK, M.; TREWIN, H.; GAWTHROP, F.; WAGSTAFF, C. Sesquiterpenoids lactones: Benefits to plants and people. **International journal of molecular sciences,** v.14, p.127, 2013

DA COSTA, F.B.; TERFLOTH, L.; GASTEIGER, J. Sesquiterpene lactone-based classification of three Asteraceae tribes: A study based on self-organizing neural networks applied to chemosystematics. **Phytochemistry** v.66, p. 345–353, 2005.

FRANK, F.M.; ULLOA, J.; CAZORLA, S.I.; MARAVILLA, G.; MALCHIODI, E.L.; GRAU, A.; MARTINO, V.; CATALAN, C.; MUSCHIETTI, L.V. Trypanocidal activity of Smallanthus sonchifolius: Identification of active sesquiterpene lactones by bioassay-guided fractionation. **Evidence-based complementary and alternative medicine** v. 2013, 2013

GIRARDI, C.; FABRE, N.; PALOQUE, L.; RAMADANI, A.P.; BENOIT-VICAL, F.; GONZALEZ-ASPAJO, G.; HADDAD, M.; RENGIFO, E.; JULLIAN, V. Evaluation of antiplasmodial and antileishmanial activities of herbal medicine Pseudelephantopus spiralis (less.) cronquist and isolated hirsutinolide-type sesquiterpenoids**. Journal of ethnopharmacology.** v.170, p.167–174. 2015.

MESHNICK, S.R. et al. Artemisinin (qinghaosu): the role of intracellular hemin in its mechanism of antimalarial action. **Molecular and Biochemical Parasitology,** v.49, p. 181-189, 1991.

ORGANIZAÇÃO MUNDIAL DA SAÚDE, Neglected Tropical Diseases, Disponível em: <http://www.who.int/neglected_diseases/en/>. Aceso em 31/10/2017.

ORTU, G.; WILLIAMS, O. Neglected tropical diseases: exploring long term practical approaches to achieve sustainable disease elimination and beyond. **Infectious Diseases of Poverty**, v.6, p. 147, 2017

PHILLIPS. M.A. et al. Malaria. **Nature Reviews Disease Primers,** v.3, 2017

SCOTTI, M.T.; EMERENCIANO, V.; FERREIRA, M.J.; SCOTTI, L.; STEFANI, R.; DA SILVA, M.S.; MENDONCA JUNIOR, F.J. Self-organizing maps of molecular descriptors for

sesquiterpene lactones and their application to the chemotaxonomy of the Asteraceae family. **Molecules**. v.17, p.4684–4702, 2012.

SEAMAN, F.C. Sesquiterpene lactones as taxonomic characters in the Asteraceae. **The Botanical Review**, v.48, p. 121-594, 1982.

SILVA, D.G. et al. New Class of Antitrypanosomal Agents Based on Imidazopyridines. **ACS Medicinal Chemistry Letters**, v. 8, p. 766-770, 2017.

SULSEN, V.P.; FRANK, F.M.; CAZORLA, S.I.; ANESINI, C.A.; MALCHIODI, E.L.; FREIXA, B.; VILA, R.; MUSCHIETTI, L.V.; MARTINO, V.S. Trypanocidal and leishmanicidal activities of sesquiterpene lactones from Ambrosia tenuifolia sprengel (Asteraceae). **Antimicrobial Agents Chemotherapies**. v.52, p.2415–2419. 2008.

TIUMAN, T.S. et al**.** Antileishmanial activity of parthenolide, a sesquiterpene lactone isolated from Tanacetum parthenium**. Antimicrobial Agents Chemotherapies,** v.49, p. 176–182, 2005.

ULIANA S.R.B.; TRINCONI C.T.; COELHO A.C., Chemotherapy of leishmaniasis: present challenges, **Parasitology,** pp. 1-17, 2017

WU, H.K.; FRONCZEK, F.R.; BURANDT, C.L.; ZJAWIONY, J.K. Antileishmanial Germacranolides from Calea Zacatechichi. **Planta Medica**. v.77, p.749–753. 2011,

Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases

Chonny Herrera Acevedo, Luciana Scotti, Mateus Feitosa Alves, Margareth De Fátima Formiga Melo Diniz and Marcus Tullius Scotti *

Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil; caherrera@ltf.ufpb.br (C.H.A.); luciana.scotti@gmail.com (L.S.); mateusfalves@gmail.com (M.F.A.); margareth@ltf.ufpb.br (M.F.F.M.D.)

*Correspondence: mtscotti@gmail.com; Tel.: +55-83-99869-0415

Artigo publicado na revista ***Molecules*** *2017, 22(1), 79.*

DOI:10.3390/molecules22010079

Fator de Impacto: 2.861

Abstract: This review presents a survey to the biological importance of sesquiterpene lactones (SLs) in the fight against four infectious neglected tropical diseases (NTDs)—leishmaniasis, schistosomiasis, Chagas disease, and sleeping sickness—as alternatives to the current chemotherapies that display several problems such as low effectiveness, resistance, and high toxicity. Several studies have demonstrated the great potential of some SLs as therapeutic agents for these NTDs and the relationship between the protozoal activities with their chemical structure. Recently, Computer-Aided Drug Design (CADD) studies have helped increase the knowledge of SLs regarding their mechanisms, the discovery of new lead molecules, the identification of pharmacophore groups and increase the biological activity by employing *in silico* tools such as molecular docking, virtual screening and Quantitative-Structure Activity Relationship (QSAR) studies.

Keywords: computer-aided drug design; neglected tropical diseases; sesquiterpene lactones; secondary metabolites; leishmaniasis; Chagas disease; sleeping sickness; schistosomiasis; QSAR studies

1. Introduction

Neglected tropical diseases (NTDs) are diseases that are present around the world in tropical and subtropical regions. They affect about one billion people, who commonly live in poverty [1], from 150 countries, causing some 140,000 deaths per year [2]. These types of diseases are attributable to inadequate access to safe water, sanitation, and appropriate housing [3], which presents difficulties in their prevention, control, and elimination, and generate costs of billions of dollars a year in those countries [1].

According to the classification of the World Health Organization, 19 NTDs exist [4], of which eight—Buruli ulcer, trachoma, leprosy, rabies, endemic treponematoses, mycetoma, dengue, and chikungunya—are caused by viruses or bacteria, while 11 others—dracunculiasis, soil-transmitted helminthiases, echinococcosis, foodborne trematodiases, lymphatic filariasis, onchocerciasis, taeniasis, schistosomiasis, American trypanosomiasis (Chagas disease), human African trypanosomiasis (sleeping sickness), and leishmaniasis—are classified as parasitic diseases. The last three, which are generated by different kinetoplastic parasite species [5], have more impact worldwide due to the approximately two million new cases and more than 50,000 deaths annually, which combined with schistosomiasis are responsible for about 60% of all the deaths resulting from NTDs [2,6].

Antimonial compounds are the main treatment against leishmaniasis, which presents high toxicity and resistance in some endemic regions. Nifurtimox and benznidazole, the drugs employed against *Trypanosoma cruzi*, present many side effects in patients. The low effectivity in the chronic phase of the disease, the high risk of death from melarsoprol, an arsenic derivative for treating the chronic phase in human African trypanosomiasis, and the increase in cases of resistance of *Schistosoma* to praziquantel treatment are some examples of chemotherapy actually used for NTDs that present several problems and prevent the control and elimination of these diseases.

Because of this, it is necessary to discover new low-cost drugs that are highly effective with low toxicity. Humans have thus turned to plants as a source of compounds for the treatment of many diseases, and in addition to being an option to find solutions for many health problems of most of the world's population, is an important industry, as plants generate over $25 billion dollars per year in value based only on the relevant pharmaceuticals derived from plants [7].

One of the most studied plant families in the natural products research field is the Asteraceae (Compositae). It is the largest family of flowering plants, formed by more than 23,000 species that are present in every continent, principally in the New World [8]. Some molecules found in this family have been successfully used in humans against parasite diseases, such artemisinin, an antimalarial sesquiterpene lactone which discovery led to the 2015 Nobel Prize for Medicine and Physiology [9,10]. This drug, which has the most rapid action against the malaria caused by *Plasmodium falciparum*, opened the panorama with respect to research into new therapeutics for leishmaniasis, trypanosomiasis, and schistosomiasis from this type of secondary metabolite, and in addition to the antimalarial activity, the drug has antitumor potential and shows prevention of neurodegeneration, antimigraine activity, analgesic, and sedative activities [11].

In recent years, cheminformatics tools have served in the taxonomic classification of the sesquiterpene lactones [12-14], cytotoxicity assays [15-17], as well as the continuous search for new drugs or lead compounds through the development of new methodologies in Computer-Aided Drug Design (CADD). These in silico studies have emerged as a good alternative in medicinal chemistry for helping to establish the experimental design, find new drugs, and compare the molecular structure and activity presented through one of the most important tools in this area, quantitative structure–activity relationship (QSAR) studies, generating successful results with low cost in a short time.

In brief, this review summarizes the principal available information with respect to the sesquiterpene lactones in the search for effective and safe drugs against four infectious NTDs (leishmaniasis, human African trypanosomiasis, Chagas disease, and schistosomiasis), highlighting the corresponding QSAR studies.

2. The Current Chemotherapy against Infectious NTDs

2.1. Leishmaniasis

Leishmaniasis are anthroponotic and zoonotic diseases caused by approximately 20 protozoan parasite species of the genus *Leishmania*, which are transmitted to humans by more than 30 different species of phlebotomine sandflies [18,19]. In accordance with the clinical manifestations, the three main forms of leishmaniasis are cutaneous, mucocutaneous, and visceral [19,20]. Visceral leishmaniasis or kala-azar is the most severe form of the disease, being fatal in the absence of treatment [21], and some studies report that even 5%–10% of patients under medication die [22].

For over five decades, the main treatment has consisted of pentavalent antimonial compounds such as sodium stibogluconate (**1**, Figure 1) and meglumine antimoniate (**2**, Figure 1), the efficacy of which is between 80% and 95% [23-25]. They are low cost antileishmanial drugs, easily available in endemic regions [26], however, resistance is observed in some regions [27,28], and toxicity [29] and the long duration of treatment (up to 28 days of parenteral administration) [30] have emerged as the principal problems for the use of those drugs, which has prompted the search for new compounds. Another drug, pentamidine (**3**, Figure 1), initially used in India for visceral leishmaniasis, was discontinued due to the many adverse reactions related with its use, however, it was recently re-introduced in Ethiopia for the treatment of HIV–cutaneous leishmaniasis co-infections [26]. Amphotericin B (**4**, Figure 1) presents the highest effectivity, but the treatment requires prolonged hospitalization in addition to producing side effects such as fever, vomiting, abdominal pain, and nephrotoxicity. The liposomal form is less toxic, but both forms are very expensive [19,20]. Miltefosine (**5**, Figure 1) and paromomycin (**6**, Figure 1), which have also been used recently as antileishmanial drugs, present several problems, such as toxicity (mainly renal), increased resistance, and teratogenic and abortifacient effects [31,32].

2.2. Trypanosomiasis

Trypanosomiasis is the name of the disease caused by a protozoan parasite of the genus *Trypanosoma*. Two types of trypanosomiasis can affect humans: Chagas disease (American trypanosomiases) and sleeping sickness (human African trypanosomiases).

Chagas disease is the highest impact parasitic disease in the Americas [33], affecting about 8 million persons, mainly in the poorest regions of Central and South America. It is attributed 10,600 deaths per year [2], usually as a result of heart failure or thromboembolic episodes. The infection is caused by the flagellated protozoan *Trypanosoma cruzi*, which is transmitted to humans by the *Triatoma* bug vector.

There is a limitation in the treatments currently used against Chagas disease as observed in leishmaniasis because several studies have shown inefficiency, resistance [34], toxic side effects [35], and reactivation of infections that mainly appear due to the increase of *T. cruzi*–HIV co-infections [36]. The chemotherapy of Chagas' disease currently uses two drugs: nifurtimox (**7**, Figure 1) and benznidazole (**8**, Figure 1) [37]. The effectivity is close to 100%, but not in all cases. Various difficulties are presented with these drugs such as low rates of reduction of parasitemia in the chronic phase and mainly in the late phase of the disease,

resistance of some strains of *T. cruzi*, and many side effects such as nausea, pain, and vomiting, in addition to a prolonged treatment time that in some cases can be up to two months. These difficulties have led to the continuous search for alternatives for the elimination of these diseases [37,38].

Sleeping sickness is transmitted by the bite of insects of the *Glossina* genus, also known as tse-tse fly, which acts as a vector for different species of *Trypanosoma brucei* [39]. In 2014, 3796 cases were reported [40] and is an NTD mainly present in sub-Saharan Africa.

The treatment of sleeping sickness is divided in accordance with the disease phases, and only a few drugs with many side effects and limited efficacy are used. For the first phase, where the parasites are restricted to the hemolymphatic system, the above-mentioned pentamidine (**3**) together with suramin (**9**, Figure 1) are the drugs used against *T. brucei gambiense* and *T. brucei rhodesiense*, respectively [41,42]. Both are relatively effective against the parasite, but they have side effects such as allergy, fatigue, neuropathy, renal problems, and nausea [42].

The treatment in the second phase presents more difficulties due to the fact *T. brucei* is present in the cerebrospinal fluid. For this reason, the drugs for the treatment in this stage must have the ability to cross the blood–brain barrier [41]. Melarsoprol (**10**, Figure 1) is an arsenic-derived compound that was for a long time the only compound available to treat patients at this stage of the disease. The drug has many adverse effects such as high risk of death, including liver toxicity, severe enterocolitis, diffuse peripheral neuropathy, and encephalopathic syndromes that occur in 2%–10% of patients, leading to death in half of these cases [41,43,44]. Eflornithine (**11**, Figure 1) is safer than melarsoprol and has emerged as an alternative because studies with patients in the second stage of human African trypanosomiasis presented better results with respect to the cutaneous and neurological adverse effects and death risk factor, but these positive results are only observed against *T. brucei gambiense* [44,45].

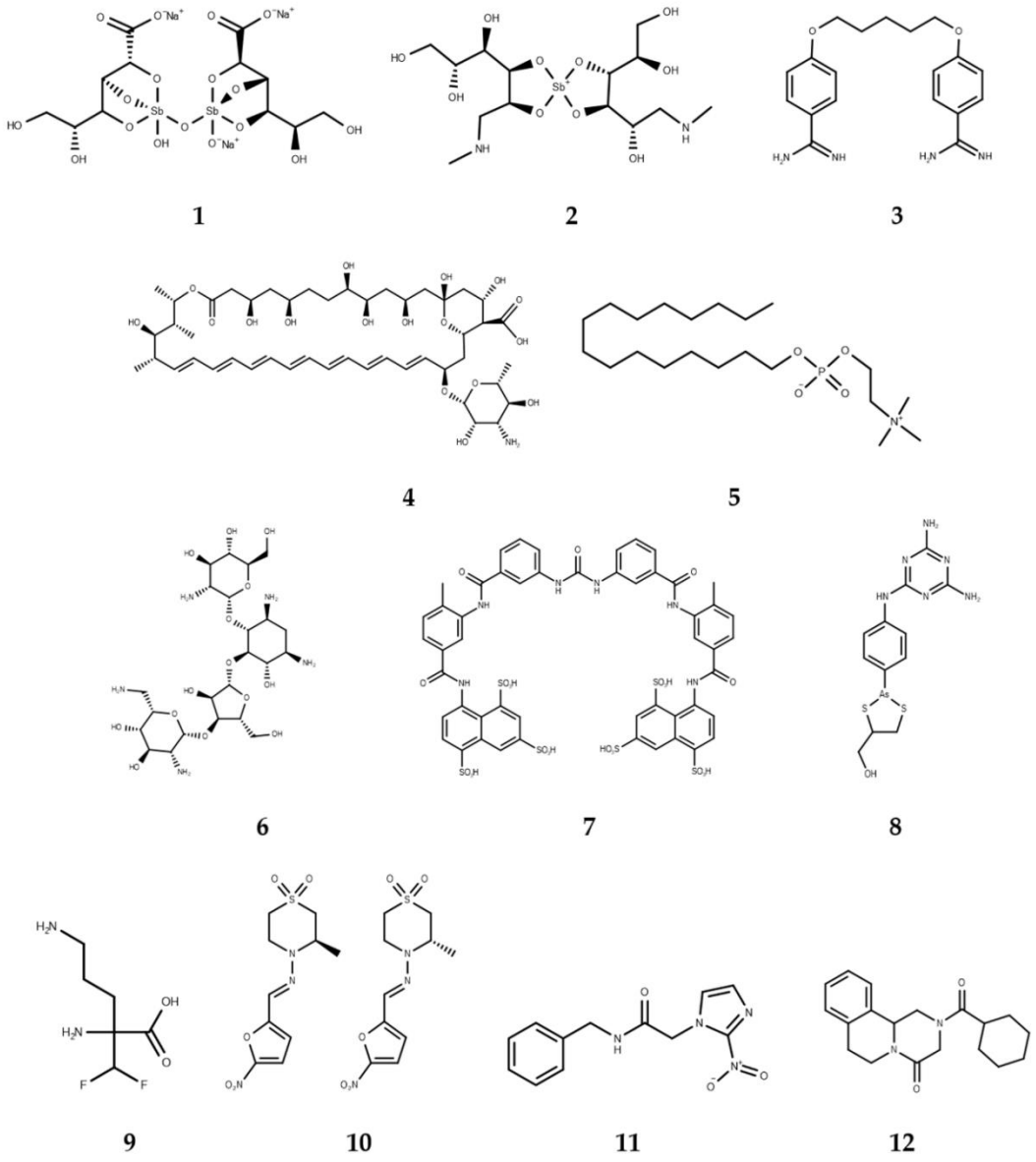


Figure 1. Chemotherapy used against four neglected tropical diseases.

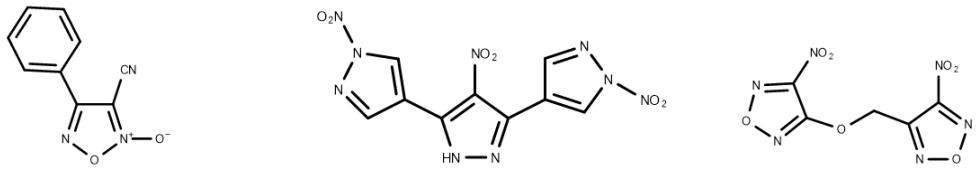
2.3. Schistosomiasis

Schistosomiasis is an NTD caused by trematode flukes of the genus *Schistosoma* spp., which affects more than 200 million people, most commonly in the sub-Saharan region in Africa [46]. The chronic infection causes anemia, neurological complications, intestinal fibrosis veins, hepatosplenomegaly, stunted growth, and other negative conditions, and this infection leads to approximately 5500 deaths each year [2,47]. It is anticipated for the morbidity caused by schistosomiasis will be controlled and eliminated as a public health problem in 2020 and 2025, respectively. To reach this goal, efforts in the social aspects of treatment such as sanitation,

provision of water and hygiene facilities, and effective treatments for the elimination of parasites are necessary [46,48].

Praziquantel (**12**, Figure 1) is the drug that has been used in chemotherapy since 1970 for treating schistosomiasis and is almost the only available drug against this disease [49,50]. In contrast to the state of antileishmanial and antitrypanosomal treatments, praziquantel, besides its high efficacy, is safe and does not have toxicity problems [49], however, in the last two decades, some studies have shown schistosomiasis cases caused by *S. haematobium* infections in which repeated standard treatment fails to clear the infection, while *S. mansoni* presents a reduced susceptibility in field isolates and may develop resistance to schistosomicidal drugs over the course of relatively few passages [50-52].

Oxadiazoles have been studied in the search for alternative drugs against praziquantel-resistant parasites. Incubation of *S. mansoni* parasites with oxadiazole 2-oxides leads to parasite death by the inhibition of thioredoxin glutathione reductase (TGR), which is a crucial enzyme whose role is in the redox homeostasis of the *Schistosoma* [53]. This antischistosomal activity was associated with the donation of nitric oxide [54]. This activity was evaluated *in vivo, S. mansoni*-infected mice were treated with 4-phenyl-1,2,5-oxadiazole-3-carbonitrile-2-oxide (13, Figure 2) and worm burden reductions of 99%, 88%, and 94% were observed from treatments against skin-, liver-, and adult-stage parasites, respectively [53]. *In silico* QSAR models using consensus analysis and virtual screening led to the proposal of two novel chemical scaffolds with antischistosomal activity: 4-nitro-3,5-bis(1-nitro-1H-pyrazol-4-yl)-1H-pyrazole (**14**, Figure 2) and 3-nitro-4-{[(4-nitro-1,2,5-oxadiazol-3-yl)oxy]meth-1,2,5-oxadiazole (**15**, Figure 2) [55].



13 **14** **15**

Figure 2. Oxadiazoles with antischistosomal activity.

3. Sesquiterpene Lactones with Activity against NTDs

3.1. Sesquiterpene Lactones with Leishmanicidal Activity

Several studies testing *in vitro* and *in vivo* activity of numerous SLs (Figure 3) have been performed in the search for new drugs against leishmaniasis. Axenic amastigotes of *L. donovani* were incubated with three anthecotulide-type linear SLs (anthecotulide (**16**), 4-hydroxyanthecotulide (**17**) and 4-acetoxyanthecotulide (**18**)) from *Anthemis auriculata* and a significant antileishmanial activity was observed; however, they are non-viable as future drugs due to the high *in vitro* cytotoxic activity values with mammalian L6 cells presented by the three molecules [56].

By evaluating antileishmanial activity against *L. donovani* looking for alternative drugs in leishmaniasis treatment, six germacranolide SLs **19–24** *of Calea zacatechichi* showed high IC_{50} values for this strain (IC_{50} as concentration that inhibits 50% of parasite growth), calealactone C (**20**) and calein D (**21**) (IC_{50} 1.9 and 2.2 μM, respectively) presented lower values than pentamidime (IC_{50} 2.9 μM), which is a drug used as positive control [57]. Inuloxins A, C, and D (**25–27**) (IC_{50} 6.9, 15.3, and 15.5 μM, respectively) also showed leishmanial activity for *L. donovani* promastigotes among 12 molecules tested, and the activity value of inuloxin A (**25**) was close to that of the positive control pentamidine (IC_{50} 4.8 μM), which is being proposed by the authors as a promising new antileishmanial lead [58].

Some SLs with antileishmanial activity for parasite forms of *L. mexicana* have been identified as promising therapies against leishmaniasis. By bioassay-guided fractionation, two SLs from *Ambrosia tenuifolia*, peruvin (**28**) and psilostachyin (**29**), besides the trypanocidal activity, showed low values of IC_{50} 1.5 μM and 0.4 μM, respectively, for the promastigote form [59]. Helenalin (**30**), together with dehydroleucodine (**31**) and mexicanin I (**32**), also have activity against L. mexicana promastigotes. For **30** and **32**, an IC_{50} value of 1.9 μM was obtained, while for **31**, it was 3.8 μM and these three drugs were more effective than ketoconazole (leishmanicidal). Low concentrations of these three SLs reduced the number of parasites per cell (Vero cells) and disturbed the growth capacity of the parasite [60]. The mechanism of these three was recently evaluated and show an increase in the generation of reactive oxygen species (ROS) by parasites and decrease the endogenous concentration of glutathione in the parasite, which has been associated with the induction of oxidative stress on promastigotes of *L. mexicana* [61].

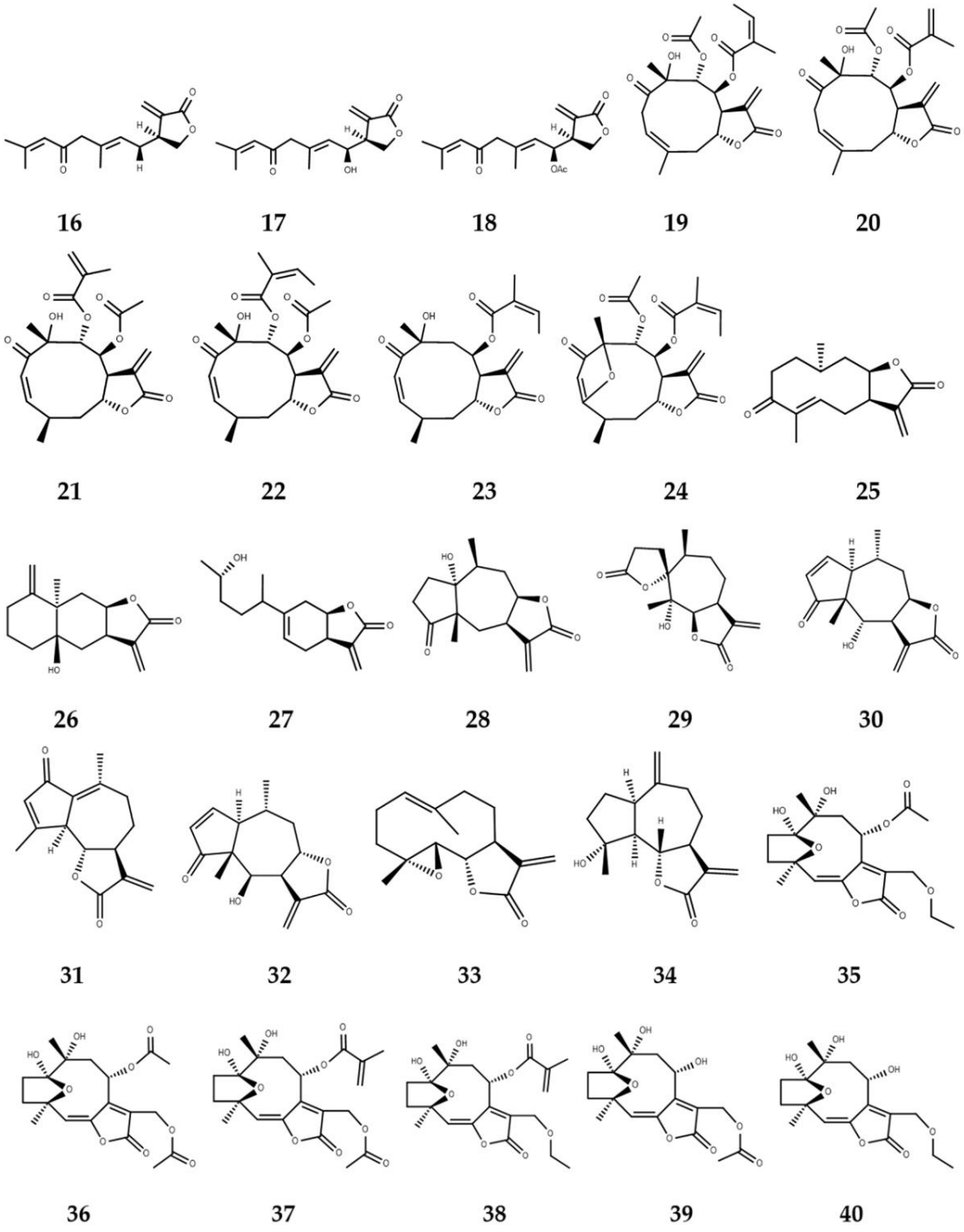


Figure 3. Sesquiterpene lactones with biological activity against NTDs

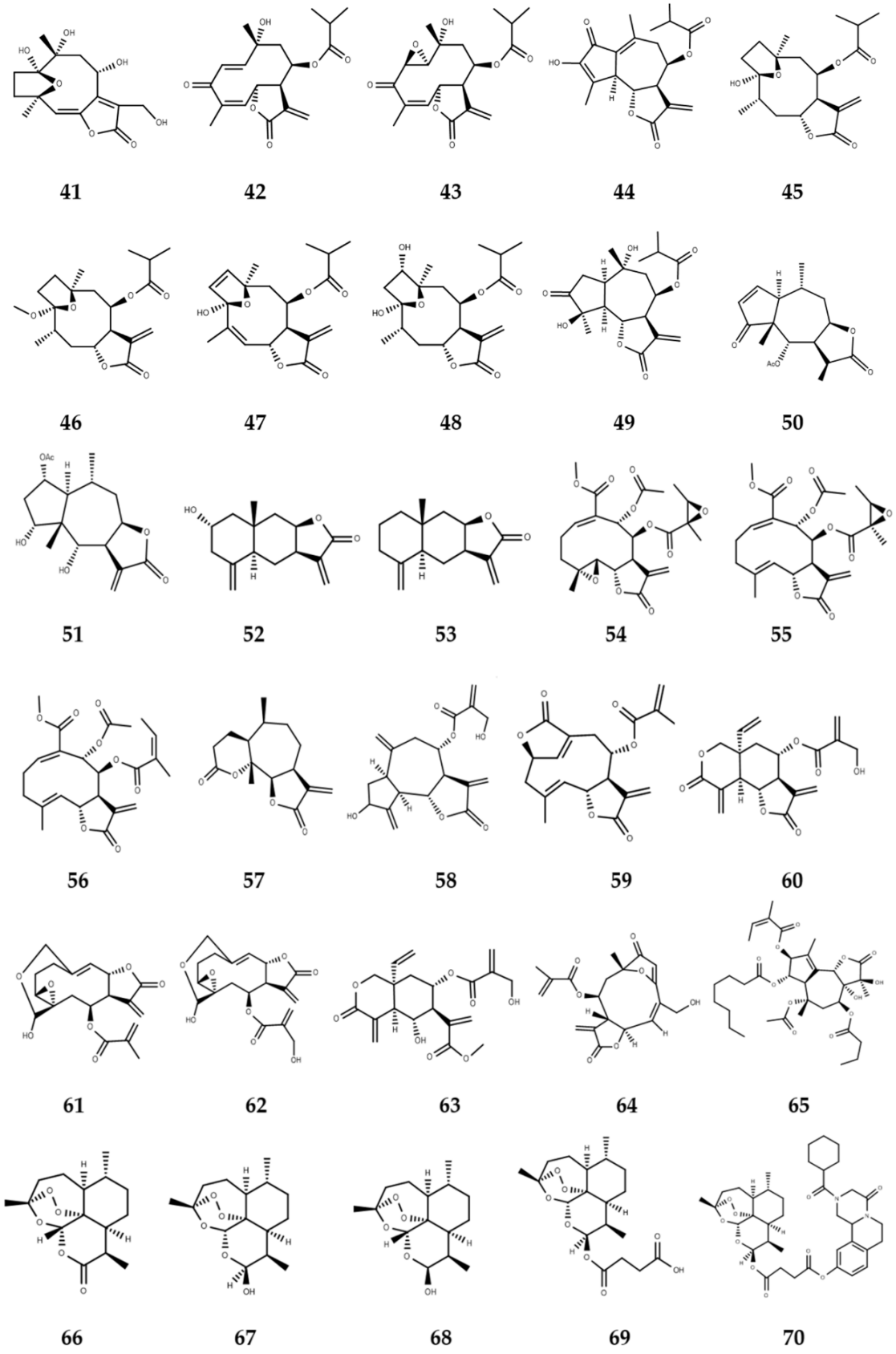


Figure 3. Sesquiterpene lactones with biological activity against NTDs.

Parthenolide (**33**) is a very common SL that is present in several species of the Asteraceae family. Tiuman et al. purified this compound from *Tanacetum parthenium*, which had previously been studied in promastigotes [62], and its leishmanial activity has been studied against axenic and intracellular amastigotes of *L. amazonensis* obtained after 72 h of incubation. IC_{50} values of 1.3 μM and 2.9 μM, respectively, were obtained, and these measurements were higher concentrations than those observed for the positive control amphotericin B, which presented an IC_{50} value of 0.22 μM. Using scanning electron microscopy (SEM), changes were observed in the parasite form, in vivo, and parthenolide concentrations of 4.0, 3.2, 2.4, and 1.6 μM reduced the proliferation of parasites into mouse macrophages and showed a survival index of 82.5%, 59.4%, 37.3%, and 6.1%, respectively [63].

In vitro and *in vivo* activity against *L. amazonensis* was tested from an SL-rich dichloromethane fraction of this species. IC_{50} values of 2.40 μg/mL and 1.76 μg/mL were obtained with promastigotes and axenic amastigotes, respectively. *In vivo*, it was observed that after four weeks of treatment with the dichloromethane fraction, a reduction in the growth and size of footpad lesions in BALB/c mice infected with promastigotes of *L. amazonensis* occurred. The main SLs present in the dichloromethane fraction are parthenolide (**33**) and 11,13-dehydrocompressanolide (**3**) and according to the differences between the plasma malondialdehyde (MDA) levels observed in the treated mice with respect to the control group, the lipid oxidation level increased; thus, these SLs cause oxidative stress by producing free radicals, which include cellular apoptosis [64].

In other studies, the activity of seven hirsutinolide-type SLs **35–41** isolated from *Pseudelephantopus spiralis* was tested with promastigotes, axenic amastigotes, and intramacrophagic amastigotes (using J774A.1 cells) of *L. infantum*. The results show that the hirsutinolides are more active against the amastigote than against the promastigote form, and compound **37** presented the lowest IC_{50} values for these two parasite forms of 9.5 μM (promastigotes) and 2.0 μM (amastigotes). A low selectivity index, however, was obtained and all SLs tested were inactive in intramacrophagic amastigotes at the evaluated concentration [65]. Antileishmanial activities of eight SLs **42–49** were tested in amastigotes and promastigotes of *L. braziliensis* as well as the cytotoxic activity in macrophages. The results show that these SLs are promising leads due to the six of seven compounds evaluated presenting activity (only **44** was not effective) while being selective against the parasite; **46–48** were the most active against promastigotes and amastigotes, and **43** had the lowest LD_{50} (the 50% lethal

dose: required dose for killing the 50% of parasites) value for promastigotes (6.0 μM) and the highest selective index (>22.7) [66].

3.2. Sesquiterpene Lactones with Trypanocidal Activity

Similarly, some SLs have shown trypanosomal activity and emerged as promising molecules for treatments against *T. cruzi* and/or *T. brucei*. For example, an *in vitro* study to evaluate the anti-*T. cruzi*, anti-*T. brucei rhodesiense*, and cytotoxic activity (L-6 cells) of six SLs (helenalin, mexicanin I, 11α,13-dihydrohelenalin acetate, chamissonolide, ivalin, and isoalantolactone) **30**, **32**, **50–53** only 5 numbers but 6 compounds was realized. Helenalin showed the lowest IC_{50} concentration against the two parasites, 0.051 μM (*T. brucei rhodesiense*) and 0.695 μM (*T. cruzi*), was related to an increase of oxidative stress inside the parasite [67]. Jimenez et al. proposed that helenalin (**30**) and dehydroleucodine (**31**) acted in *T. cruzi* induced programmed cell death of replicative epimastigotes and infective trypomastigotes [68].

New antitrypanosomal SLs have been isolated through bioassay-guided techniques from the organic extract of *Smallanthus sonchifolius*, which provided three active compounds against epimastigotes of *T. cruzi*, enhydrin (**54**) (IC_{50} 0.84 μM), uvedalin (**55**) (IC_{50} 1.09 μM), and polymatin B (**56**) (IC_{50} 4.90 μM), of which only the first two compounds presented activity against *T. cruzi* trypomastigotes 33.4 μM and 25.0 μM, respectively, being proposed as lead molecules to develop new drugs for Chagas disease [69]. By this same methodology, peruvin (**28**) and psilostachyin (**29**) were tested with *T. cruzi* epimastigotes and gave low values of IC_{50} 6.2 μM and 4.4 μM, respectively, where **29** was more active on the trypomastigotes with an IC_{50} of 2.7 μM compared with 200 μM reached by **28** [59].

The compound 11,13-dehydrocompressanolide (**34**), a guaianolide isolated from *Tanacetum parthenium*, showed activity against multiplicative epimastigotes (IC_{50} 18.1 μM), amastigotes (IC_{50} 66.6 μM), and trypomastigotes (EC_{50} 5.7 μM) of *T. cruzi* (EC_{50}: concentration that lyses 50% of the parasites), and a synergistic effect against epimastigotes in the combinational treatment with benznidazole was observed [70]. Recently, two other SLs with similar structures presented an additive effect on the activity against *T. cruzi* acting on different targets. Psilostachyin (**29**), from *Ambrosia tenuifolia*, interacts with hemin, while psilostachyin C (**57**) (*Ambrosia scabra*) inhibits the synthesis of sterols; in the two cases, the parasite death is caused by apoptosis after 4 h of treatment [71].

Otherwise, from a screening of 1800 plant and fungal extracts with *in vitro* activities against malaria, Chagas disease, and human African trypanosomiasis (two strains) parasites, cynaropicrin (**58**) was considered as a promising molecule for the development of a new antitrypanosomal drug due to the low values of IC_{50} presented for *T. brucei rhodesiense* (IC_{50} 0.3 μM) and *T. brucei gambiense* (IC_{50} of 0.2 μM) [72]. Later, the activity of this SL together with psilostachyin (**29**) against *T. cruzi* was tested *in vitro* and *in vivo*. Cynaropicrin had a similar effectivity as the control (benznidazole) against trypomastigotes *in vitro*; however, the *in vivo* testing was unable to suppress parasitemia or protect against mortality of the infected mice [73].

In another study, deoxyelephantopin (**59**) showed an IC_{50} value of 0.070 μM against *T. brucei rhodesiense* after being isolated by ethyl acetate partition from *Elephantopus scaber Linn* methanolic extract, which presented the highest activity value from 70 extracts screened from Malaysian plants [74].

3.3. Sesquiterpene Lactones with Anti-Schistosoma Activity

In the same way, some sesquiterpene lactones have been studied in the search for alternative drugs to manage schistosomiasis. Vernodalin (**60**), obtained from *Vernonia amygdalina*, showed *in vitro* activity against *S. japonicum* and inhibited movement and egg laying at 0.055 μM, while another three SLs **61–63** from the same species only presented activity at 0.5 μM. Vernodalin was lethal in infected mice with *S. japonicum* at doses up to 5 mg (oral administration), while parasites are not affected at a non-lethal dose of 2.5 mg [75].

Goyazensolide (**64**), a heliangolide extracted from *Eremanthus goyazensis*, shows antischistosomal activity *in vitro*, and reduced the motility of the worms of *S. mansoni* and caused 90% mortality in the first 24 h of exposure with 9.7–11.0 μM. Similarly, female schistosomas were significantly more susceptible than male worms under the same culture conditions and the compound showed a dramatic reduction of egg production in concentrations up to 2.2 μM [76]. In addition, thapsigargin (**65**) a SL isolated from *Thapsia garganica*, inhibits *S. mansoni* ecto-ATPdiphosphohydrolase (Ki ~= 20 μM) [77].

Artemisinin (**66**), an SL present in *Artemisia annua*, in addition to its antimalarial activity, has shown schistosomal activity. Teguments worms were altered in males and females, 30 or 45 days postinfection with *S. mansoni* in mice that were treated for 15 days with 300 or 500 mg/kg of artemisinin [78]; however, due to artemisinin's low solubility, its derivatives

(artemether, dihydroartemisinin, artesunate, and DW-3-15, **67–70**) have been studied more often and has shown positive results against *S. haematobium, S. japonicum*, and *S. mansoni* [79]; thus, these drugs are show positive results against schistosomiasis and are alternative therapies to praziquantel [80,81].

4. CADD Studies against NTDs

In addition, CADD studies have correlated biological activities of some sesquiterpene lactones against NTD-causing parasites (*Trypanosoma* and *Leishmania*) with their chemical structure and developed QSAR models using topological and three-dimensional descriptors. Similarly, there are molecular docking studies testing the interactions of these secondary metabolites (Figures 3 and 4) with target enzymes. Some of these models are purely explanatory and have helped to increase the knowledge of the SLs with respect to their action mechanism, identification of pharmacophore groups, and increase of the biological activity of these molecules. However, it is important to highlight those *in silico* studies which have found promising molecules with anti-parasitic activity from highly predictive QSAR models and its subsequent experimental validation such as the discovery of 4,15-isoatripicolide tiglate, a SL which possesses the highest *in vitro* activity against *T. brucei rhodesiense* [82].

4.1. Anti-Leishmania In Silico Studies

A series of 40 SLs with activity against *T. brucei, T. cruzi, L. donovani*, and *P. falciparum* together with their cytotoxicity activities (L6) was evaluated through use of Hologram QSAR (HQSAR) models with 16 series of fragment distinctions and fixed fragment size (four to seven atoms). The PLS models generated were evaluated through cross-validation (leave-one-out and leave-n-out) and external validation, using different latent variables. The five best HQSAR models presented values between $0.637 \leq Q^2_{LOO} \leq 0.78$ and $0.653 \leq r^2_{ext} \leq 0.944$, being the highest coefficients for *L. donovani* [83].

All models using fragment distinction were based on atoms and connections; however, only the activity against *L. donovani* and L6 cytotoxicity is clearly influenced by chirality, H-bond donor, and acceptor group parameters. Afterwards, fragment distinction was fixed and fragment size was varied in eight groups (each 1 of 3 atoms), for the five best models, from 1 to 4 atoms to 8 to 11 atoms [83].

The α, β-unsaturated groups such as 2-methylen-γ-lactone and cyclopentanone are fundamental for the biological activity, as was previously observed [84]. The presence of an

epoxide group affects the activity against *T. brucei*. Molecules with a methylcycloheptane ring present higher levels of biological activity, and these values are related to cellular permeation mechanisms. Meanwhile, six-member rings have a relationship with the toxicity by decreasing the protozoal activity at the same time while increasing the toxicity. These results show that it is possible to obtain information respect to the evaluated molecular fragments and the relationships of these with the antiprotozoal and cytotoxic activity observed [83].

A set of 123 sesquiterpenoids with leishmanicidal activity, which was subdivided according to their skeleton as sesquiterpene-coumarins, agarofurans, drimanes, pseudoguaianolides, germacranolides, guaianolides, eudesmanolides, and xanthanolides, was studied by molecular docking against four target enzymes: pteridine reductase (*PTR1*, PDB: 2QHX), cysteine synthase (*Lm*CS, PDB: 4AIR), trypanothione synthetase (*TryS*, PDB: 2VOB) of *L. major*, and N-myristoyl transferase (NMT, PDB: 2WUU) of *L. donovani* to obtain more molecular details of the binding modes of these types of metabolites [85].

Some sesquiterpene lactone skeletons presented low docking energy values with the tested enzymes. For PTR1, two xanthanolides, pungiolide A (**71**, Figure 4) and pungiolide B (**72**, Figure 4) exhibited the lowest values of docking (−10.6 kcal/mol), with even lower values than the control used (DB07765, −10.2 kcal/mol). Additionally, a similar situation was observed with respect to *Lm*CS. The docking energy value of pungiolide A was one of the lowest (−11.4 kcal/mol). In addition, the principal component analysis revealed a higher affinity of neurolenin A (**73**, Figure 4, a germacranolide) for *Lm*CS, concluding that these compounds are promising leads for further studies toward the inhibition of the studied enzymes [85]

Recently, *in vitro* leishmanicidal activity against promastigotes of *L. amazonensis* and *L. braziliensis* of 17 sesquiterpene lactones isolated from five species of the tribe Vernonieae (Asteraceae) was evaluated. An energy minimization of the whole set of molecules was performed by a semi-empirical PM3 method in MOPAC software. With the IC_{50} values, three quadratic QSAR models were obtained where descriptors related to the partition coefficient and polarizability (bpol) presented good correlation with the experimental activities r^2 = 0.82 (CLogP), 0.81 (ALogP) and 0.83 (bpol) [86]. The authors, however did not report any type of validation such as cross-validation or external test.

The compounds with the best leishmanicidal activity (IC_{50} 1.45 µM) against *L. braziliensis*, isodeoxyelephantopin (**74**, Figure 4), deoxyelephantopin (**58**, Figure 4, IC_{50} 1.34 µM) and *L. amazonensis*, centratherin (**75**, Figure 4, IC_{50} 1.60 µM) showed similar bpol values of 34.9,

32.5, and 34.8 cm^3, respectively. These results might be related to the molecules' penetration through the cellular membrane of the parasite and the receptor–ligand recognition where the van der Waals interactions are involved. Likewise, the six most active compounds against the two strains, the three above mentioned compounds together with 2-hydroxy-2,5-epoxy-8-angeloxygermacra-3Z,11(13)-dien-6,12-olide (**76**, Figure 4), 2-methoxy-2,5-epoxy-8-methacryloxygermacra-3Z,11(13)-dien-6,12-olide (**77**, Figure 4), and isocentratherin (**78**, Figure 4), showed a small range of log p values (1 to 3), suggesting that hydrophobicity-specific conditions are necessary for those sesquiterpene lactones to display better antiprotozoal effects [86].

Ogungbe et al. evaluated 28 eudesmanolide-type, 25 guaianolide-type, and 29 germacranolide-type SLs in a molecular docking study of terpenoids with several enzymes of *Leishmania*. The germacranolide-type structures exhibited the lowest docking energies with the *Leishmania* enzymes tested. 4α,5β -Epoxy-8-epi-inunolide (**79**, Figure 4) presented the lowest docking energy for *L. major* methionyl t-RNA synthetase (*Lmaj*MetRS) and *L. major* dihydroorotate dehydrogenase (*Lmaj*DHODH) being selective for these two enzymes. 16,17-Dihydrobrachycalyxolide (**80**, Figure 4) with docking values below −96.1 kJ/mol showed the lowest docking energy for *Lmaj*MetRS (−152.9 kJ/mol) and *L. mexicana* phosphor-mannomutase (*Lmex*PMM) (−136.3 kJ/mol), but it was less selective [87].

Interaction of the guaianolides and eudesmanolides was stronger with *Lmaj*MetRS and *Lmaj*DHODH. In overall terms, SLs with the lowest docking energies were observed with *Lmaj*DHODH due to the presence in the active site of a sulfur atom of the cysteine-131 and hydroxyl group of serine-69 that favor the formation of covalent bonds with the electrophilic carbon of the α-methylene lactone groups. Some guaianolides also presented very low values of docking, such as is the case of diguaiaperfolin (**81**, Figure 4) with *Lmaj*DHODH (−154.5 kJ/mol) and *Lmaj*UGPase (−151.8 kJ/mol) and 8β-[4-hydroxy-5-(5-hydroxytigloyloxy)tigloyl] santamarin (**82**, Figure 4) exhibiting selectivity to *Lmaj*MetRS (−153.1 kJ/mol) [87].

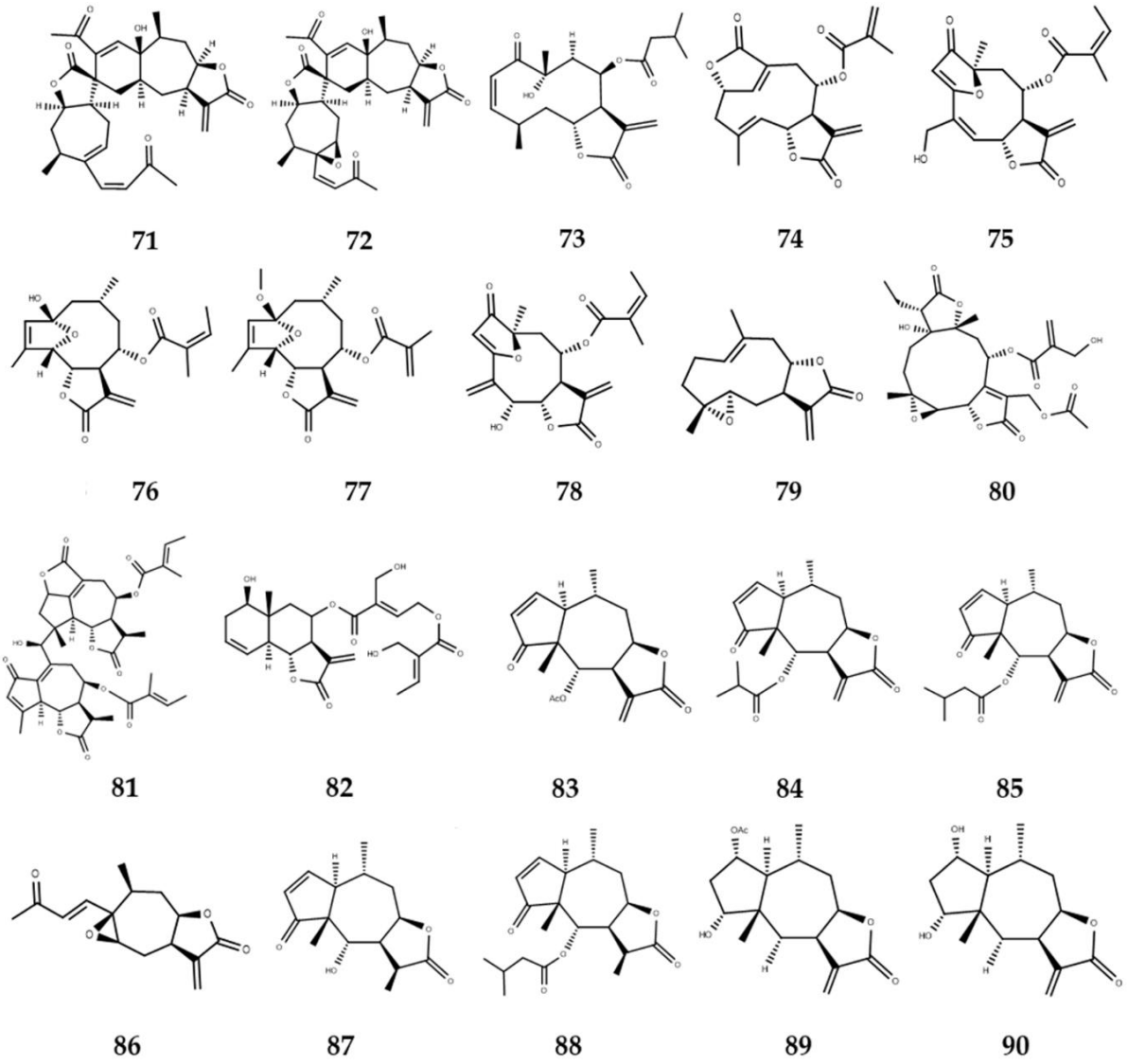


Figure 4. Sesquiterpene lactones studied by *in silico* methodologies.

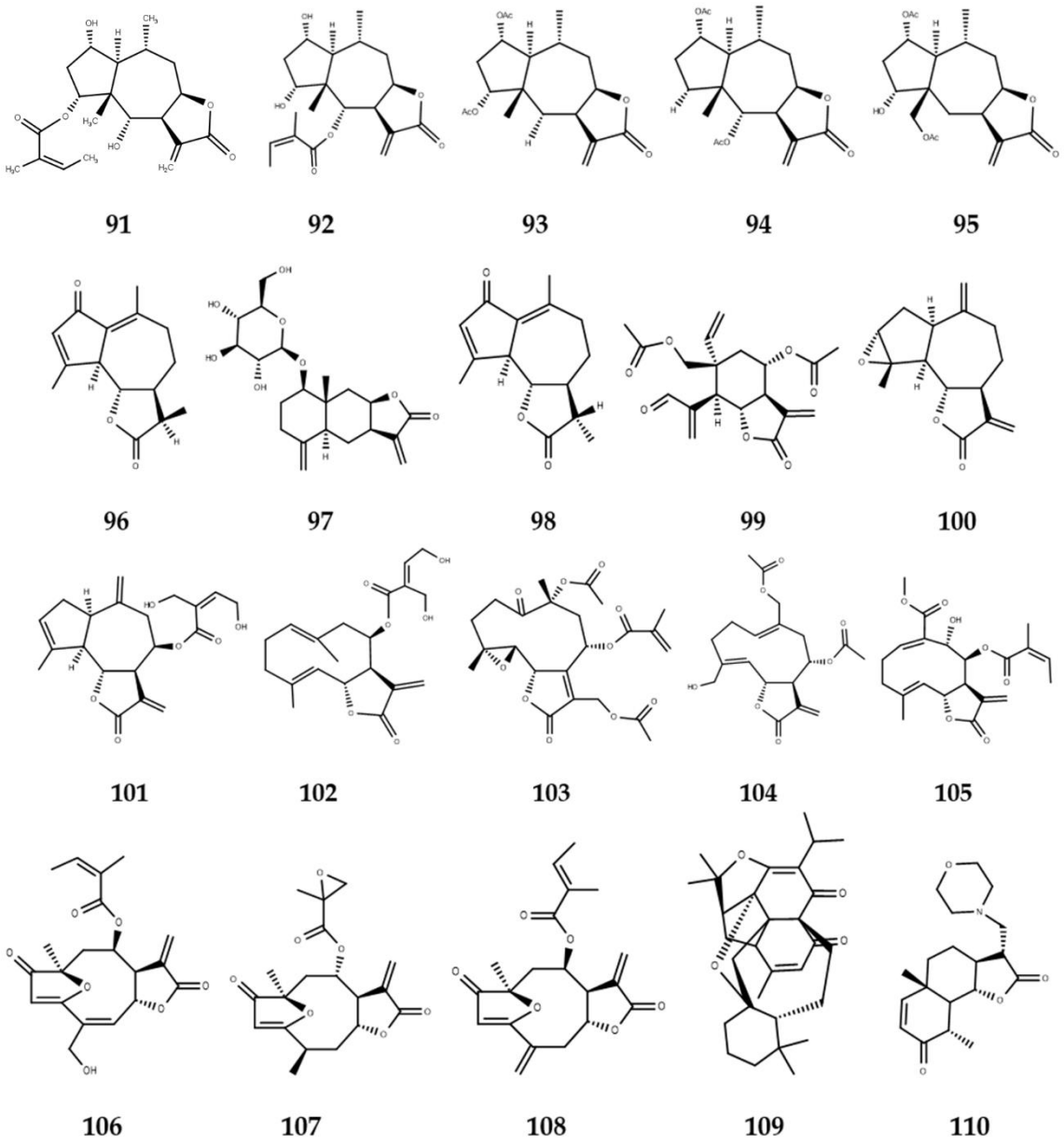


Figure 4. Sesquiterpene lactones studied by *in silico* methodologies.

4.2 Anti-T. brucei and anti-T. cruzi in Silico Studies

Schmidt et al. tested the in vitro activity of 40 sesquiterpene lactones (16 pseudoguaianolides, four xanthanolides, four modified xanthanolides, eight eudesmanolides, and eight germacranolides) against four protozoan pathogens—*T. brucei*, *T. rhodesiense*, *T. cruzi*, and *L. donovani*—as well as *Plasmodium falciparum* and the cytotoxic activity against L6 rat skeletal myoblast cells. QSAR studies were performed with those data using molecular modeling and multivariate data analysis tools to evaluate the variance in the biological activity in terms of the relative influences of various molecular descriptors. Helenalin (**30**, Figure 3)

and its derivatives **83–85** (Figure 4) show high activity against the parasites, and even some helenalin congeners have lower IC_{50} values than the used control benznidazole for *T. cruzi*. Another compound, 8-epixanthatin-1,5-epoxide (**86**, Figure 4), a xanthanolide, presents considerable activity against *T. brucei* and *L. donovani* [84].

The influence of presence or absence of α-methylene-γ-lactone and α, β-unsaturated ketone in activity against *T. brucei* and L6-cytotoxicity is studied through the variation in the coefficients of determination (r^2) obtained by multiple linear regression. The whole molecule set generated low r^2 values for *T. brucei* and L6-cytotoxicity (0.61 and 0.41, respectively), and the significance of these two descriptors was confirmed by *F*- and *t*-tests. While only considering 16 related compounds of the helenanolides **30, 32, 50, 51, 83–85**, and **87–95** series, higher r^2 values against *T. brucei* and L6-cytotoxicity (0.87 and 0.79, respectively) were achieved. The authors interpreted that in the case of compounds possessing very similar molecular structure, the modulating influence of other structural factors on bioactivity is small compared with the major impact of chemical reactivity (i.e., presence of enone and methylene lactone groups) [84,88,89].

Three QSAR models was evaluated. The AM1 semi-empirical method was initially used to optimize the molecule structure for posteriorly calculating 44 descriptors in the MOE software to generate an initial model. PLS analysis was performed using two dependent variables (L6 cytotoxicity and *T. brucei* activity). The final model contained three latent variables from 20 selected descriptors and presented a Q^2 value for *T. brucei* and L6 cells of 0.71 (for both), which is significantly higher than the model obtained using the whole set of molecules and seven latent variables from 44 descriptors that presented a Q^2 value of 0.52 and 0.41, respectively. However, PLS models only were evaluated through internal validation (cross validation, leave-one-out) without any external test [84].

Helenalin (**30**) was also studied computationally through molecular docking using two sirtuins (silent-information regulator 2 proteins), *T. cruzi* Sir2RP1, and Sir2RP3, which act as antiparasitic targets. The PLP score was used to rank and evaluate the docking poses and these were compared to select inhibitors of sirtuin classes I and III, AGK2, and thiobarbiturate 6. PLP docking scores obtained were 40.6 (Sir2RP1) and 49.2 (Sir2RP3), and intermediate values with respect to the best scores were reached by anacardic acid (8E, 11E) of 80.7 for Sir2RP1 and aculeatin D of 90.3 for Sir2RP3 [90].

A purely explanatory QSAR study was performed for 15 sesquiterpene lactones **28, 29, 33, 55, 59, 96–105** with *in vitro* activity against *T. cruzi* epimastigotes previously determined. Initially, the geometry of all molecules was pre-optimized through Molecular Mechanics Force Field (MM+) and refined by the semi-empirical AM1 method. A set of one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) descriptors was calculated through PaDEL 2.11. Additionally, Weighted Holistic Invariant Molecular (WHIM) 3D molecular descriptors used (*x, y, z*) coordinates of a molecule were used within different weighting schemes to generate a value that represents its structural features with respect to size, shape, symmetry, and atom distribution [91]. The QSAR model was evaluated by internal validation, (cross validation, leave-one-out) and determination coefficient r^2. SLs studied were divided in two groups, a training group (composed by 12 SLs) and a test group (conformed by **28, 29,** and **59**) whose selection method was not reported. pIC_{50} values predicted by the best QSAR model (r^2 = 0.99), which relates two descriptors, LAC (present or absence of α, β-unsaturated lactones) and W_{vol} (a holistic descriptor related to steric factors), are close to the experimental values reported for these 3 SLs. Cross-validation coefficient values are not reported. The authors found that a correlation of the observed antitrypanosomal activity existed with the steric parameters of the set of molecules tested and concluded that SLs have a high potential as lead molecules against Chagas disease [92].

An excellent example of the discovery of new compounds with high antiparasitic activity from *in silico* researches was found from the prediction of *in vitro* activity of 1750 SLs structures of a virtual library through a QSAR model performed with 69 molecules. Because of the high levels of activity against *T. brucei rhodesiense*, furanoheliangolide-type compounds appear as interesting leads against human African trypanosomiasis. From the three-dimensional structures of the 69 molecules, whose energies were minimized through AM1 semi-empirical method, 123 molecular descriptors were calculated. The whole set of molecules was divided into a training set (46 molecules) and test set (23 molecules), and this division was based on a structural diversity classification. A QSAR model generated, which is correlated with the presence of reactive structure elements and their electronic properties, is important to the hydrophobic volume balance, and is codified by VolSurf descriptors with a significant contribution in the model [93]. Internal validation (cross-validation, leave-one-out) presented a 0.65 Q^2 value [82].

In the prediction of activity against *T. brucei rhodesiense*, 71 molecules presented pIC_{50} values greater than or equal to 7.0, and 15 compounds corresponded to furanoheliangolide-type

SLs. Therefore, the antitrypanosomial activities of four of these were evaluated, goyazensolide (**64**, Figure 3, predicted pIC_{50}, 7.62), budlein A (**106**, Figure 4, predicted pIC_{50}, 6.60), and 4,5-dihydro-2′,3′-epoxy-15-deoxygoyazensolide (**107**, Figure 4), which presented a lower activity than the others according to the authors due to hydrogenation of the conjugated γ,δ-double bond between C-4 and C-5 (predicted pIC_{50}, 6.17) and 4,15-isoatripicolide tiglate (**108**, Figure 4, predicted pIC_{50}, 7.04), and experimental pIC_{50} values of 7.14, 7.14, 6.70, and 7.82, respectively, were obtained. The QSAR model underestimated the activities of budlein A and 4,15-isoatripicolide tiglate, while it overestimated goyazensolide [82].

From this QSAR model and subsequent validation through in vitro tests performed, furanoheliangolide (**108**, Figure 4) 4,15-isoatripicolide tiglate was found as the most potent sesquiterpene lactone against *T. brucei rhodesiense*, with a IC_{50} value of 15 nM and a selectivity index of 77 (cytotoxic assay was performed with L6 rat skeletal myoblasts), which is an interesting lead against *T. brucei rhodesiense* [82].

Finally, an interesting study was performed by Hu et al. who tested the potential adverse effects of 89 natural products and derivatives (including some SLs) with activity against *Trypanosoma* in VirtualToxLab, an *in silico* tool for estimating the toxic potential of drugs, chemicals, and natural products. The results obtained showed a wide range of potential toxicity, from 0.161 classified as benign of germacranolide (**109**, Figure 4) to 0.593, potentially harmful, of compound **110**, indicating a moderate risk to develop adverse effects, which might be due to continued exposure to some of the 89 compounds evaluated. In addition, a fraction of the tested compounds bound to nuclear receptors, favoring the development of endocrine dysregulation, while other compounds showed enzyme inhibition of the cytochrome P450 family [94].

5. Conclusions and Perspectives

In this review, we have covered the problematic aspects of the chemotherapies currently used against leishmaniasis, Chagas disease (American trypanosomiasis), sleeping sickness (human African trypanosomiasis), and schistosomiasis, and we showed how natural products have emerged as alternatives for the development of new medications against these NTDs. Specifically, sesquiterpene lactones, and secondary metabolites characteristic of the Asteraceae family have shown promising *in vitro* and *in vivo* activity, and some promising molecules have emerged from these studies.

Recently, CADD studies (mainly molecular docking and QSAR) have appeared with the aim to obtain new molecules by modify some existing molecules and understanding the mechanism of action of the SLs, which also increases the knowledge of this topic. *In vitro* tests for confirming the mechanism of action of the most active selected SLs in the docking studies here referenced were not reported. Hence, the obtained information from those studies are inconclusive regarding its enzymatic activities and demand further study.

This research area is still developing and an example of that is the case of schistosomiasis, where there have been no *in silico* studies with SLs. Therefore, it is important to continue performing CADD studies to find solutions with low cost and few side effects for the health and social problems caused by NTDs.

Acknowledgments: We would like to thank the CNPq and Capes for financial Support, and the Student Agreement Program of Graduate—PEC-PG of CNPq—Brazil. This review was performed as part of the activities of the Research Network Natural Products against Neglected Diseases (ResNet-NPND), http://resnetnpnd.org/.

Author Contributions: All authors contributed to the writing of this review

Conflicts of Interest: The authors declare no conflict of interest

References

1. World Health Organization (WHO). Investing to Overcome the Global Impact of Neglected Tropical Diseases: Third Who Report on Neglected Diseases; WHO Document Production Services: Geneva, Switzerland, 2015.
2. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the global burden of disease study 2013. Lancet 2015, 385, 117–171.
3. Feasey, N.; Wansbrough-Jones, M.; Mabey, D.C.W.; Solomon, A.W. Neglected tropical diseases. Br. Med. Bull. 2010, 93, 179–200.
4. WHO. Neglected Tropical Diseases. Available online: http://www.who.int/neglected_diseases/diseases/en/ (accessed on 10 September 2016).
5. Khare, S.; Nagle, A.S.; Biggart, A.; Lai, Y.H.; Liang, F.; Davis, L.C.; Barnes, S.W.; Mathison, C.J.; Myburgh, E.; Gao, M.Y., et al. Proteasome inhibition for treatment of leishmaniasis, chagas disease and sleeping sickness. Nature 2016, 537, 229–233.
6. World Health Organization (WHO). Research priorities for chagas disease, human african trypanosomiasis and leishmaniasis. World Health Organ. Tech. Rep. Ser. 2012, 975, 1–100.
7. De Luca, V.; Salim, V.; Atsumi, S.M.; Yu, F. Mining the biodiversity of plants: A revolution in the making. Science 2012, 336, 1658–1661.

8. Bohm, B.A.; Stuessy, T.F. Flavonoids of the Sunflower Family (Asteraceae); Springer: Vienna, Austria, 2001.

9. Tu, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat. Med. 2011, 17, 1217–1220.

10. Nobel Prize. The Nobel Prize in Physiology or Medicine 2015, Press Release: Stockholm, Sweden, 2015. Available online: https://www.nobelprize.org/nobel_prizes/medicine/laureates/2015/press.html (accessed on 11 September 2016).

11. Chadwick, M.; Trewin, H.; Gawthrop, F.; Wagstaff, C. Sesquiterpenoids lactones: Benefits to plants and people. Int. J. Mol. Sci. 2013, 14, 12780–12805.

12. Scotti, M.T.; Emerenciano, V.; Ferreira, M.J.; Scotti, L.; Stefani, R.; da Silva, M.S.; Mendonca Junior, F.J.
Self-organizing maps of molecular descriptors for sesquiterpene lactones and their application to the chemotaxonomy of the asteraceae family. Molecules 2012, 17, 4684–4702.

13. Hristozov, D.; Da Costa, F.B.; Gasteiger, J. Sesquiterpene lactones-based classification of the family asteraceae using neural networks and k-nearest neighbors. J. Chem. Inf. Model. 2007, 47, 9–19.

14. Da Costa, F.B.; Terfloth, L.; Gasteiger, J. Sesquiterpene lactone-based classification of three asteraceae tribes: A study based on self-organizing neural networks applied to chemosystematics. Phytochemistry 2005, 66, 345–353.

15. Scotti, M.T.; Fernandes, M.B.; Ferreira, M.J.; Emerenciano, V.P. Quantitative structure-activity relationship of sesquiterpene lactones with cytotoxic activity. Bioorg. Med. Chem. 2007, 15, 2927–2934.

16. Heilmann, J.; Wasescha, M.R.; Schmidt, T.J. The influence of glutathione and cysteine levels on the cytotoxicity of helenanolide type sesquiterpene lactones against KB cells. Bioorg. Med. Chem. 2001, 9, 2189–2194.

17. Woerdenbag, H.J.; Merfort, I.; Passreiter, C.M.; Schmidt, T.J.; Willuhn, G.; van Uden, W.; Pras, N.; Kampinga, H.H.; Konings, A.W. Cytotoxicity of flavonoids and sesquiterpene lactones from arnica species against the GLC4 and the COLO 320 cell lines. Planta Med. 1994, 60, 434–437.

18. Chappuis, F.; Sundar, S.; Hailu, A.; Ghalib, H.; Rijal, S.; Peeling, R.; Alvar, J.; Boelaert, M. Visceral leishmaniasis: What are the needs for diagnosis, treatment and control? Nat. Rev. Microbiol. 2007, 11, 873–882.

19. Pearson, R.D.; Sousa, A.Q. Clinical spectrum of leishmaniasis. Clin. Infect. Dis. 1996, 22, 1–13.

20. McGwire, B.S.; Satoskar, A.R. Leishmaniasis: Clinical syndromes and treatment. QJM 2014, 107, 7–14.

21. Sundar, S.; Rai, M. Laboratory diagnosis of visceral leishmaniasis. Clin. Diagn. Lab. Immunol. 2002, 9, 951–958.

22. Belo, V.S.; Struchiner, C.J.; Barbosa, D.S.; Nascimento, B.W.; Horta, M.A.; da Silva, E.S.; Werneck, G.L. Risk factors for adverse prognosis and death in American visceral leishmaniasis. PLoS Negl. Trop. Dis. 2014, 8, e2982.

23. Vasconcellos Ede, C.; Pimentel, M.I.; Schubach Ade, O.; de Oliveira Rde, V.; Azeredo-Coutinho, R.B.; Silva Fda, C.; Salgueiro Mde, M.; Moreira, J.S.; Madeira Mde, F.; Baptista, C.; et al. Intralesional meglumine antimoniate for treatment of cutaneous leishmaniasis patients with contraindication to systemic therapy from Rio De Janeiro (2000 to 2006). Am. J. Trop. Med. Hyg. 2012, 87, 257–260.

24. Yesilova, Y.; Surucu, H.A.; Ardic, N.; Aksoy, M.; Yesilova, A.; Oghumu, S.; Satoskar, A.R. Meglumine antimoniate is more effective than sodium stibogluconate in the treatment of cutaneous leishmaniasis. J. Dermatol. Treat. 2016, 27, 83–87.

25. Asilian, A.; Sadeghinia, A.; Faghihi, G.; Momeni, A.; Amini Harandi, A. The efficacy of treatment with intralesional meglumine antimoniate alone, compared with that of cryotherapy combined with the meglumine antimoniate or intralesional sodium stibogluconate, in the treatment of cutaneous leishmaniasis. Ann. Trop. Med. Parasito.l 2003, 97, 493–498.

26. Sundar, S.; Singh, A. Recent developments and future prospects in the treatment of visceral leishmaniasis. Ther. Adv. Infect. Dis. 2016, 3, 98–109.

27. Lira, R.; Sundar, S.; Makharia, A.; Kenney, R.; Gam, A.; Saraiva, E.; Sacks, D. Evidence that the high incidence of treatment failures in Indian Kala-Azar is due to the emergence of antimony-resistant strains of Leishmania donovani. J. Infect. Dis. 1999, 180, 564–567.

28. Sundar, S. Drug resistance in Indian visceral leishmaniasis. Trop. Med. Int. Health 2001, 6, 849–854.

29. Burrows, J.N.; Elliott, R.L.; Kaneko, T.; Mowbray, C.E.; Waterson, D. The role of modern drug discovery in the fight against neglected and tropical diseases. Med. Chem. Comm. 2014, 5, 688–700.

30. Croft, S.L.; Coombs, G.H. Leishmaniasis—Current chemotherapy and recent advances in the search for novel drugs. Trends Parasitol. 2003, 19, 502–508.

31. Singh, N.; Kumar, M.; Singh, R.K. Leishmaniasis: Current status of available drugs and new potential drug targets. Asian Pac. J. Trop. Med. 2012, 5, 485–497.

32. Clementi, A.; Battaglia, G.; Floris, M.; Castellino, P.; Ronco, C.; Cruz, D.N. Renal involvement in leishmaniasis: A review of the literature. NDT Plus 2011, 4, 147–152.

33. Tarleton, R.L. Chagas disease: A solvable problem, ignored. Trends Mol. Med. 2016, 22, 835–838.

34. Bermudez, J.; Davies, C.; Simonazzi, A.; Real, J.P.; Palma, S. Current drug therapy and pharmaceutical challenges for chagas disease. Acta Trop. 2016, 156, 1–16.

35. Castro, J.A.; de Mecca, M.M.; Bartel, L.C. Toxic side effects of drugs used to treat chagas' disease (American trypanosomiasis). Hum. Exp. Toxicol. 2006, 25, 471–479.

36. Teixeira, A.R.L.; Hecht, M.M.; Guimaro, M.C.; Sousa, A.O.; Nitz, N. Pathogenesis of chagas' disease: Parasite persistence and autoimmunity. Clin. Microbiol. Rev. 2011, 24, 592–630.

37. Apt, W. Current and developing therapeutic agents in the treatment of chagas disease. Drug Des. Dev. Ther. 2010, 4, 243–253.

38. Viotti, R.; Vigliano, C.; Lococo, B.; Alvarez, M.G.; Petti, M.; Bertocchi, G.; Armenti, A. Side effects of benznidazole as treatment in chronic chagas disease: Fears and realities. Expert Rev. Anti Infect. Ther. 2009, 7, 157–163.

39. Brun, R.; Blum, J.; Chappuis, F.; Burri, C. Human African trypanosomiasis. Lancet 375, 148–159.

40. Jones, A.J.; Avery, V.M. Future treatment options for human African trypanosomiasis. Expert Rev. Anti Infect. Ther. 2015, 13, 1429–1432.

41. Nok, A.J. Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. Parasitol. Res. 2003, 90, 71–79.

42. Babokhov, P.; Sanyaolu, A.O.; Oyibo, W.A.; Fagbenro-Beyioku, A.F.; Iriemenam, N.C. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. Pathog. Glob. Health 2013, 107, 242–252.

43. Raber, G.; Raber, T.; Raml, R.; Murko, M.; Magnes, C.; Francesconi, K.A. Determination of the trypanocidal drug melarsoprol and its conversion products in biological fluids with HPLC–ICPMS/ESMS. Talanta 2013, 116, 876–881.

44. Chappuis, F. Melarsoprol-free drug combinations for second-stage gambian sleeping sickness: The way to go. Clin. Infect. Dis. 2007, 45, 1443–1445.

45. Chappuis, F.; Udayraj, N.; Stietenroth, K.; Meussen, A.; Bovier, P.A. Eflornithine is safer than melarsoprol for the treatment of second-stage trypanosoma brucei gambiense human african trypanosomiasis. Clin Infect Dis 2005, 41, 748–751.

46. Colley, D.G.; Bustinduy, A.L.; Secor, W.E.; King, C.H. Human schistosomiasis. Lancet 2014, 383, 2253–2264.

47. Zoni, A.C.; Catala, L.; Ault, S.K. Schistosomiasis prevalence and intensity of infection in latin america and the caribbean countries, 1942–2014: A systematic review in the context of a regional elimination goal. PLoS Negl. Trop. Dis. 2016, 10, doi:10.1371/journal.pntd.0004493.

48. Fenwick, A.; Jourdan, P. Schistosomiasis elimination by 2020 or 2030? Int J Parasitol 2016, 46, 385–388.

49. Cioli, D.; Pica-Mattoccia, L.; Basso, A.; Guidi, A. Schistosomiasis control: Praziquantel forever? Mol. Biochem. Parasit. 2014, 195, 23–29.

50. Wang, W.; Wang, L.; Liang, Y.S. Susceptibility or resistance of praziquantel in human schistosomiasis: A review. Parasitol. Res. 2012, 111, 1871–1877.

51. Alonso, D.; Muñoz, J.; Gascón, J.; Valls, M.E.; Corachan, M. Failure of standard treatment with praziquantel in two returned travelers with schistosoma haematobium infection. Am. J. Trop. Med. Hyg. 2006, 74, 342–344.

52. Fallon, P.G.; Doenhoff, M.J. Drug-resistant schistosomiasis: Resistance to praziquantel and oxamniquine induced in schistosoma mansoni in mice is drug specific. Am. J. Trop. Med. Hyg. 1994, 51, 83–88.

53. Sayed, A.A.; Simeonov, A.; Thomas, C.J.; Inglese, J.; Austin, C.P.; Williams, D.L. Identification of oxadiazoles as new drug leads for the control of schistosomiasis. Nat. Med. 2008, 14, 407–412.

54. Rai, G.; Sayed, A.A.; Lea, W.A.; Luecke, H.F.; Chakrapani, H.; Prast-Nielsen, S.; Jadhav, A.; Leister, W.; Shen, M.; Inglese, J.; et al. Structure mechanism insights and the role of nitric oxide donation guide the development of oxadiazole-2-oxides as therapeutic agents against schistosomiasis. J. Med. Chem. 2009, 52, 6474–6483.

55. Melo-Filho, C.C.; Dantas, R.F.; Braga, R.C.; Neves, B.J.; Senger, M.R.; Valente, W.C.; Rezende-Neto, J.M.; Chaves, W.T.; Muratov, E.N.; Paveley, R.A.; et al. Qsar-driven discovery of novel chemical scaffolds active against Schistosoma Mansoni. J. Chem. Inf. Model. 2016, 56, 1357–1372.

56. Karioti, A.; Skaltsa, H.; Kaiser, M.; Tasdemir, D. Trypanocidal, leishmanicidal and cytotoxic effects of anthecotulide-type linear sesquiterpene lactones from Anthemis Auriculata. Phytomedicine 2009, 16, 783–787.

57. Wu, H.K.; Fronczek, F.R.; Burandt, C.L.; Zjawiony, J.K. Antileishmanial Germacranolides from Calea Zacatechichi. Planta Med. 2011, 77, 749–753.

58. Avolio, F.; Rimando, A.M.; Cimmino, A.; Andolfi, A.; Jain, S.; Tekwani, B.L.; Evidente, A. Inuloxins A-D and derivatives as antileishmanial agents: Structure-activity relationship study. J. Antibiot. 2014, 67, 597–601.

59. Sulsen, V.P.; Frank, F.M.; Cazorla, S.I.; Anesini, C.A.; Malchiodi, E.L.; Freixa, B.; Vila, R.; Muschietti, L.V.; Martino, V.S. Trypanocidal and leishmanicidal activities of sesquiterpene lactones from Ambrosia tenuifolia sprengel (asteraceae). Antimicrob. Agents Chemother. 2008, 52, 2415–2419.

60. Barrera, P.A.; Jimenez-Ortiz, V.; Tonn, C.; Giordano, O.; Galanti, N.; Sosa, M.A. Natural sesquiterpene lactones are active against leishmania mexicana. J. Parasitol. 2008, 94, 1143–1149.

61. Barrera, P.; Sulsen, V.P.; Lozano, E.; Rivera, M.; Beer, M.F.; Tonn, C.; Martino, V.S.; Sosa, M.A. Natural sesquiterpene lactones induce oxidative stress in leishmania mexicana. Evid. Based Complement. Altern. Med. 2013, 2013, doi:10.1155/2013/163404.

62. Tiuman, T.S.; Ueda-Nakamura, T.; Garcia Cortez, D.A.; Dias Filho, B.P.; Morgado-Diaz, J.A.; de Souza, W.; Nakamura, C.V. Antileishmanial activity of parthenolide, a sesquiterpene lactone isolated from Tanacetum parthenium. Antimicrob. Agents Chemother. 2005, 49, 176–182.

63. Tiuman, T.S.; Ueda-Nakamura, T.; Alonso, A.; Nakamura, C.V. Cell death in amastigote forms of leishmania amazonensis induced by parthenolide. BMC Microbiol. 2014, 14, doi:10.1186/1471-2180-14-152.

64. Rabito, M.F.; Britta, E.A.; Pelegrini, B.L.; Scariot, D.B.; Almeida, M.B.; Nixdorf, S.L.; Nakamura, C.V.; Ferreira, I.C. In vitro and in vivo antileishmania activity of sesquiterpene lactone-rich dichloromethane fraction obtained from Tanacetum parthenium (L.) schultz-bip. Exp. Parasitol. 2014, 143, 18–23.

65. Girardi, C.; Fabre, N.; Paloque, L.; Ramadani, A.P.; Benoit-Vical, F.; Gonzalez-Aspajo, G.; Haddad, M.; Rengifo, E.; Jullian, V. Evaluation of antiplasmodial and antileishmanial activities of herbal medicine pseudelephantopus spiralis (less.) cronquist and isolated hirsutinolide-type sesquiterpenoids. J. Ethnopharmacol. 2015, 170, 167–174..

66. de Toledo, J.S.; Ambrosio, S.R.; Borges, C.H.; Manfrim, V.; Cerri, D.G.; Cruz, A.K.; Da Costa, F.B. In vitro leishmanicidal activities of sesquiterpene lactones from tithonia diversifolia against leishmania braziliensis promastigotes and amastigotes. Molecules 2014, 19, 6070–6079.

67. Schmidt, T.J.; Brun, R.; Willuhn, G.; Khalid, S.A. Anti-trypanosomal activity of helenalin and some structurally related sesquiterpene lactones. Planta Med. 2002, 68, 750–751.

68. Jimenez, V.; Kemmerling, U.; Paredes, R.; Maya, J.D.; Sosa, M.A.; Galanti, N. Natural sesquiterpene lactones induce programmed cell death in Trypanosoma cruzi: A new therapeutic target? Phytomedicine 2014, 21, 1411–1418.

69. Frank, F.M.; Ulloa, J.; Cazorla, S.I.; Maravilla, G.; Malchiodi, E.L.; Grau, A.; Martino, V.; Catalan, C.; Muschietti, L.V. Trypanocidal activity of smallanthus sonchifolius: Identification of active sesquiterpene lactones by bioassay-guided fractionation. Evid. Based Complement. Altern. Med 2013, 2013, doi:10.1155/2013/627898.

70. Cogo, J.; Caleare, A.d.O.; Ueda-Nakamura, T.; Filho, B.P.D.; Ferreira, I.C.P.; Nakamura, C.V. Trypanocidal activity of guaianolide obtained from Tanacetum parthenium (L.) Schultz-Bip. And its combinational effect with benznidazole. Phytomedicine 2012, 20, 59–66.

71. Sulsen, V.P.; Puente, V.; Papademetrio, D.; Batlle, A.; Martino, V.S.; Frank, F.M.; Lombardo, M.E. Mode of action of the sesquiterpene lactones psilostachyin and psilostachyin C on Trypanosoma cruzi. PLoS ONE 2016, 11, e0150526.

72. Zimmermann, S.; Kaiser, M.; Brun, R.; Hamburger, M.; Adams, M. Cynaropicrin: The first plant natural product with in vivo activity against trypanosoma brucei. Planta Med. 2012, 78, 553–556.

73. da Silva, C.F.; Batista Dda, G.; De Araujo, J.S.; Batista, M.M.; Lionel, J.; de Souza, E.M.; Hammer, E.R.; da Silva, P.B.; De Mieri, M.; Adams, M.; et al. Activities of psilostachyin a and cynaropicrin against Trypanosoma cruzi in vitro and in vivo. Antimicrob. Agents Chemother. 2013, 57, 5307–5314.

74. Zahari, Z.; Jani, N.A.; Amanah, A.; Latif, M.N.; Majid, M.I.; Adenan, M.I. Bioassay-guided isolation of a sesquiterpene lactone of deoxyelephantopin from elephantopus scaber linn. Active on trypanosome brucei rhodesience. Phytomedicine 2014, 21, 282–285.

75. Jisaka, M.; Kawanaka, M.; Sugiyama, H.; Takegawa, K.; Huffman, M.A.; Ohigashi, H.; Koshimizu, K. Antischistosomal activities of sesquiterpene lactones and steroid glucosides from vernonia amygdalina, possibly used by wild chimpanzees against parasite-related diseases. Biosci. Biotechnol. Biochem. 1992, 56, 845–846.

76. Barth, L.R.; Fernandes, A.P.M.; RibeiroPaes, J.T.; Rodrigues, V. Effects of goyazensolide during in vitro cultivation of schistosoma mansoni. Mem. Inst. Oswaldo Cruz 1997, 92, 427–429.

77. Martins, S.M.; Torres, C.R.; Ferreira, S.T. Inhibition of the ecto-ATPdiphosphohydrolase of schistosoma mansoni by thapsigargin. Biosci. Rep. 2000, 20, 369–381.

78. De Oliveira, C.N.F.; de Oliveira, R.N.; Frezza, T.F.; Rehder, V.L.G.; Allegretti, S.M. Tegument of schistosoma mansoni as a therapeutic target. In Parasitic Diseases—Schistosomiasis; InTech - Open Access Publisher:, 2013.

79. Neves, B.; Andrade, C.; Cravo, P. Natural products as leads in schistosome drug discovery. Molecules 2015, 20, doi:10.3390/molecules20021872.

80. Madbouly, N.A.; Shalash, I.R.; El Deeb, S.O.; El Amir, A.M. Effect of artemether on cytokine profile and egg induced pathology in murine schistosomiasis Mansoni. J. Adv. Res. 2015, 6, 851–857.

81. Dong, L.L.; Duan, W.W.; Chen, J.L.; Sun, H.; Qiao, C.H.; Xia, C.M. An artemisinin derivative of praziquantel as an orally active antischistosomal agent. PLoS ONE 2014, 9, e112163.

82. Schmidt, T.J.; Da Costa, F.B.; Lopes, N.P.; Kaiser, M.; Brun, R. In silico prediction and experimental evaluation of furanoheliangolide sesquiterpene lactones as potent agents against trypanosoma brucei rhodesiense. Antimicrob. Agents Chemother. 2014, 58, 325–332.

83. Trossini, G.H.; Maltarollo, V.G.; Schmidt, T.J. Hologram qsar studies of antiprotozoal activities of sesquiterpene lactones. Molecules 2014, 19, 10546–10562.

84. Schmidt, T.J.; Nour, A.M.; Khalid, S.A.; Kaiser, M.; Brun, R. Quantitative structure—Antiprotozoal activity relationships of sesquiterpene lactones. Molecules 2009, 14, 2062–2076.

85. Bernal, F.A.; Coy-Barrera, E. In-silico analyses of sesquiterpene-related compounds on selected leishmania enzyme-based targets. Molecules 2014, 19, 5550–5569.

86. Sosa, A.M.; Amaya, S.; Salamanca Capusiri, E.; Gilabert, M.; Bardon, A.; Gimenez, A.; Vera, N.R.; Borkosky, S.A. Active sesquiterpene lactones against leishmania amazonensis and leishmania Braziliensis. Nat. Prod. Res. 2016, 1–5.

87. Ogungbe, I.V.; Setzer, W.N. In-silico leishmania target selectivity of antiparasitic terpenoids. Molecules 2013, 18, 7761–7847.

88. Schmidt, T.J.; Heilmann, J. Quantitative structure-cytotoxicity relationships of sesquiterpene lactones derived from partial charge (q)-based fractional accessible surface area descriptors (q_frasas). Quant. Struct.-Act. Relationsh. 2002, 21, 276–287.

89. Schmidt, T. Quantitative structure-cytotoxicity relationships within a series of helenanolide type sesquiterpene lactones. Pharm. Pharmacol. Lett. 1999, 9, 9–13.

90. Sacconnay, L.; Angleviel, M.; Randazzo, G.M.; Marçal Ferreira Queiroz, M.; Ferreira Queiroz, E.; Wolfender, J.-L.; Carrupt, P.-A.; Nurisso, A. Computational studies on sirtuins from Trypanosoma cruzi: Structures, conformations and interactions with phytochemicals. PLoS Negl. Trop. Dis. 2014, 8, e2689.

91. Todeschini, R.; Gramatica, P.; Provenzani, R.; Marengo, E. Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polyaromatic hydrocarbons. Chemom. Intell. Lab. 1995, 27, 221–229.

92. Fabian, L.; Sulsen, V.; Frank, F.; Cazorla, S.; Malchiodi, E.; Martino, V.; Lizarraga, E.; Catalan, C.; Moglioni, A.; Muschietti, L.; et al. In silico study of structural and geometrical requirements of natural sesquiterpene lactones with trypanocidal activity. Mini Rev. Med. Chem. 2013, 13, 1407–1414.

93. Cruciani, G.; Crivori, P.; Carrupt, P.A.; Testa, B. Molecular fields in quantitative structure–permeation relationships: The volsurf approach. J. Mol. Struct. 2000, 503, 17–30.

94. Hu, Z.; Wahl, J.; Hamburger, M.; Vedani, A. Molecular mechanisms of endocrine and metabolic disruption: An in silico study on antitrypanosomal natural products and some derivatives. Toxicol. Lett. 2016, 252, 29–41.

CAPÍTULO III

Recentemente o número de novas estruturas assim como de estudos relacionados a atividade biológica destas têm aumentando de maneira exponencial, sendo fundamental a classificação e organização das informações. A partir disso, as bases de dados têm fornecido uma coleta sistemática de informações sobre produtos naturais e seus derivados, incluindo estrutura, fonte e mecanismos de ação, provindo uma ajuda significativa na descoberta moderna de drogas (XIE, 2015).

Na área da quimioinformática, as bases de dados de produtos naturais desempenham um papel preponderante, permitindo uma pesquisa rápida e o desenvolvimento de métodos computacionais para extrair ou prever rapidamente informações úteis para cada molécula, incluindo suas propriedades físicas, químicas e biológicas (CHEN, 2005), isto representa uma diminuição no tempo dos experimentos assim como nos custos econômicos respeito ao observado no trabalho tradicional em produtos naturais.

Atualmente podem ser encontradas múltiplas bases de dados de produtos naturais com informações associadas para as estruturas o que facilita as buscas diretas, fornecendo propriedades mais específicas e úteis para a área de pesquisa de desenvolvimento de medicamentos, NAPRALERT (http://www.napralert.org), Dictionary of Natural Products—DNP (http://dnp.chemnetbase.com) ou $NuBBE_{DT}$ o qual incorpora várias classes de metabolitos secundários e derivados da biodiversidade do Brasil (VALLI, 2013), são alguns exemplos.

No entanto, o problema não deve limitar-se ao referido com a quantidade de informações contidas num banco de dados. Outros aspectos como a facilidade da busca, consistência entre diversidade dos sistemas operacionais atualmente disponíveis (Microsoft Windows, Mac e Linux) e uma interface agradável e de uso fácil são igualmente importantes.

Por isso, neste capitulo é apresentado a parte da implementação assim como o gerenciamento dos dados, de uma moderna e inovadora base de dados de produtos naturais denominada, SistematX (http://sistematx.ufpb.br), desenvolvida no Laboratório de Quimioinformática de Universidade Federal da Paraíba, a qual permite a consulta via internet assim como o gerenciamento a partir de qualquer computador credenciado.

REFERÊNCIAS

CHEN, J.; et al. ChemDB: a public database of small molecules and related chemoinformatics resources. **Bioinformatics**, v. 21, n. 22, p. 4133-4139, 2005.

VALLI, M.; et al. Development of a natural products database from the biodiversity of Brazil. **Journal of natural products**, v. 76, n. 3, p. 439-444, 2013.

XIE, T.; et al. Review of natural product databases. **Cell proliferation**, v. 48, n. 4, p. 398-404, 2015.

SistematX, an online web-based cheminformatics tool for data management of secondary metabolites

Marcus Tullius Scotti[1]*, Chonny Herrera-Acevedo[1], Tiago Oliveira Branquinho[2,3], Renan Paiva Oliveira Costa[1], Silas Yudi Konno de Oliveira Santos[1], Ricardo Pereira Rodrigues[1], Luciana Scotti[1], Fernando Batista Da-Costa[3]

[1]Postgraduate Program in Natural Products and Synthetic Bioactive, Federal University of Paraíba, IPeFarM, Campus I, Cidade Universitária, 58051-900, João Pessoa - PB, Brazil; mtscotti@gmail.com (M.T.S.), chonny622@gmail.com (C.H.A.), renan0paiva@hotmail.com (R.P.O.C.), syudik12@gmail.com (S.Y.K.O.S) ricardo.p.rodrigues@ufes.br (R.P.R.), Luciana.scotti@gmail.com (L.S.)

[2]Department of Pharmacy, Federal University of Sergipe (UFS-SE), Av. Marechal Rondon, s/n, Jd. Rosa Elze, 49100-000 São Cristóvão, SE, Brazil; tiagobranquinho.ufs@gmail.com (T.B.O)

[3]AsterBioChem research team, University of São Paulo (USP), School of Pharmaceutical Sciences of Ribeirão Preto, Laboratory of Pharmacognosy, Av. do Café, s/n 14040-903 Ribeirão Preto – SP, Brazil; febcosta@fcfrp.usp.br (F.B.D.C)

*Correspondence: mtscotti@gmail.com; Tel.: +55-83-99869-0415

Artigo submetido na revista ***Molecules***.

Fator de Impacto: 2.861

Abstract: The traditional work of a natural products researcher consists in large part of time-consuming experimental work, collecting biota to prepare and analyze extracts and to identify innovative metabolites. However, along this long scientific path, much information is lost or restricted to a specific niche. The large amounts of data already produced and the science of metabolomics reveal new questions: Are these compounds known or new? How fast can this information be obtained? To answer these and other relevant questions, an appropriate procedure to correctly store information on the data retrieved from the discovered metabolites is necessary. The SistematX homepage is http://sistematx.ufpb.br, and the interface is implemented considering the following aspects: a) the ability to search by structure, SMILES (Simplified Molecular Input Line Entry) code, compound name and species; b) the ability save chemical structures found by searching; c) compound data results include important

characteristics for natural products chemistry; and d) the user can find specific information for taxonomic rank (from family to species) of the plant from which the compound was isolated, the searched-for molecule, and the bibliographic reference and Global Positioning System (GPS) coordinates. The SistematX homepage allows the user to log into the data management area using a login name and password and gain access to administration pages. In this article, we introduced a modern and innovative web interface for the management of a secondary metabolite database. With its multiplatform design, it is able to be properly consulted via the internet and managed from any accredited computer. The interface provided by SistematX contains a wealth of useful information for the scientific community about natural products, highlighting the locations of species from which compounds are isolated.

Keywords: SistematX, secondary metabolites, data management, online web-based tool

1. Introduction

The traditional work of a natural products researcher can be summarized as collection of biological samples, preparation of extracts for biological screening or bioassay-guided fractionation, and isolation and purification of (bioactive or not) compounds. However, the first question that may arise is the following: are these compounds known or new? In addition, metabolomics studies have introduced a new question: how fast can this information be obtained? [1]

The stage of dereplication, a process known as the rapid characterization of previously known compounds in mixtures without their prior purification, has become a strategically important area for natural products research involved in screening programs in several commercial and non-commercial databases [2-4]. These databases can be searched with minimal information, such as structural chemical and biological data from compounds; however, dereplication now requires additional information, such as biogeographical and taxonomic information, or the presence of a certain compound (new or known) in other individuals of the same species, genus, subfamily, and family. This information can also help to reduce the number of hits during chemical identification by dereplication.

Large structure-based data collections, such as ChemSpider, Pubchem, ChemBl, and ZINC [5–8] can be used for this purpose [9, 10]. However, searching for chemical structures in these

databases is costly and often generates false targets among compounds of natural and synthetic origin.

For this reason, a number of specialized natural products databases were developed that are commercially or freely available and only contain restricted information, for example, the Dictionary of Natural Products (DNP) [11], NAPRALERT [12], Marinlit for marine natural products [13], and Antibase for microorganisms and higher fungi materials. Nevertheless, none of these provide structural collections in a format that can be rapidly integrated into software such as ACD/Structure Elucidator and others [9].

Other natural products databases provide natural products extracted from various resources and contain various associated information such as toxicity prediction, but so far, little or nothing is known about these resources, for example, SUPER NATURAL II [14]. Natural products databases exhibit a huge range of structural complexity and thus are expected to contribute to the ability of such databases to provide positive hits [2, 15]. These structures are available in regional databases, for example, NUBBEdb [16], SANCDB [17], TM-CM [18], TCM-Database@Taiwan [19], NANPDB [20] and TCMID [21]. Many have been used in virtual screening research studies. In addition to the database information described above that uses two-dimensional (2D) structures, several databases have selected methods and tools for generating three-dimensional (3D) structures of small organic molecules, often for use in structure-based drug design.

In addition, databases of natural products with a focus on metabolomic studies with relationships between species-metabolites include the KNApSAcK Family [22], TIPdb-3D [23], and AsterDB [24], which enable searches for chemical structures by plant species names and other taxonomic information. Nevertheless, some data are still lacking for the purpose of exact dereplication. Information such as exact mass and geographic data can be very important for this type of study [25–27].

It is not enough to simply focus on the information contained in a database. A clean and user-friendly interface, fast search, and consistency between currently available operational systems (Microsoft Windows, Mac, and Linux) can be just as important. For this purpose, the SistematX software was developed to provide the abovementioned information for chemosystematics studies, dereplication, and botanical correlations.

2. Results and discussion

2.1. Utility and discussion

The SistematX homepage is shown in Figure 1A. After the user enters the website (http://sistematx.ufpb.br), the “Structure search” option is seen with the MarvinJS API at the top of the screen. Another three search options can be exhibited in the interface. The initial screen of the system also shows the SMILES code (Figure 1B), compound name (Figure 1C) and plant species search modes (Figure 1D).

In the first option, the user can perform the search using the drawn full structure or molecular skeleton, fragments or substructures, which is important in cases when the user only knows a structural characteristic of the structure such as functional groups or when studies require structural similarity, structural groups or families of compounds. In addition, it is possible to search by SMILES code, a chemical notation system capable of representing even the most complex organic compounds using a simple grammar; by common (usual) name or IUPAC name (or part of one of these); and by species, although in this option, it is necessary to first insert the name of the genus (which presents an autocompletion option). After being selected, the system presents all species available for the user to select for the search

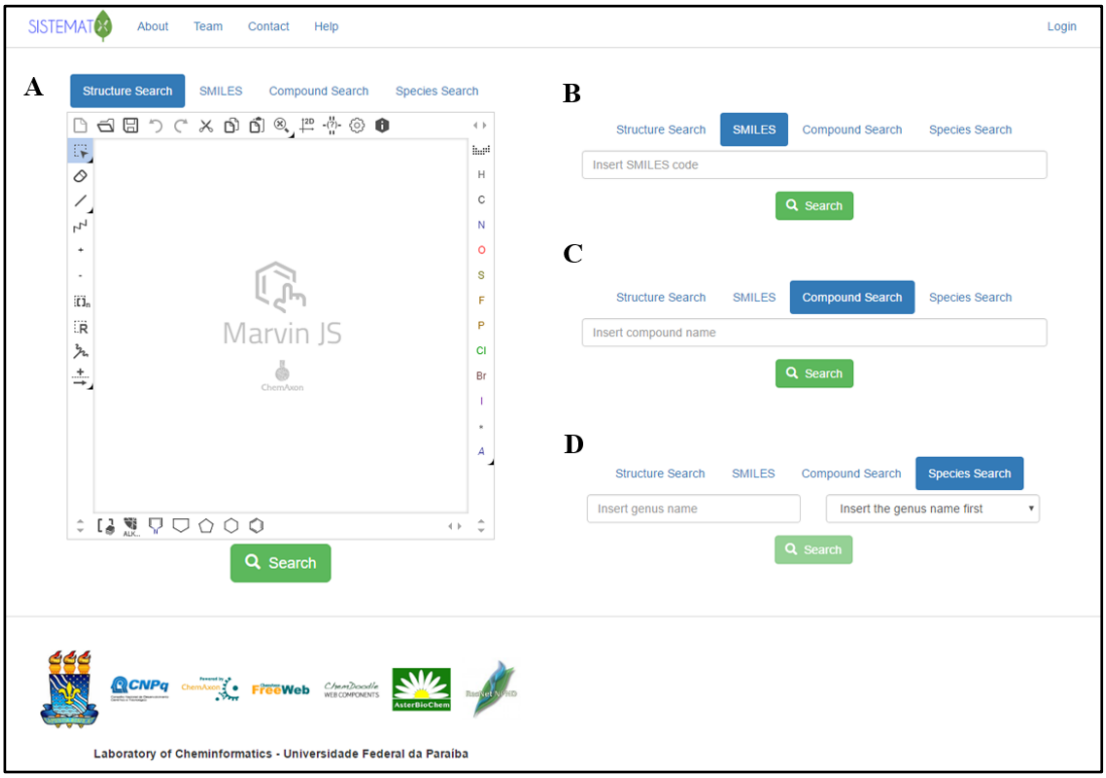


Figure 1. SistematX homepage with different search options: A) by structure, B) by SMILES, C) by compound name and D) by plant species.

When performing a search, the mechanism generates a search results page (six results per page), using common names; if the compound does not have one, it shows the IUPAC name (Figure 2). When a result is selected, the user has access to the data for that molecule, which are classified into six different groups (Figure 3).

The first group of results that appears is related to the structural representation of the searched molecule. The 2D structure is observed in the interface; on top, this appears as the option to amplify. After this is clicked, the system displays the visualization of the molecule in 2D and 3D (ChemDoodle) and an additional option for saving the 2D or 3D structure in an MDL (Molecular Design Limited) Molfile. The second type of result exhibited by the systems associated with compound identification, such as common name, SMILES code, IUPAC name, InChI code, InChIKey code and CAS number. Except for common name, which is optional and registered by the administrator, all parameters are provided by JChem API

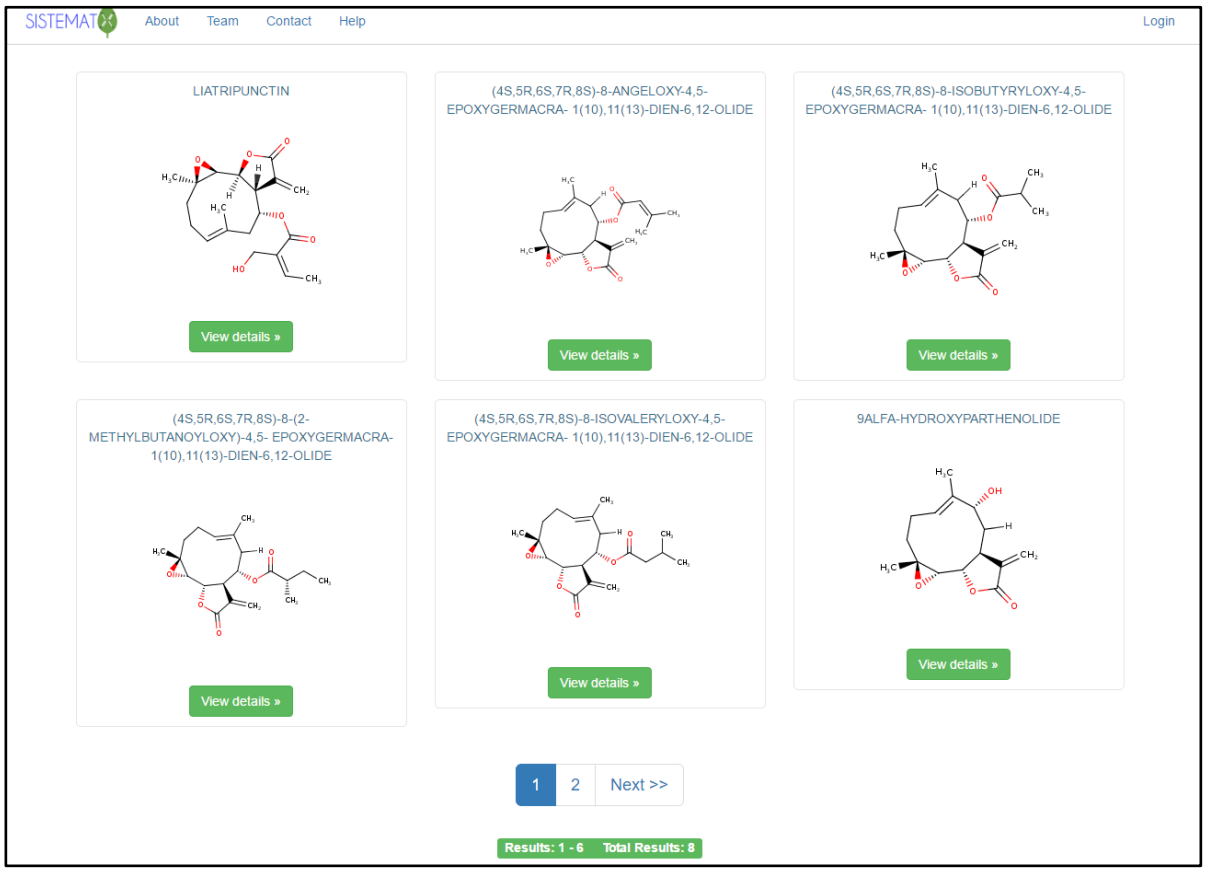


Figure 2. SistematX results page.

Compound data results include important characteristics for natural products chemistry. The class of secondary metabolite of the searched molecule and its skeleton provides information

about its biosynthetic pathway and assists in chemosystematics and chemotaxonomic studies. Oxidation number (NOX), which is calculated based on Hendrickson rules [28], has been fundamental in chemotaxonomy since Gottlieb related the oxidation grade of molecules to species evolution [29]. Molecular mass is calculated using the most abundant isotope of each element (exact mass) and the average atomic mass of each element (relative mass); these data are important for users working on purification processes and for structural elucidation of molecules, due to the mass information, which is essential for determining the purity of secondary metabolites.

In the botanical data field, the user can find specific information such as the taxonomic rank (from family to species) of the plant from which the compound was isolated, the searched molecule, and the bibliographic reference, which includes journal name, volume, page and year. Because many different species can biosynthesize the same molecule, there is one register per species. Meanwhile, the biological data exhibit results obtained in studies related to the biological activity of the searched molecule; the type of activity, system, units, activity value and bibliographic reference are available in this section.

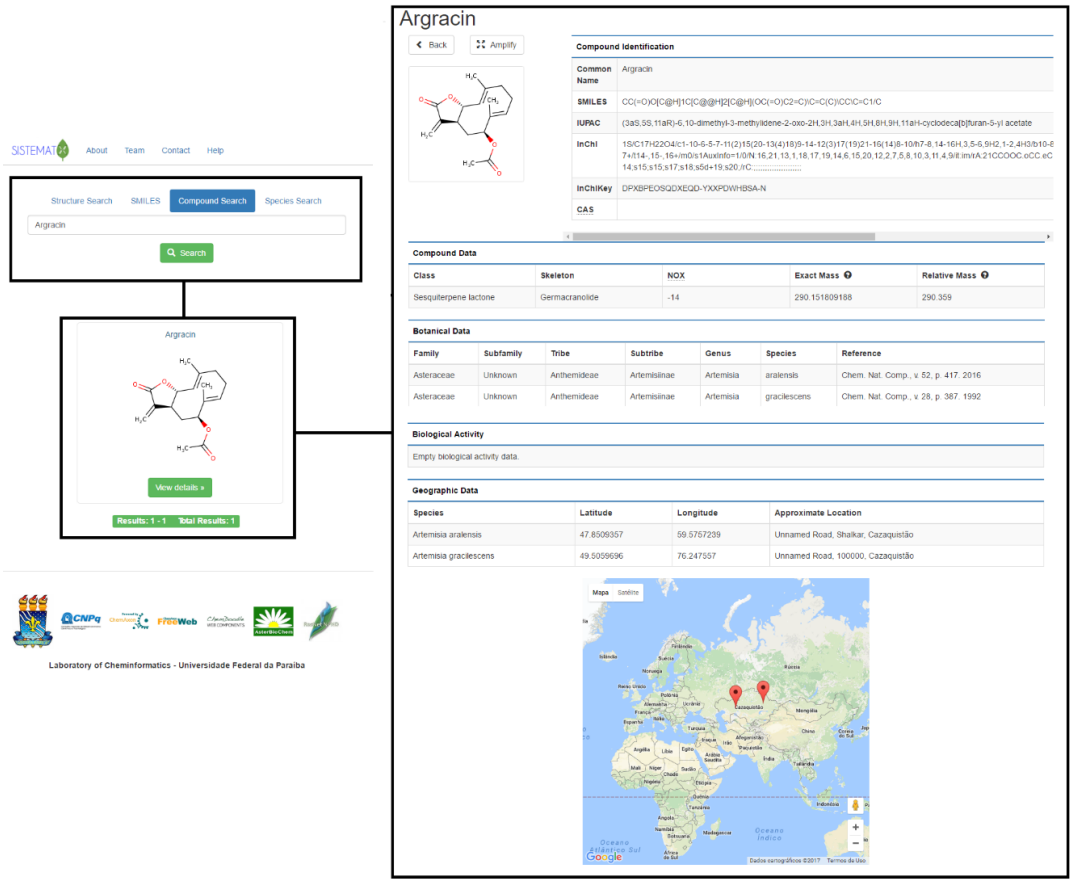


Figure 3. SistematX screen for molecular data.

Plant species have revealed clear genetic signals for local adaptation [30]. One species can synthesize a secondary metabolite depending on its location, and there are observed variations in compound concentrations at different sites. Because geographical data is an important parameter in natural products research, SistematX shows geographical coordinates (latitude and longitude) for a searched molecule and an approximate location of the species from which this metabolite was isolated. Using the Google maps API, the user can observe the species location on the world map.

2.2. Data management

On the SistematX homepage, the user can also log into the data management area using login name and password (Figure 4A) and from there access the administration pages to or edit register new molecules. Once the correspondent information has been accepted, the data management interface appears.

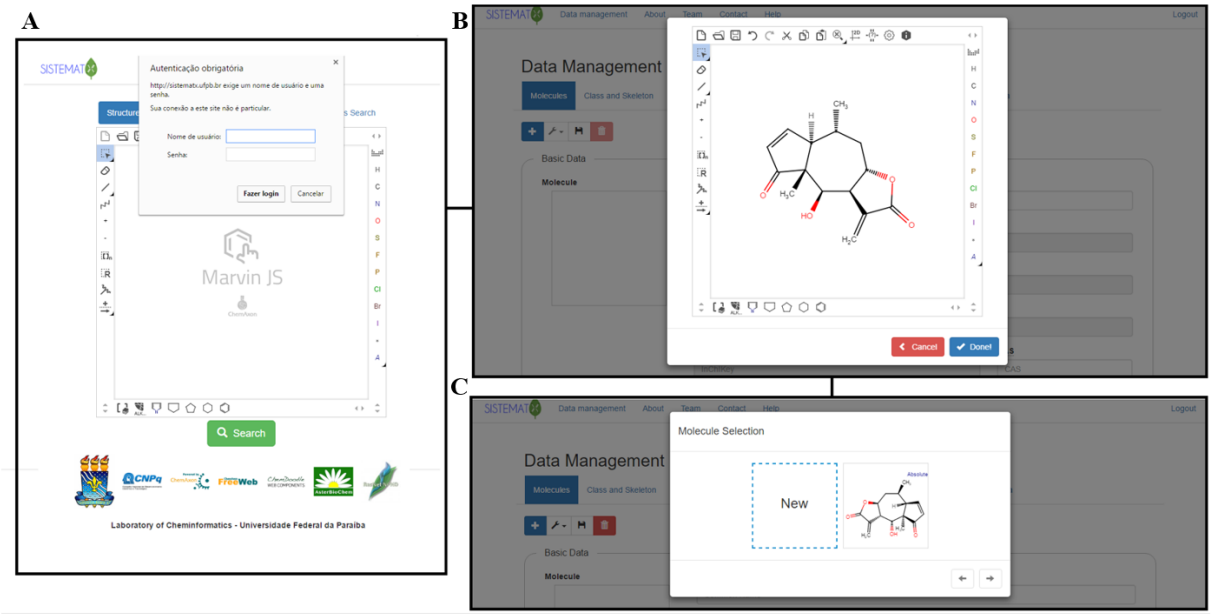


Figure 4. SistematX creates new registers through an administrator: A) login and password option, B) structure view and C) molecule selection.

The first requirement to register a new molecule is to insert the structure in the MarvinJS API (Figure 4B). Several methods can be used to accomplish this step: drawing, copying SMILES code, or importing the molecule in a compatible format (e.g., sdf, cdx, mol, mol2). The molecule selection option then appears (Figure 4C), and if the molecule is new to the system, it appears with the option New. If this option is selected, a blank register page with four

subdivisions is shown: basic data, extra data, botanical data, and geographical data. However, if the molecule already exists in the system, another box with the drawn molecule appears, and after choosing this box, the register page for the structure containing all previously registered information appears (Figure 5).

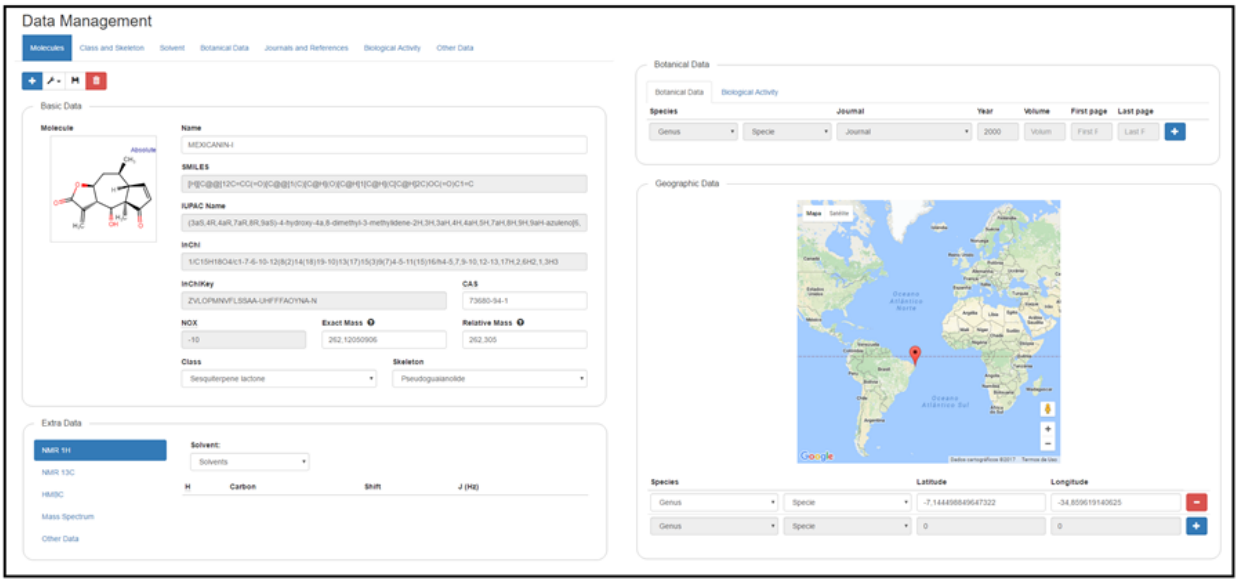


Figure 5. SistematX data management interface.

Immediately after the New option is selected, it appears in the register page, including some basic data generated by MarvinJS API: SMILES, IUPAC name, InChI, InChIKey, NOX, exact mass and relative mass. The class and skeleton of metabolites must be chosen to register a new structure in the system; this is a sequential process, and thus, class must first be selected to make the skeleton structure option available. Classes and skeletons not already registered in the system can be registered via the Class and Skeleton tab on the top of the screen. Common name and CAS are optional registration options.

The Extra Data subdivision allows insertion of all structural spectroscopic information, such as ^{1}H and ^{13}C NMR and mass spectra. In addition, it is possible to find 2D NMR information through the HMBC technique to establish the relationship between ^{13}C and ^{1}H shifts. In the NMR data, the administrator must first select the deuterated solvent used in the spectroscopic studies. If this information does not exist in the options, it can be registered by selecting the Solvent tab on the top of the screen. After the structure appears with an atomic numeration assigned by MarvinJS API (identified as Atom), it is always necessary to verify these numbers for the biogenetic numeration (identified as “Biogenetic”) and to finally add the chemical shift

value for each atom. For ^{1}H NMR, it is also possible to register H-H coupling (J in Hz). Mass spectrum information, molecular mass and intensity of fragments can also be being registered.

To create a new registry for botanical data of a certain species, the following information is required: journal, year, volume, first page and last page information. Journal, genus and species are drop-down lists, and these last two must be filled in this order. If any information respective to these three fields does not exist, it can be inserted by clicking in Journal and References and writing the journal name (autocomplete tab) or by clicking in Botanical Data, where taxonomic data are registered. Biological activity data appear as activity, system, system type, value, journal, year and pages. Drop-down lists can be filled with previously registered data or by entering new data in the Biological Activity tab.

Finally, the administrator can register geographical data using the Google maps API. For any structure, it is possible to register latitude and longitude of the corresponding species studied. The genus and species boxes are filled in the same manner described above. Longitude and latitude can be inserted in two ways, first by writing the coordinates in the spaces; once registered, they appear on the world map as a red indicator showing the location. Another method is to select the place by clicking on the red pin on the map; when the pin is released, the values of longitude and latitude appear in the respective boxes.

3. Materials and methods

3.1. *Implementation*

SistematX was developed in Java programming language version 8 or higher, using JSP (JavaServer Pages) technology version 2.1 or higher and MySQL database version 5.5.46-0 for Linux [31] to maintain the system data. An initial version of SistematX web was published in the proceedings of MOL2NET in 2015 to demonstrate its functionalities to the academic community and, with feedback, to improve old functionalities and add new tools [32].

The system uses JSP to create pages with specific information for each molecule and dynamic page change by clicking on certain buttons. Intermediary pages are used to recover information from the database and insert it in the JSP. The system creates a DAO (Data Access Object) to organize the data on intermediary pages, working like a bridge from the DAO to JSP. For each database table needed in a request, a DAO is created specifically for that table containing all attributes from the database table.

Bootstrap version 3.3.5 [33], a graphical interface web framework that uses HTML (HyperText Markup Language), CSS (Cascading Style Sheets) and JavaScript, is employed for a better appearance on HTML pages, adapting some functions and styles to aid in the website design. The system also uses jQuery framework version 2.1.4 [34], a powerful JavaScript tool to manipulate HTML DOM (Document Object Model) events. Another framework, jQuery AutoComplete version 1.2.18 [35], is used to generate autocomplete inputs.

Several API are used in SistematX implementation (Table 1). MarvinJS version 15.7.20, from ChemAxon [36], is the drawing API (Application Programming Interface) that is integrated with ChemAxon JChem WebService [37], an external online service that transforms the drawn structure into SMILES (Simplified Molecular-Input Line-Entry System) code, after which a JChem API function turns it to a binary fingerprint. This fingerprint is used to search for molecules in the database via their structure. The molecule converted to the fingerprint is used as a fragment in the search, comparing it to the database molecule's fingerprints to determine if it exists inside as a fragment. This API utilizes HTML, CSS and JavaScript to perform the transformations.

MarvinJS is used to create a 3D structure from a 2D molecule, and ChemDoodle Web Components [38] shows this view as JavaScript in the browser. This API is able to generate several 2D and 3D molecule graphical views using pure JavaScript.

ChemAxon API allows the visualization of compound characteristics. SistematX displays general nomenclature information such as the common name, SMILES code, IUPAC name, InChI, InChIKey, CAS registry number and properties such as oxidation number (NOX), exact mass and relative mass.

Table 1. Summary of the API implemented by SistematX

API	Description	Engine
	1. Structure	
2D drawing	Allows drawing and visualization of chemical structures	ChemAxon
3D generator	Uses 2D drawing to generate a 3D representation of the molecule	ChemAxon
3D	Graphical visualization of 3D molecules with JavaScript	ChemDoodle
	2. Compound Identification	
SMILES	Simplified Molecular Input Line Entry System	ChemAxon
IUPAC	IUPAC Nomenclature	ChemAxon
InChI	IUPAC International Chemical Identifier	ChemAxon
InChIKey	InChIKey is a compact format of InChI code	ChemAxon
CAS	Chemical Abstracts Service Registry Number	ChemAxon
	3. Compound Data	
NOX	Oxidation number (NOX) of an organic compound	ChemAxon
Exact Mass	Uses the mass of the most abundant isotope of each element	ChemAxon
Relative Mass	Uses the average atomic mass of each element	ChemAxon
	4. Geographic data	
Latitude	Can be inserted by the administrator or appears by clicking in the world	Google, Inc.
Longitude	Can be inserted by the administrator or appears by clicking in the world	Google, Inc.
Approximate	Using the latitude and longitude, appears as an approximate location of the	Google, Inc.
Visualization	Uses the world map to visualize the localization of the species	Google, Inc.

Additionally, Google Maps, from Google Inc. [39], an API used to prepare maps and locations, is used in the system to show the registered metabolite location on the world map. The API draws the map and receives locations from the database, which are two variables representing a latitude and a longitude. A registered molecule may have multiple locations and a species linked to it. Using a JavaScript function, it graphically sets the locations on the map. When registering by clicking on the map, it sets a marker at the mouse location and adds a line to the coordinates list below the map for each marker on the map, allowing it to automatically change the position when the value in the latitude or longitude boxes is changed. The coordinates are also transformed into an approximate address, using reverse geocode, a function from the Google Maps API.

4. Conclusions

In this article, we introduce a web interface for managing a secondary metabolite database, which is multiplatform and able to be consulted via the internet and managed from any accredited computer. The interface provides a wealth of useful information for the scientific community about natural products, highlighting the location of species from which the compounds were isolated.

SistematX is freely accessible on the homepage http://sistematx.ufpb.br.

JSP: Java Server Pages, DAO: Data Access Object, API: Application Programming Interface HTML: HyperText Markup Language, CSS: Cascading Style Sheets, DOM: Document Object Model, SMILES: Simplified Molecular Input Line Entry Specification, NOX: Oxidation number, NMR: Nuclear Magnetic Resonance, HMBC: Heteronuclear Multiple Bond Correlation, IUPAC: International Union of Pure and Applied Chemistry, 3D: Three-dimensional, 2D: Two-dimensional, InChI: IUPAC International Chemical Identifier. CAS: Chemical Abstracts Service.

Acknowledgments: We would like to thank the Student Agreement Program of Graduate—PEC-PG of CNPq, Brazil and the Brazilian National Council for Scientific and Technological Development (CNPq) – Award Number - 461093/2014-6.

Author Contributions: S.Y.K.O.S., R.P.O.C. and M.T.S. carried out the design, programing and drafted the manuscript. C.H.A, T.B.O., F.B.D.C., L.S. and R.P.R. carried out the design, tests and drafted the manuscript. C.H.A, L.S., F.B.D.C. and M.T.S. revised the manuscript critically.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blunt, J.W.; Munro, M.H.G. Is There an Ideal Database for Natural Products Research? In: Natural Products: Discourse, Diversity, and Design. Eds.; Osbourn A., Goss R. J. and Carter G. T.; John Wiley & Sons, Inc., Hoboken, NJ, USA; 2014. pp. 413–431. DOI: 10.1002/9781118794623.ch22
2. Harvey, A.L.; Edrada-Ebel, R.; Quinn, RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat. Rev. Drug. Discov. 2015, 14, 111-129 DOI:10.1038/nrd4510
3. Corley, D.G.; Durley, R.C. Strategies for Database Dereplication of Natural Products. J. Nat. Prod. 1994, 57, 1484–1490. DOI:10.1021/np50113a002
4. Oliveira, T.; Chagas-Paula, D.; Rosa, A.; Gobbo-Neto, L.; Schmidt, T.J., Da Costa F.B. Temporal characteristics of a natural products in-house database. Planta Med. 2013, 79, 1113–1114. DOI: 10.1055/s-0033-1351852
5. Pence, H.; Williams, A. ChemSpider: an online chemical information resource. J. Chem. Educ. 2010, 87, 1123–1124. DOI: 10.1186/1758-2946-2-3.
6. Bolton, E.E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. Annu. Reports. Comput. Chem. 2008, 4, 217–241. DOI: 10.1016/S1574-1400(08)00012-1

7. Degtyarenko, K.; de Matos, P.; Ennis, M., Hastings, J.; Zbinden, M.; McNaught, A.; Alcántara, R.; Darsow, M.; Guedj, M.; Ashburner, M. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008, 36, D344–D350. DOI: 10.1093/nar/gkm791

8. Irwin, J.J.; Shoichet, B.K. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Info Model, 2005, 45, 177-182. DOI: 10.1021/ci049714+

9. Williams, R.B.; O'Neil-Johnson, M.; Williams, A.J.; Wheeler, P.; Pol, R.; Moser, A. Dereplication of natural products using minimal NMR data inputs. Org. Biomol. Chem. 2015, 13, 9957–9962. DOI: 10.1039/C5OB01713K

10. Eugster, P.J.; Boccard, J.; Debrus, B.; Bréant, L.; Wolfender, J.L.; Martel, S.; Carrupt, P.A. Retention time prediction for dereplication of natural products (CxHyOz) in LC–MS metabolite profiling. Phytochemistry. 2014, 108, 196–207. DOI: 10.1016/j.phytochem.2014.10.005

11. DNP Database, Dictionary of Natural Products. CRC Press. Available online: http://dnp.chemnetbase.com/. (Accessed on 4 Mar 2017).

12. Graham, J. G.; Farnsworth, N. R. The NAPRALERT database as an aid for discovery of novel bioactive compounds. In: Comprehensive Natural Products II: Chemistry and Biology, Elsevier Ltd, 2010, Volume 3, pp. 81-94; ISBN: 9780080453828

13. Dabb, S.; Blunt, J.; Munro, M. MarinLit: Database and essential tools for the marine natural products community. Abstr. Pap. Am. Chem. Soc. 248th National Meeting of the American-Chemical-Society (ACS), San Francisco, CA, USA, 10-14 Aug 2014.

14. Banerjee, P.; Erehman, J.; Gohlke, B-O.; Wilheim, T.; Preissner, R.; Dunkel, M. Super Natural II--a database of natural products. Nucleic Acids Res. 2015, 43, D935–D939. DOI: 10.1093/nar/gku886

15. Drewry, D.H.; Macarron, R. Enhancements of screening collections to address areas of unmet medical need: an industry perspective. Curr. Opin. Chem. Biol. 2010, 14, 289–298. DOI: 10.1016/j.cbpa.2010.03.024

16. Valli. M.; dos Santos, R.N.; Figueira, L.D.; Nakajima, C.H.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Development of a natural products database from the biodiversity of Brazil. J. Nat. Prod. 2013, 76, 439–444.

17. Hatherley, R.; Brown, D.K.; Musyoka, T.M.; Penkler, D.L.; Faya, N.; Lobb, K.A.; Tastan Bishop, Ö. SANCDB: a South African natural compound database. J. Cheminform. 2015, 7, 29. DOI: 10.1186/s13321-015-0080-8

18. Kim, S.K.; Nam, S.; Jang, H.; Kim, A.; Lee, J.J. TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. BMC Complement. Altern. Med. 2015, 15, 218. DOI: 10.1186/s12906-015-0758-5

19. Chen, CY-C. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One, 2011, 6, e15939. DOI: 10.1371/journal.pone.0015939

20. Ntie-Kang, F.; Telukunta, K.K.; Döring, K.; Simoben, C.V.; A Moumbock, A.F.; Malange, Y.I.; Njume, L.E.; Yong, J.N.; Sippl, W.; Günther, S. NANPDB: A Resource for Natural Products from Northern African Sources. J. Nat. Prod. 2017, 80: 2067-2076. DOI: 10.1021/acs.jnatprod.7b00283

21. Xue, R.; Fang, Z.; Zhang, M.; Yi, Z.; Wen, C.; Shi, T. TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res. 2013, 41, D1089–D1095. DOI: 10.1093/nar/gks1100

22. Afendi, F.M.; Okada, T.; Yamazaki, M.; Hirai-Morita, A.; Nakamura, Y.; Nakamura, K.; Ikeda, S.; Takahashi, H.; Altaf-Ul-Amin, M.; Darusman, L.K.; Saito, K.; Kanaya, S. KNApSAcK Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research. Plant Cell Physiol, 2012, 53, e1. DOI: 10.1093/pcp/pcr165

23. Tung, C.W.; Lin, Y.C.; Chang, H.S.; Wang, C.C.; Chen, I.S.; Jheng, J.L.; Li, J.H. TIPdb-3D: the three-dimensional structure database of phytochemicals from Taiwan indigenous plants. Database-Oxford. 2014, bau055. DOI:10.1093/database/bau055.

24. AsterDB, AsterBioChem in house database. Available online: http://www.asterbiochem.org/asterdb. (Accessed on 4 Mar 2017)

25. Sampaio, B.L.; Edrada-Ebel, R.; Da Costa, F.B. Effect of the environment on the secondary metabolic profile of Tithonia diversifolia: a model for environmental metabolomics of plants. Sci. Rep. 2016, 6, 29265. DOI: 10.1038/srep29265

26. Schmidt, T.J.; Rzeppa, S.; Kaiser, M.; Brun, R. Larrea tridentata—Absolute configuration of its epoxylignans and investigations on its antiprotozoal activity. Phytochem. Lett. 2012, 5, 632–638. DOI: 10.1016/j.phytol.2012.06.011

27. Gaquerel, E.; Kuhl, C.; Neumann, S. Computational annotation of plant metabolomics profiles via a novel network-assisted approach. Metabolomics, 2013, 9, 904–918. DOI: 10.1007/s11306-013-0504-2

28. Hendrickson, J.B.; Cram, D.J.; Hammond, G.S. Organic chemistry, 3rd ed.; McGraw-Hill: New York, USA 1970; ISBN: 07-028150-5

29. Gottlieb, O. The role of oxygen in phytochemical evolution towards diversity. Phytochemistry, 1989, 28, 2545-2558. DOI: 10.1016/S0031-9422(00)98039-7

30. Züst, T.; Heichinger, C.; Grossniklaus, U.; Harrington, R.; Kliebenstein, D.J.; Turnbull, L.A. Natural enemies drive geographic variation in plant defenses. Science, 2012, 338, 116-119. DOI: 10.1126/science.1226397.

31. MySQL. Available online: http://dev.mysql.com/downloads/mysql/ (Accessed on 31 Jan 2017).

32. Scotti, M.; Oliveira Da Silva Junior, R.; Yudi, S.; Brayner, R.; Scotti, L. SISTEMAT X - A web tool to manage databases of secondary metabolites. In Proceedings of the MOL2NET, 5–15 December 2015; Sciforum Electronic Conference Series, Vol. 1, 2015, F013; DOI:10.3390/MOL2NET-1-F013.

33. Bootstrap. Available online: http://getbootstrap.com/ (Accessed on 10 Jan 2017).

34. jQuery. Available online: http://jquery.com/ (Accessed on 17 Jan 2017).

35. jQuery autocomplete. Available online: http://jqueryui.com/autocomplete/ (Accessed on 20 Jan 2017).

36. ChemAxon, Marvin JS. Available online: http://chemaxon.com/products/marvin/marvin-js/ (Accessed on 24 Jan 2017).

37. ChemAxon. Jchem web services. Available online: http://chemaxon.com/products/jchem-web-services/ (Accessed on 24 Jan 2017).

38. ChemDoodle by iChemLabs. Available online: http://web.chemdoodle.com/ (Accessed on 24 Jan 2017).

39. Google maps by Google, Inc. Available online: http://developers.google.com/maps/ (Accessed on 13 Jan 2017).

CAPÍTULO IV

A Doença de Chagas é uma das principais doenças tropicais negligenciadas, foi descoberta em 1909 por Carlos Chagas, afeta cerca de 8 milhões de pessoas, principalmente nas regiões de maior pobreza de América central e América do Sul, causando mais de 10.000 mortes por ano (GBD, 2013; TARLETON, 2016) A infecção é causada pelo protozoário flagelado *Trypanosoma cruzi,* o qual é transmitido aos humanos mediante vetores conhecidos como "barbeiros", os quais são insetos triatomíneos, comumente as espécies *Triatoma infestans, Triatoma dimidiata* e *Triatoma brasiliensis* (BARGUES, 2008)

Dois medicamentos são atualmente usados como quimioterapia contra a doença de Chagas: Nifurtimox e Benznidazol (APT, 2010). Sua eficácia está próxima ao 100%, porém, o tratamento pode durar até dois meses, gerando náuseas, dores, vomito, entre outros (CASTRO, 2006; VIOTTI, 2009; APT, 2010), além de infecções pós-tratamento (principalmente quando existe coinfecção HIV-*T. Cruzi*) (TEIXEIRA, 2011) já foram reportados casos de resistência de algumas cepas de *T. cruzi* a estes compostos (BERMUDEZ, 2016).

Sendo necessária a busca de novos tratamentos para o controle e eliminação desta doença, diversas investigações realizadas têm apresentado sesquiterpenos lactonizados com atividade *in vitro* contra as diferentes formas parasitárias de *T. cruzi.* Schmidt e colaboradores (SCHMIDT, 2002) determinaram que helenalina apresenta o menor valor de IC_{50}, dentro de uma series de 6 sesquiterpenos lactonizados testados, sendo posteriormente determinado que esta estrutura induz a morte celular programa dos epimastigotas replicativas assim como de tripomastigotas infetivas (JIMENEZ, 2014).

Igualmente um estudo com Yacón (*Smallanthus sonchifolius*), uma planta nativa dos Andes de América do Sul, permitiu determinar a atividade *in vitro* de enidrina, uvedalina e polimatina B contra epimastigotas e tripomastigotas de *T. cruzi.* Os resultados obtidos mostraram valores de IC_{50} menores a 5uM para epimastigotas, entanto para a forma parasitaria tripomastigota foi observado que somente enidrina e uvedalina apresentam atividade, sendo propostas estas duas como promissórias estruturas antichagásicas (FRANK, 2013). Usando *Ambrosia tenuifolia,* um estudo similar foi realizado, determinando-se que Psilostatina C possui atividade contra tripomastigotas e epimastigotas de *T. cruzi* com valores de IC_{50} inferiores a 5 µM, a sua vez, peruvina somente apresentou atividade contra epimastigotas (SÜLSEN, 2008).

Com estes antecedentes, os sesquiterpenos lactonizados das espécies da família Asteraceae emergem como uma interessante alternativa na busca de moléculas com atividade antichagásica. Neste capítulo, é apresentado um estudo usando ferramentas quimioinformáticas

no qual foi realizada uma aproximação combinada de duas metodologias de triagem virtual (baseada na estrutura do ligante e do receptor) que foram desenvolvidas para um banco de 1.306 sesquiterpenos lactonizados da família Asteraceae, os quais estão disponíveis no seguinte endereço: http://sistematx.ufpb.br. O objetivo, é propor estruturas com potencial atividade antichagásica, estabelecendo o possível mecanismo de ação destas.

Para a triagem virtual baseada na estrutura do ligante, foram desenvolvidos modelos *random forest* para cada uma das três formas parasitárias de *T. cruzi* (amastigota, tripomastigota e epimastigota) sendo classificados como ativas, 34, 17 e 420 moléculas, respetivamente. Por sua vez, a triagem virtual baseada na estrutura do receptor (*docking* molecular), foi feita usando oito estruturas cristalinas de proteínas alvo de *T. cruzi*, sendo classificadas como ativas aquelas estruturas com valores de probabilidade acima de 0,5 e energia menor ao ligante reportado no PDB. Finalmente, mediante análises de consenso das metodologias usadas, tendo em conta a taxa de falsos positivos dos modelos, foram propostos os sesquiterpenos lactonizados com promissória atividade antichagásica.

REFERENCIAS

APT, W. Current and developing therapeutic agents in the treatment of Chagas disease. **Drug design, development and therapy**, v. 4, p. 243-253, 2010.

BARGUES, M. D.; et al. Phylogeography and genetic variation of Triatoma dimidiata, the main Chagas disease vector in Central America, and its position within the genus Triatoma. **PLoS neglected tropical diseases**, v. 2, n. 5, p. e233, 2008.

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

CASTRO, J. A.; DEMECCA, M. M.; BARTEL, L. C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). **Human & experimental toxicology**, v. 25, n. 8, p. 471-479, 2006.

FRANK, F.M.; ULLOA, J.; CAZORLA, S.I.; MARAVILLA, G.; MALCHIODI, E.L.; GRAU, A.; MARTINO, V.; CATALAN, C.; MUSCHIETTI, L.V. Trypanocidal activity of Smallanthus sonchifolius: Identification of active sesquiterpene lactones by bioassay-guided fractionation. **Evidence-based complementary and alternative medicine** v. 2013, 2013

GBD 2013. Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the global burden of disease study 2013. **Lancet,** v. 385, p.117-171, 2015.

JIMENEZ, V.; KEMMERLING, U.; PAREDES, R.; MAYA, J.D.; SOSA, M.A.; GALANTI, N. Natural sesquiterpene lactones induce programmed cell death in Trypanosoma cruzi: A new therapeutic target? **Phytomedicine**, v. 21, p. 1411–1418, 2014.

SCHMIDT, T.J.; et al. Anti-trypanosomal activity of helenalin and some structurally related sesquiterpene lactones**. Planta medica**, v. 68, n. 8, p. 750-751, 2002.

SÜLSEN, V.P.; FRANK, F.M.; CAZORLA, S.I.; ANESINI, C.A.; MALCHIODI, E.L.; FREIXA, B.; VILA, R.; MUSCHIETTI, L.V.; MARTINO, V.S. Trypanocidal and

leishmanicidal activities of sesquiterpene lactones from Ambrosia tenuifolia sprengel (asteraceae). **Antimicrobial Agents and Chemotherapy**, v. 52, n. 7, p. 2415-2419, 2008

TARLETON, R.L. Chagas disease: A solvable problem, ignored. **Trends in molecular medicine**, v. 22, n. 10, p. 835-838, 2016

TEIXEIRA, A.R.L.; et al. Pathogenesis of chagas' disease: parasite persistence and autoimmunity. **Clinical microbiology reviews**, v. 24, n. 3, p. 592-630, 2011.

VIOTTI, R.; et al. Side effects of benznidazole as treatment in chronic Chagas disease: fears and realities. **Expert review of anti-infective therapy**, v. 7, n. 2, p. 157-163, 2009.

In silico studies designed to select sesquiterpene lactones with potential antichagasic activity from an in-house Asteraceae database

Chonny Herrera Acevedo, [a] Luciana Scotti [a] and Marcus Tullius Scotti*[a]

[a] C. Herrera-Acevedo, Dra. L. Scotti, Dr. M.T. Scotti
Post-Graduate Program in Natural and Synthetic Bioactive Products
Federal University of Paraíba
Cidade Universitária- Castelo Branco III, João Pessoa, PB, Brazil
E-mail: mtscotti@gmail.com

Artigo submetido na revista ***ChemMedChem***.

Fator de Impacto: 3.225

Abstract: Chagas disease is an endemic disease caused by *Trypanosoma cruzi*, which affects more than eight million people, mostly in the Americas. A search for new treatments is necessary to control and eliminate this disease. Sesquiterpene lactones (SLs) are an interesting group of secondary metabolites characteristic of Asteraceae that have presented a wide range of biological activities. From the ChEMBL database, we selected a diverse set of 4,452, 1,635 and 1,322 structures with tested activity against the three *T. cruzi* parasitic forms, amastigote, trypomastigote and epimastigote, respectively, to create random forest (RF) models with an accuracy of greater than 74% for cross-validation and test sets. Afterwards, a ligand-based virtual screen of the entire SLs of Asteraceae database stored in SistematX (1,306 structures) was performed. In addition, a structure-based virtual screen was also performed for the same set of SLs using molecular docking. Finally, using an approach combining ligand-based and structure-based virtual screening along with the equations proposed in this study to normalize the probability scores, we verified potentially active compounds and established a possible mechanism of action.

Keywords: Asteraceae, Chagas disease, Ligand-based virtual screening, Structure-based virtual screening, Sesquiterpene lactones, Machine learning.

1. Introduction

Chagas disease (American trypanosomiasis) is a neglected tropical disease (NTD) that mainly affects populations living in poverty and low sanitary conditions. This disease causes more than 10,000 deaths annually and is the parasitic disease with the greatest impact on the Americas. This disease is an anthropozoonosis caused by *Trypanosoma cruzi*, a flagellated protozoan, which is transmitted in endemic areas by several species of triatomine insects, such as *Triatoma*, *Panstrongylus* and *Rhodnius*[1]. The life cycle of *T. cruzi* alternates between noninfective and infective forms that involve both human and insect hosts, beginning with the entry of trypomastigotes (infective—nonreplicative form) into human hosts who have been bitten by a triatomine insect, and the subsequent differentiation of the parasites into amastigotes (noninfective—replicative form) within mammalian cells. After the amastigotes replicate, they transform into trypomastigotes with capacity to infect new cells, starting a new replicative cycle, or are ingested along with the blood meal by a triatomine insect, enabling differentiation into a third parasitic form, epimastigotes (noninfective—replicative form), whose replicative stage occurs in the gut of the insect vector[1a, 2]

Chemotherapies currently used against *T. cruzi* are two nitro compounds: benznidazole and nifurtimox. Although these drugs are greater than 80% effective against the acute phase of the disease, they present several problems in the chronic phase, where their effectiveness is limited [3]. Undesirable side-effects are observed in approximately 30% of the patients, including nausea, vomiting, weight loss, dermatitis, fever, muscular pain, and cardiomyopathy, among many others, further prolonging the duration of the treatment (30–60 days)[4]; additionally, the risk of reactivation in immunosuppressed patients [1a, 5], creates a complex scenario. Therefore, novel antichagasic drugs that allow the control and elimination of this disease must be developed.

In an attempt to discover new compounds with antitrypanosomal activity, several enzymes of *T. cruzi* have been studied as targets: trans-sialidase (E.C.3.2.1.18), sterol 14α-demethylase (CYP51, E.C. 1.14.13.70), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH, E.C. 1.2.1.12) [6]. However, as the major cysteine protease of *T. cruzi*, Cruzain (E.C. 3.4.22.51) is the main target of the research aimed at the development of antichagasic drugs, as this enzyme is expressed in all stages of the parasite life cycle and is essential for the invasion of host cells, evasion of the immune system, nutrition, and reproduction, among other functions [7]. Currently, diverse inhibitors of this protease with different scaffolds and catalytic mechanisms display substantial activity against the parasite *in vitro* and *in vivo*[8].

On the other hand, natural products have been an invaluable source of inspiration for the development of therapeutic agents. In recent years, their use has increased again after several decades in which it had diminished due to technical barriers. Despite these limitations, natural products were used as the starting material and were a major source of diverse chemicals during the 20th century in studies searching for new drugs. As many of the limitations described above have already been overcome by recent technical advances, strategies for screening natural products have recently re-emerged and play a critical role in the therapeutic area. More than 50% of new small-molecule drugs approved by the FDA from 1981 to 2014 (n = 1,211) were estimated to be directly or indirectly obtained from these metabolites[9].

Because of the exponential increase in new structures and related activity studies, the classification and organization of this information in databases has become essential. This systematization has allowed the development of new techniques in silico, such as virtual screening, which has helped to optimize the use of compounds, mainly where very small amounts of these compounds exist [9b]. SistematX (http://www.sistematx.ufpb.br), which was developed at the Laboratory of Cheminformatics of the Federal University of Paraíba, the Dictionary of Natural Products—DNP (http://dnp.chemnetbase.com) and NAPRALERT (http://www.napralert.org) are some examples of natural product databases that have compiled structural information to facilitate directed searches. In addition, these databases provide more specific and useful properties for the drug development research area, such as ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) properties [10].

Sesquiterpene lactones (SLs) are one of those interesting small molecules identified in the search for new chemotherapies against infectious diseases [11]. SLs are compounds with 15 carbon atoms that include one or more lactone rings, are abundant in a wide range of natural sources, and are chemotaxonomic markers of the Asteraceae family, which is composed of more than 23,000 species worldwide [1b]. SLs present a wide range of biological activities, including anticancer, anti-inflammatory, and antimalarial activities, as is the case for artemisinin from Artemisia annua (Nobel Prize in Physiology or Medicine, 2015) [12]. The entire SL database stored in SistematX was used in the present study.

In recent years, Computer Aided Drug Design (CADD) approaches have increased our knowledge about the class of secondary metabolites. For example, chemotaxonomic studies of the Asteraceae family have been performed using self-organizing maps and neural networks, and QSAR studies have identified some sesquiterpene lactones with promising activities against

infectious diseases, such as 4,15-isoatripicolide tiglate, which was the most potent structure against *T. brucei* in a purely *in silico* study of 1,750 SLs. This result was subsequently validated in an *in vitro* test (IC_{50} = 1.5 μM) [13].

Some studies have shown how ligand-based QSAR models complement structure-based approaches by boosting the prediction performances when used in tandem; in addition, ligand-based modeling has been used to select a more appropriate protein conformation for docking and assess the reliability of the docking experiment [14]. From this perspective, this paper presents an approach combining ligand-based virtual screening (VS, using the three *T. cruzi* parasitic forms: amastigote, trypomastigote and epimastigote) and structure-based virtual screening (employing eight target proteins) with the aim of selecting structures with potential antichagasic activity from the 1,306 SLs in the SistematX database.

2. Results and Discussion

In the present study, the potential antichagasic activity of 1,306 SLs obtained from an in-house database (SistematX) was predicted using two VS approaches performed simultaneously: ligand-based VS using three (amastigote, promastigote and epimastigote) RF models generated using VolSurf descriptors and structure-based VS using molecular docking with eight *T. cruzi* target enzymes.

Figure 1 shows the general scheme for the virtual screening methodologies used in the present study. Here, we followed the good practices of QSAR and Docking reported in the literature [15]. Initially, data were curated from the ChEMBL database. Subsequently, for the constructed RF models, applicability domains of both the test set and SL structures were evaluated, and all models were rigorously evaluated using and internal and external validation. In the same way, for structure-based VS, the main premises for best practices of docking were also respected: availability of data, realistic representations of the chemical structures, exhaustive sampling of the protein-ligand conformational space and the correct ranking with a score representing the physico-chemical features of the interaction.

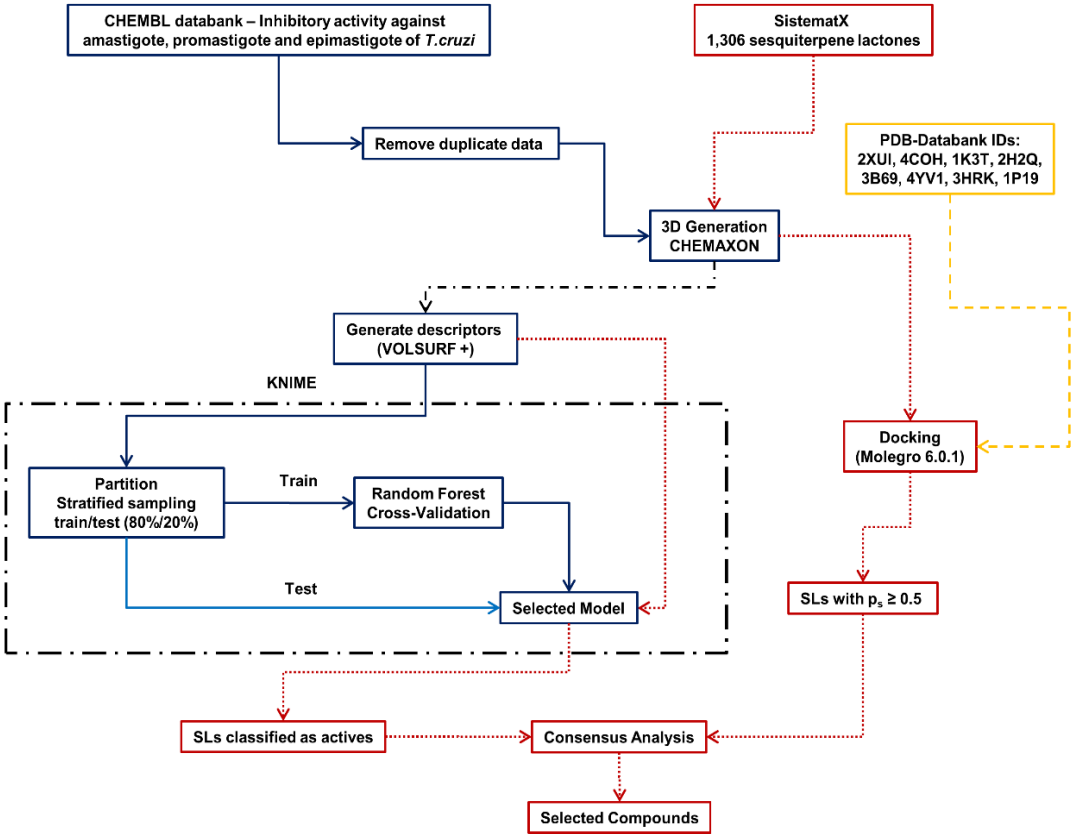


Figure 1: Virtual screening methodology used in this study. Solid blue lines represent the three sets of compounds used to generate the RF model for *T. cruzi* amastigotes, trypomastigotes and epimastigotes, and their validation (clear blue line – external test set). Red dotted lines represent the SLs of Asteraceae obtained from SistematX (in-house database). The black dashed-dotted line represents both datasets (CHEMBL and SistematX). The yellow dashed line represents the eight *T. cruzi* protein structures extracted from the PDB databank (PDB ID: 2XUI, 4COH, 1K3T, 2H2Q, 3B69, 4YV1, 3HRK, 1P19). The dash-dot border delimits the process performed using KNIME software.

2.1. Ligand-based virtual screening

The calculation of 128 descriptors by the Volsurf+ program for each structure dataset composed of 4,452, 1,635 and 1,322 structures for amastigotes, trypomastigotes and epimastigotes, respectively, as well as for the 1,306 SLs in the dataset required approximately 2 hours and 30 minutes to complete using a computer with an i7-4790 processor, running at 3.6 GHz, and equipped with 16 GB of RAM memory. The descriptors, together their dependent variable (binary classification—active or inactive), were used as input data in the KNIME program for generating the Random Forest (RF) model. The cut-off used in the present study

($pIC_{50}\geq5$) was selected to maximize the representativeness of the chemical space for each class of structures (active – inactive) as well as diminish the false positive rate of the model. Using a cut-off of $pIC_{50}\geq4$ for amastigotes, only 250 of the 3,352 structures obtained from ChEMBL were classified as inactive, and thus the CV false positive rate was greater than 65%. Even when a cut-off point of $pIC_{50}\geq4.7$ was used, the model still reached a false positive rate of approximately 50%. Using a cut-off of $pIC_{50}\geq5$, all models showed CV false positive rates of less than 26%. Several random splits were tested by maintaining the same proportions of 80% - 20% for the training and test sets; however, stratified sampling was finally selected since the false positive rate was lower for cross-validation and test sets (Supporting information). The results of the validation of the RF models are summarized in Table 1.

Table 1. Summary of internal cross-validation and test results and corresponding match results obtained using the RF algorithm for the total set of 4452 compounds (3562 in the training set and 890 in the test set) for amastigotes, 1635 compounds (1308 in the training set and 327 in the test set) for trypomastigotes and 1322 compounds (1058 in the training set and 264 in the test set) for epimastigotes.

		Cross-validation			Test		
Model		**Samples**	**Match**	**% Match**	**Samples**	**Match**	**% Match**
Amastigotes	Active	1859	1426	**76.7**	465	368	**79.1**
	Inactive	1703	1271	**74.6**	425	318	**74.8**
	Total	3562	2697	**75.7**	890	686	**77.1**
Trypomastigotes	Active	594	418	**70.4**	149	97	**65.1**
	Inactive	714	581	**81.4**	178	150	**84.3**
	Total	1308	999	**76.4**	327	247	**75.5**
Epimastigotes	Active	386	259	**67.1**	96	62	**64.6**
	Inactive	672	575	**85.6**	168	153	**91.1**
	Total	1058	834	**78.8**	264	215	**81.4**

The training set hit-rate values for the three RF models are quite close to or exactly 100%; nevertheless, for cross-validation and test sets values range from 64.6% to 91.1%, with epimastigote and trypomastigote models serving as better predictors of inactive molecules than the amastigote model, in which the hit percentage of active compounds is slightly higher than the hit rate for inactive compounds. Therefore, these models were selected based on two parameters: specificity (true negative rate) and sensitivity (true positive rate). These parameters are fundamentally required to diminish the number of false positive compounds identified (1-

specificity). The specificity of the epimastigote model is better than the other two models, as the percentage of true negative compounds predicted in the test set (91.1%) was higher than the cross-validation set (85.6%). The amastigote model is the most sensitive of the three, presenting a true positive prediction rate of 76.7% and 79.1% for the cross-validation and test sets, respectively. In turn, the models for the two other parasitic forms were approximately 10% less sensitive to the values reached in the amastigote model. Notably, for the smaller ChEMBL series (epimastigote and trypomastigote), structures with pIC_{50} values ranging from 4.7 to 5.0 (range of 0.3 units) were excluded to avoid edge effects and improve the predictive capacity of the models, since it minimizes the differences in the activity values resulting from the errors and different experimental protocols[16]. This criterion was not applied to the amastigote series because it contained a much higher number of molecules than the other two models and therefore a greater number of structures exists for each pIC_{50} value.

Two parameters were calculated to evaluate the quality of these binary models: Receiver Operating Characteristic (ROC) curve [17] and Matthews's correlation coefficient (MCC) [18]. Initially, ROC curves for the three models were generated and the area under the curve of the plot of sensitivity versus 1-specificity was analyzed. In all cases, a good separation of values was observed between the two groups, without perfect overlap of distributions. The correlation with the training set and values for the area under the curve were 0.85, 0.86 and 0.84 (cross-validation) and 0.85, 0.89 and 0.84 for the amastigote, epimastigote and trypomastigote models, respectively (Figure 2).

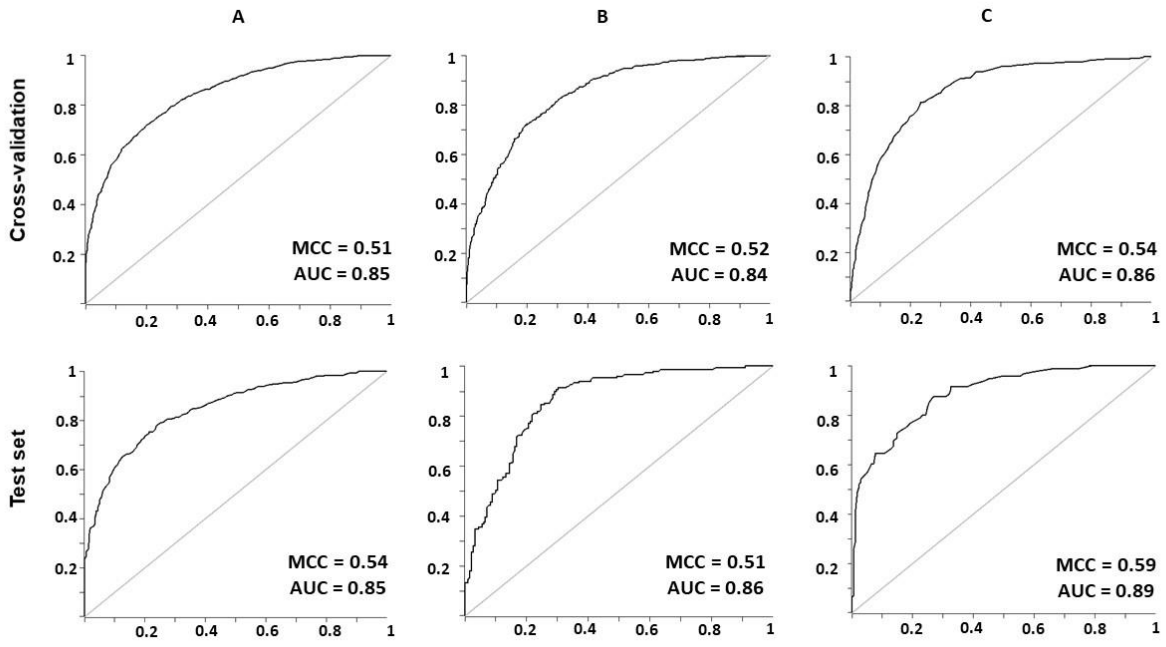


Figure 2. ROC plot, sensitivity versus 1-specificity, generated for the selected RF model for cross-validation and test sets: A) amastigote, B) trypomastigote and C) epimastigote. AUC = value of the area under the curve; MCC = Matthews Correlation Coefficient.

Likewise, the MCC parameter that relates all values of the confusion matrix was calculated using the following equation:

Equation 1:

$$\mathrm{MCC} = \frac{\mathrm{TP\ x\ TN} - \mathrm{FP\ x\ FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

where TP = true positive rate, TN = true negative rate, FP = false positive rate and FN = false negative rate [18].

Here, an MCC value equal to 1 indicates a perfect correlation, 0 indicates a random prediction and –1 indicates total disagreement between the prediction and observation [18]. MCC values for training, cross-validation and test sets in the RF models were obtained. For all three models, a perfect correlation (MCC = 1) was observed for the training set, with the epimastigote model presenting the highest values 0.54 (cross-validation) and 0.59 (test). All MCC values are shown in Figure 2.

The applicability domain was used to assess the reliability of the predictions for the samples in the test and SL sets. For the three test sets more than 99.2% of molecules were classified as

reliable, and only two molecules in each model were classified as unreliable. Meanwhile, for the SL set, more than 99.6% of molecules were classified as reliable in each model, and three, four, and five molecules were classified as unreliable for amastigote, trypomastigote and epimastigote RF models, respectively.

Afterwards, predictions for the three models were performed using previously reported SL structures with trypanocidal activity. The applicability domain for the three series was calculated, and all molecules were classified as reliable in each RF model tested. In the three cases, the models were highly restrictive, which is a fundamental concept for this type of study to minimize the probability of selecting inactive molecules as actives compounds (FP) and therefore recording a high TN rate.

Specifically, for trypomastigotes, a correlation between the active probability value determined using RF model and experimental pIC_{50} was observed ($r^2 = 0.69$). For the four tested compounds, the observed *p* values were less than 0.38, with a high rate of FN structures (three structures). However, this rate is not bad for VS models with the purpose of selecting active molecules. Similarly, five structures with experimental pIC_{50} values for amastigotes were evaluated in the RF model and the results showed similar *p* values for structurally related SLs. However, the only active molecule of the series showed the lowest active probability value and was classified as an FN. Finally, the series of 11 molecules with pIC_{50} values for *T. cruzi* epimastigotes showed an FP rate of less than 10% (CV FP rate was of 14.4%, Table 1) and thus one structure was classified as active when it should have been classified as inactive. Notably, two compounds with *in vitro* activity against this parasitic form were also classified as active in the RF models (Supporting information).

Therefore, using this machine learning algorithm, a virtual screen was performed on a set comprising 1,306 molecules obtained from SistematX, an in-house database (some molecules were excluded due to structural problems). For amastigotes, 34 SLs were predicted to be antichagasic compounds, with probability values ranging from 0.50 to 0.58. Some common structural features are observed among the structures with higher probability values, SLs **1**–**5** (Figure 3). All five are acetylated molecules and four **(1**–**4)** are germacranolides. Likewise, with the exception of compound **3**, the other four structures contain an epoxide moiety in their structures.

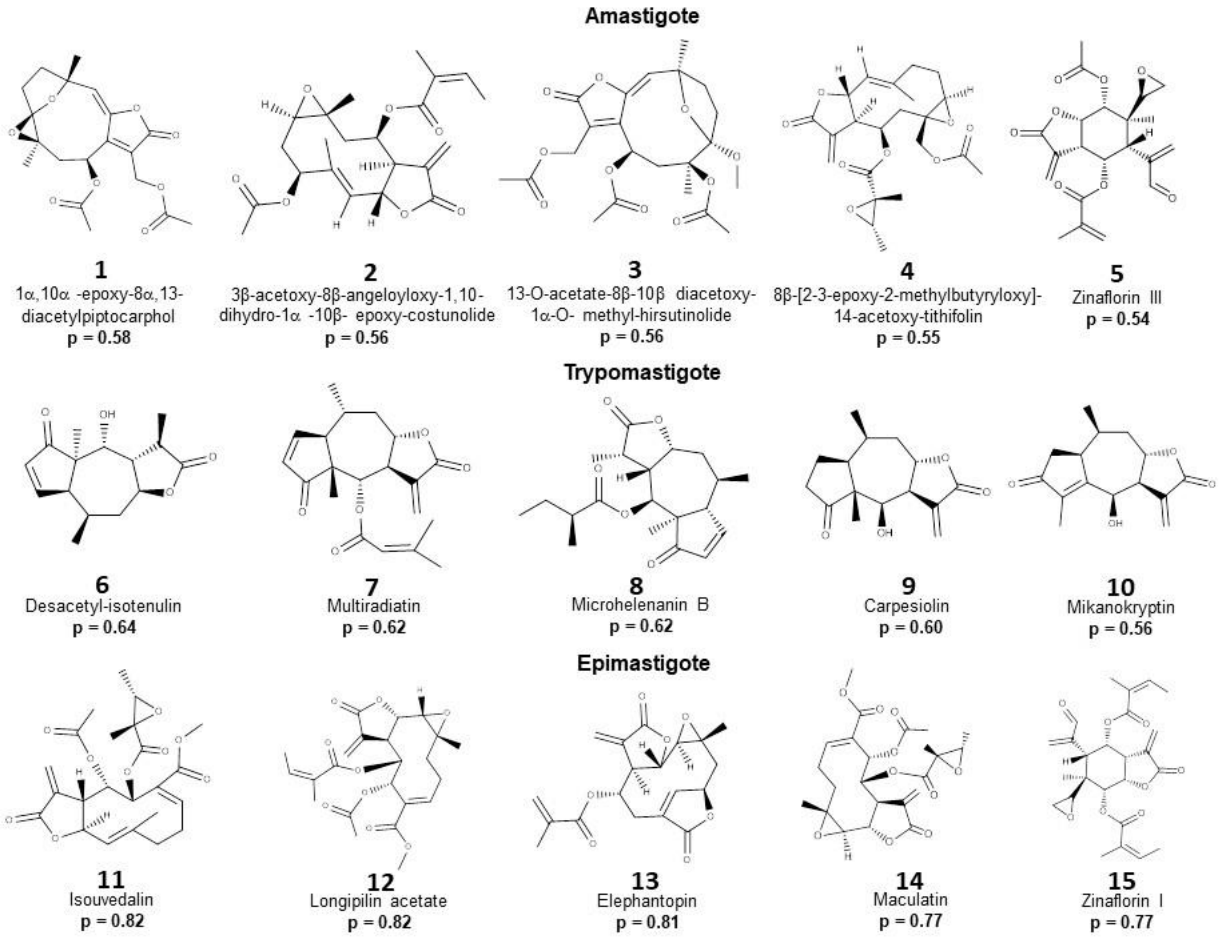


Figure 3. Potentially active sesquiterpene lactones (best five ranked) identified using a ligand-based virtual screen of each parasitic form of *T. cruzi*; *p*= active probability value.

Otherwise, 17 SLs were predicted to be anti-*T. cruzi* compounds for the trypomastigote parasitic form, with probability values ranging from 0.50 to 0.64. Desacetyl-isotenulin (SL **6**, Figure 3) was the structure with the highest probability value. The structures of the active molecules **(6-10)** are similar (guaianolides). In all cases, a ketone group is present in ring A; however, in carpesiolin **9** (p = 0.60), a conjugated system is not present in this ring (α,β-unsaturated carbons). Finally, the epimastigote model was less selective than the other two models, as 420 active molecules were predicted, with probability values ranging from 0.50 to 0.82. As in the amastigote model, structural similarity was observed between SLs with higher probability values (**11–15**), such as the presence of an epoxide moiety in all compounds and a common skeleton (germacranolide) for four of the five structures, including isouvedalin **11,** which has the highest *p* value (0.82).

2.2. ***Structure-based virtual screening***

Initially, the molecular docking was validated by redocking of the original ligand for each one of eight *T. cruzi* proteins used in the present study. These scores are listed in Table 2 along with their respective RMSD values. Then, using the same parameters for all proteins, a virtual screen of 1,306 SLs was performed. Based on the binding energy values, all tested molecules were ranked using the following probability calculation (p_s, Equation 2):

Equation 2:

$$p_s = \frac{\mathrm{E_i}}{\mathrm{E_{min}}} \quad \text{if} \quad \mathrm{E_i} < \mathrm{E_{ligand}}$$

where p_s = structure-based probability; E_i = docking energy of compound *i*, and *i* ranges from 1 to 1306 (SLs dataset); E_{min} = the lowest energy value of the dataset; E_{ligand} = the ligand energy from protein crystallography.

This equation proposed for structure-based VS aims to normalize scores obtained from molecular docking to generate values comparable to the active probability values obtained from ligand-based VS. Additionally, a principle of selection is that the structures must have an energy lower than the value obtained for the ligand in the crystallography study. SLs were classified as active if the structure-based probability values are greater than or equal to 0.5.

The numbers of compounds with probability values greater than 0.5 and binding energy values less than the ligand were: Cruzain (753), sterol 14α-demethylase (642), glyceraldehyde-3-phosphate dehydrogenase (379), dihydrofolate reductase-thymidylate synthase (311), trans-sialidase (211), and hypoxanthine phosphoribosyl transferase (452). For histidyl-tRNA synthetase and spermidine synthase (E.C.2.5.1.16), 122 and 46 compounds have p_s>0.5, respectively, but the binding energy values for all SLs are higher than the PDB ligand (S4M–301). The best-ranked molecules for each enzyme are listed in Table 2 and Figure 4. All binding energy values are presented in the Supporting information.

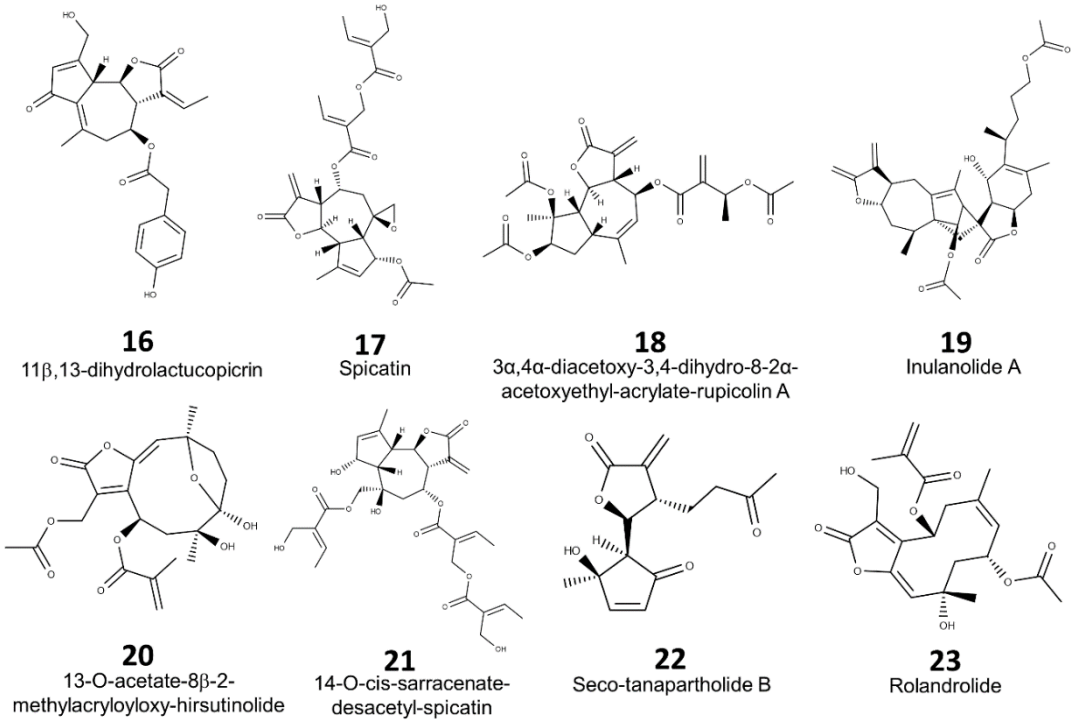


Figure 4. Antichagasic SLs (best ranked) identified in the structure-based virtual screen; p_s = active probability value obtained in the structure-based approach.

Table 2. The docking energy (kJ/mol) of the best-ranked SLs from the structure-based approach for each of the eight *T. cruzi* enzymes studied. Ligand = energy (kJ/mol) for the PDB ligand and the RMSD values obtained from the redocking procedure.

***T. cruzi* protein**	**SL**	**Ligand**	**Redocking RMSD**
Cruzain	**16** (-91.4)	(-80.0)	0.79
Sterol 14α-demethylase CYP51	**17** (-113.6)	(-106.8)	0.70
Dihydrofolate Reductase-Thymidylate synthase	**18** (-82.1)	(-74.0)	0.16
Glyceraldehyde-3-phosphate dehydrogenase	**19** (-94.6)	(-26.6)	0.48
Hypoxanthine phosphoribosyltransferases	**20** (-107.5)	(-88.4)	0.56
Histidyl-tRNA synthetase	**21** (-133.6)	(-165.0)	0.75
Spermidine synthase	**22** (-107.9)	(-167.1)	0.15
Trans-sialidase	**23** (-103.2)	(-26.9)	0.26

Some structural features are observed among the SLs with the highest active probability values for the eight *T. cruzi* enzymes selected in this study. The guaianolide skeleton is observed in five SLs (**16-19, and 21**), whereas the germacranolide skeleton (**20 and 23**) was present in

the best ligand conformation for hypoxanthine phosphoribosyltransferase and trans-sialidase. Only spermidine synthase displays a different pattern, for which seco-tanapartholide B **22** represents the most potential active sesquiterpene lactone (Table 2).

The structures 11β, 13-dihydrolactucopicrin **16** (Figure 4) and crepidiaside B **24** (Figure 5A) are two guaianolide SLs extracted from *Lactuca georgica*, a species distributed in the Euxinian-Hyrcanian region of Southwest Asia, i.e., Caucasia, Northeast Anatolia and North Iran [19], and present the highest active probability values in structure-based VS for Cruzain (PDB ID: 4XUI). Figure 5 shows the conformations of both SLs in the active site of Cruzain, as well as the hydrogen-bonding (H-bond) interactions of compound **16** (Figure 5B) with residues Cys 25, Trp 26 and Trp 184. Molecule **24** also participated in H-bond interactions with Cys 25 and Trp 184. In both SLs, an H-bond was observed between the carbonyl moiety of carbon-2 with Trp 184.

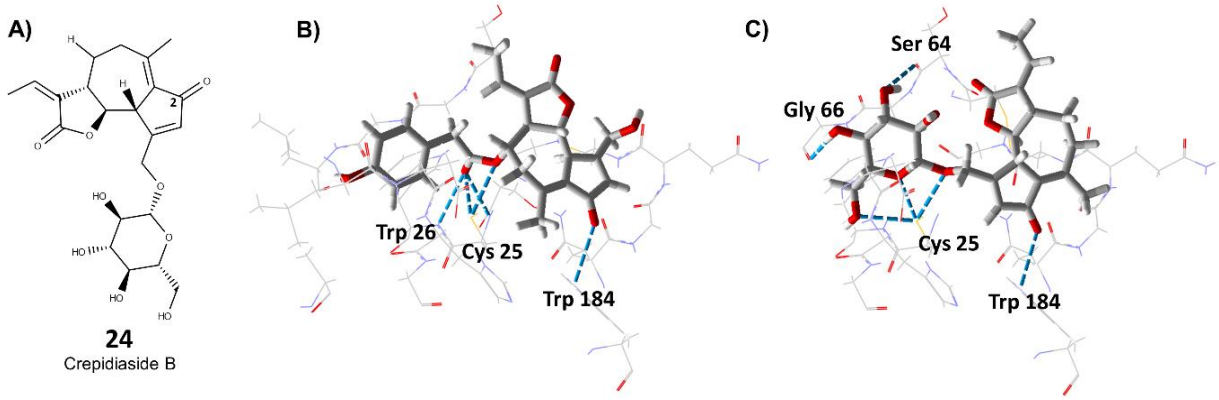


Figure 5. A) Structure of crepidiaside B (**24**). Docking conformations of 11β, 13-dihydrolactucopicrin (**16**) B) and crepidiaside B (**24**) C) in the pocket of *T. cruzi* Cruzain (PDB ID: 4XUI). The blue dotted line represents H-bond interactions between SLs **16** and **24** with Cruzain residues (black labels).

For dihydrofolate reductase-thymidylate synthase (PDB ID: 2H2Q), two SLs with different skeletons, a guaianolide obtained from *Erlangea inyangana*, 3α,4α-diacetoxy-3,4-dihydro-8-2α-acetoxyethyl-acrylate-rupicolin A (**18),** and a germacranolide extracted of *Chresta sphaerocephala*, chrestanolide (**25**), displayed the highest active probability values. Figure 6 presents H-bond interactions for the PDB ligand 2'-deoxyuridine-5'-monophosphate and the two best-ranked SLs (**18** and **25**). Some critical residues for the interaction of the three structures, Arg 423, Ser 424 and Asn 434, are identified. Interestingly, despite their structural

differences, both SL **18** and **25** interact in the same manner with Asn 134 through an H-bond formed with the carbonyl group (carbon-12) of the lactone ring.

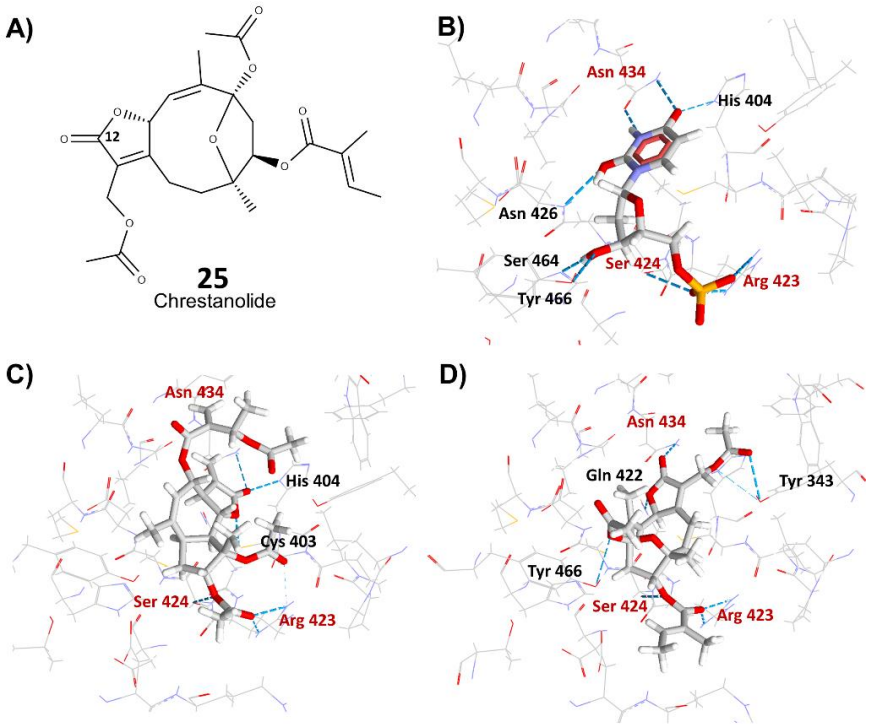


Figure 6. A) Structure of chrestanolide (**25**). Docking conformations of 2'-deoxyuridine-5'-monophosphate (PDB ID: DU) B), 3α,4α-diacetoxy-3,4-dihydro-8-2α-acetoxyethyl-acrylate-rupicolin A (**18**) C) and chrestanolide (**25**) D) with *T. cruzi* dihydrofolate reductase-thymidylate synthase (PDB ID: 2H2Q). Hydrogen-bonding interactions are illustrated as blue dotted lines. Critical residues Arg 423, Ser 424 and Asn 434 are highlighted in red.

2.3. Combined approach Structure-Based virtual screening and Ligand-Based virtual screening

Using an approach combining structure-based and ligand-based virtual screening, we verified potentially active molecules as well as their possible mechanism of action, facilitating the identification of potential multitarget compounds. In addition, this approach also tries to minimize the probability of selecting false positive compounds because SLs are classified as potential active structures through a new probability score (p_c, Equation 3), which is determined using the scores (*p* values) of both virtual screening methodologies and relates them to the true negative rate.

Equation 3:

$$p_c = \frac{p_s + (1 + TN) \times p}{2 + TN}$$

where p_c = combined probability p_s = structure-based probability; TN = true negative rate; p = ligand-based probability.

We propose a new equation based on the fact that ligand-based virtual screening is a methodology that uses known active and inactive structures (pIC_{50} experimental values); the score (p) has a high weight in this calculation with respect to the docking probabilities (p_s), which only relates interactions between the protein and ligand. In Equation 3, the weight of the ligand-based score is related to a decrease in the false positive rate (increment of TN). In this approach, the probability of selecting inactive molecules as active compounds (FP) is minimized, since the selection of this type of molecule is performed with a significant increase in time and cost, which contradicts the aims of VS studies.

Table 3 summarizes the results for the best-ranked SLs obtained using the combined approach. Some structures that previously displayed a high active probability value in the ligand-based virtual screen appear to be interesting potential multitarget structures (acting on two or more proteins of *T. cruzi*) for each parasitic form, e.g., structures **1**, **7** and **8**. Likewise, SLs **26-35** (Figure 7), which were not clearly highlighted either in structure-based or ligand-based virtual screening but were classified as active by the two methodologies (p and p_s values>0.5), now appear as potential antichagasic SLs, with the exception of compound **28**.

Table 3. Summary of the best-ranked structures obtained using an approach combining ligand-based and structure-based virtual screening; p = active probability value in ligand-based VS; p_s = active probability value in structure-based VS. p_c = combined probability value

	Amastigote				Trypomastigote				Epimastigote			
Protein	**Structure**	***p***	***p*$_s$**	***p*$_c$**	**Structure**	***p***	***p*$_s$**	***p*$_c$**	**Structure**	***p***	***p*$_s$**	***p*$_c$**
Cruzain	**1**	0.58	0.83	0.67	**7**	0.62	0.64	0.63	**32**	0.73	0.91	0.79
Sterol 14α-demethylase CYP51	**26**	0.51	0.87	0.64	**8**	0.62	0.84	0.70	**13**	0.81	0.78	0.80
Dihydrofolate Reductase-Thymidylate synthase	**1**	0.58	0.71	0.63	**30**	0.51	0.72	0.59	**33**	0.70	0.66	0.72
glyceraldehyde-3-phosphate dehydrogenase	**1**	0.58	0.73	0.64	**8**	0.62	0.54	0.59	**11**	0.82	0.68	0.77
Histidyl-tRNA synthetase	**27**	0.50	0.77	0.60	**6**	0.64	0.63	0.64	**6**	0.68	0.63	0.66
Hypoxanthine phosphoribosyltransferases	**27**	0.50	0.56	0.54	**8**	0.62	0.57	0.60	**34**	0.67	0.86	0.74
Spermidine synthase	**28**	0.39*	0.85	0.56	**31**	0.52	0.61	0.55	**35**	0.55	0.72	0.61
Trans-sialidase	**29**	0.51	0.73	0.59	**7**	0.62	0.53	0.59	**12**	0.82	0.56	0.73

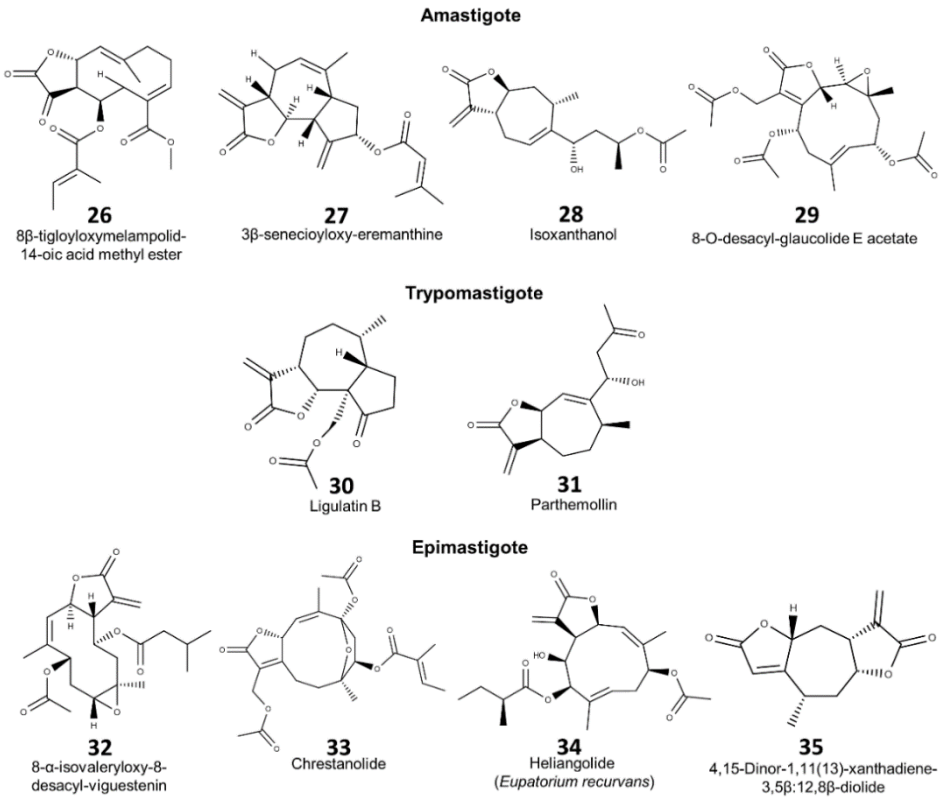


Figure 7. Representations of the best-ranked structures identified using an approach combining ligand-based and structure-based virtual screening for each parasitic form.

Among the molecules listed in Table 3, 3β-senecioyloxy-eremanthine (**27,** Figure 7)**,** an SL isolated from *Zinnia acerosa*, is highlighted as a new potential multitarget structure acting on histidyl-tRNA synthetase and hypoxanthine phosphoribosyltransferases. On the other hand, desacetyl-isotenulin (**6**, Figure 3) a guaianolide characteristic of genera *Arnica* and *Helenium* that presented the highest probability value in the ligand-based VS for trypomastigotes (p = 0.64), is the only potential active SL for two *T. cruzi* parasitic forms: epimastigote (p_c = 0.66) and trypomastigotes (p_c = 0.64).

Another approach combining two ligand-based virtual screens was used to predict potentially antichagasic SLs. In addition to the model based on parasitic forms, an RF model was developed with Cruzain as the target. We propose Equation 4 to calculate the consensus probability value (p_{LB}) for the two RF models for the entire set of 1,306 SLs. In this equation, both scores have the same weight for calculations based on the TN rate, because we combined two ligand-based VS. The aim of this score is to minimize the probability of selecting an FP compound.

Equation 4:

$$p_{LB} = \frac{(p_z \times TN_z) + (p \times TN)}{TN_z + TN}$$

where p_{LB} = combined probability p_z = ligand-based Cruzain probability; TN = CV true negative rate for the parasitic form model; p = ligand-based parasitic form probability; TN_z = true negative rate for the Cruzain model.

For the epimastigote parasitic form, 110 compounds display p_{LB} values greater than 0.5; however, only 28 were predicted to be active structures in both models. Structure **36**, 9α-linolenoyloxy-15-hydroxy-8β-2-methylbutyryloxy-14-oxo-acanthospermolide, (Figure 8) has the highest combined probability value (p_{LB} = 0.66). This structure is a characteristic secondary metabolite of *Acanthospermun hispidium*, a plant native to Central and South America, and it contains a linolenic acid ester as substituent in its germacranolide skeleton.

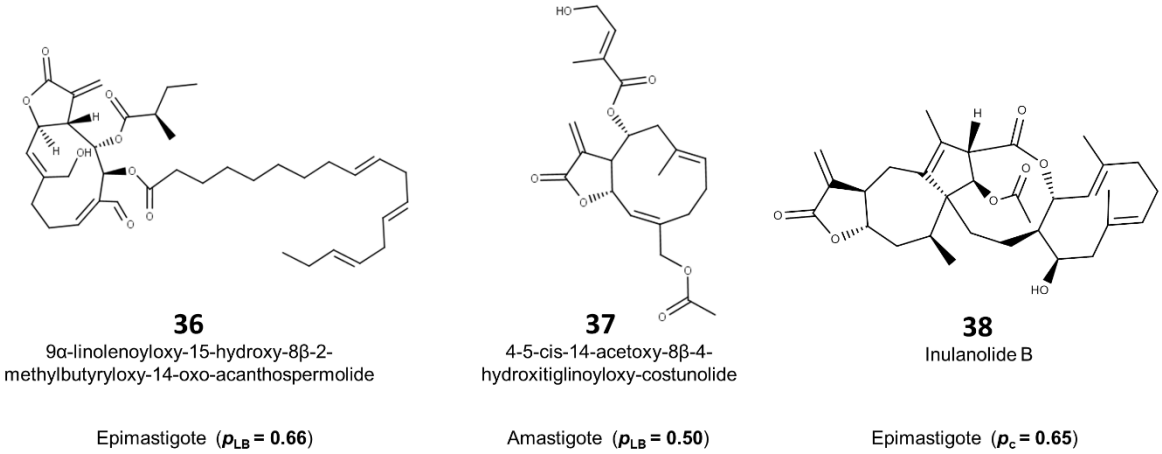


Figure 8: Potential SL inhibitors of *T. cruzi* Cruzain (**36-38**) identified by the approach combining ligand-based and structure-based virtual screens. p_{LB} = combined probability value. p_c = combined probability value.

On the other hand, using this approach combining two ligand-based virtual screens, only one active SL against amastigote (**37**, Figure 8) was identified. The compound 4-5-cis-14-acetoxy-8β-4-hydroxitiglinoyloxy-costunolide (**37,** p_{LB} value = 0.50) is a heliangolide from *Eupatorium hyssopifolium* and represents a promising structure targeting the amastigote form because it is considered active in the RF model (p = 0.52). However, in neither case is the best-ranked compound active in the two individual models, with combined ligand-based probability values (p_z) less than 0.5. All results are shown in the Supporting information. For trypomastigotes, no SL reached a $p_{LB}\geq 0.50$ in the combined model, but compounds **7** and **8** represent the structures with the highest p_{LB} values (0.46 and 0.47). These two molecules are also the most active for several targets in the combined approach (structure-based VS, ligand-based VS).

Likewise, after combining the three models (two ligand-based and structure-based VS), compound **38** (Figure 8) has the highest probability value (p_c =0.65) among 6 structures that appeared to be active in all three models. Thus, this molecule in an interesting antichagasic hit for epimastigotes. For the other two parasitic forms, 31 compounds (amastigote) and one (trypomastigote) molecule showed probability values (p_c) greater than 0.50; however, no SL structures were classified as active by the three models (Supporting information).

In addition, an optimization of best-ranked structures by molecular dynamics using PELE (Protein Energy Landscape Exploration), a technology based on protein structure prediction

algorithms and a Monte Carlo sampling, is capable of efficiently and rapidly modeling the dynamic all-atom protein–ligand interactions, with a two order of magnitude reduction in computational cost compared with traditional molecular dynamics techniques. This web server combines a Monte Carlo stochastic approach with protein structure prediction algorithms and is capable of accurately reproducing long time scale processes [20] .

The ligand binding refinement of the best conformations for each parasitic form was performed. The results show an RMSD value for the minimized conformation with respect to the docking conformation (Supporting information). Spermidine synthase structures present the lowest values for RMSD (less than 1.79), whereas hypoxanthine phosphoribosyltransferases and trans-sialidase show the high RMSD values (greater than 4.53).

In Figure 9, minimized (green) and docking conformations (yellow) for 4,15-Dinor-1,11(13)-xanthadiene-3,5β:12,8β-diolide (Figure 7, compound **35**) are presented in the active site of spermidine synthase. The docking conformation of compound **35** only presented an H-bond interaction with Gly150, whereas two additional H-bond interactions with Gly96 and Gly97 are observed in the minimized structure. A slight spatial rotation of this structure enables the interaction of the second lactone ring with these two residues. For the other minimized SLs, alterations in the observed interactions were also noted.

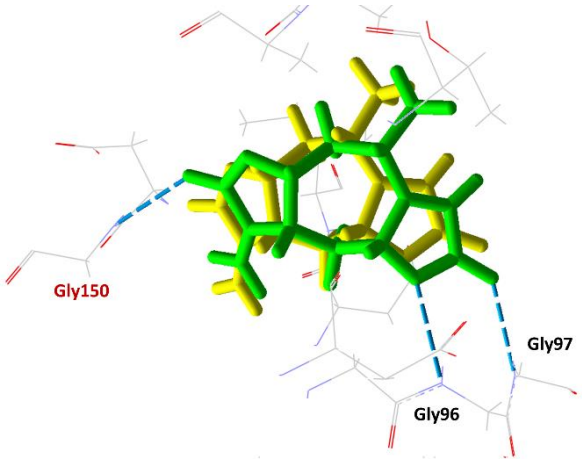


Figure 9: Superposition of the docking conformation (yellow) and minimized conformation (green) in the pocket of *T. cruzi* spermidine synthase. The blue dotted lines represent H-bond interactions between the minimized conformation and spermidine synthase residues (black labels). The common interaction with Gly150 is highlighted in red.

3. Conclusions

In the present study, we performed ligand-based and structure-based virtual screens of 1,306 SLs of *Asteraceae* obtained from an in-house database. Potential antichagasic SLs for the three parasitic forms and some structural features, such as skeleton type and the presence of epoxide moiety, were determined from RF models of *T. cruzi*. In addition, a structure-based virtual screen using PDB structures of eight *T. cruzi* proteins for the entire SL set allowed the selection of 13 potential inhibitors of these enzymes. Finally, an approach combining structure-based and ligand-based virtual screening enabled the identification of promising multitarget antichagasic SLs.

Therefore, these virtual screening approaches appear to be interesting alternatives as an initial step in further studies of these secondary metabolites aimed at identifying new, potentially active antichagasic compounds, as well as their possible mechanism of action (both single and multitarget compounds).

4. Experimental Section

4.1.Database

From the ChEMBL database (https://www.ebi.ac.uk/chembl/), we selected a diverse set that was initially classified according to the *T. cruzi* parasitic form: amastigote (4,455 structures), trypomastigote (1,820 structures) and epimastigote (1,591 structures). Compounds were classified using values of $-\log IC_{50}$ (mol/L) = pIC50, which led us to divide them into active ($pIC_{50} \geq 5$) and inactive ($pIC_{50} < 5$) structures. Due to the variability of the experimental protocols in the ChEMBL database, a qualitative pattern was used in the present study. Differences in activity values related to experimental protocols and even different strains are partially minimized in the binary classification. Here, IC_{50} represented the concentration required for 50% inhibition of parasite growth. For each ChEMBL dataset, 15 (amastigote set), 40 (trypomastigote set) and 11 (epimastigote set) SL structures [11, 21], that are not included in the secondary metabolite database were submitted to ligand-based VS, to which their pIC_{50} values were added to increase the representativeness of the generated models regarding the chemical space of this class of secondary metabolites. Afterwards, the applicability domain (APD) based on the Euclidean distances was used to identify compounds in the test set for which predictions may be unreliable if the values are greater than $APD = d + Z\sigma$, where d and σ are the average Euclidian distance and standard deviation of the set of samples in the training set with lower

average Euclidian distance values than all samples in the training set. The parameter Z is an empirical cut-off value, and 0.5 was used as the default value [22]. For the smaller series, structures with pIC_{50} values ranging from 4.7 to 5.0 (range of 0.3 units) were excluded, with the aims of avoiding edge effect and improving the predictive capacity of the models since it minimizes the differences in the activity values resulting from errors and different experimental protocols [16].

Data from datasets were curated according to procedures suggested in the literature [23]. Initially, mixtures and organometallic compounds were excluded from the database. All ions were removed from the remaining structures. Standardizer software (Jchem, version 16.11.28 (2016), a calculation module developed by ChemAxon, http://www.chemaxon.com/) was used to canonize all SMILES codes, and duplicate structures and structures with higher pIC_{50} values were removed in an attempt to construct more restrictive models, since only lower activity values are used. Some molecules were excluded due to structural problems. Finally, the number of structures for each dataset was 2,324, 743 and 482 active compounds, and 2,128, 892 and 840 inactive compounds for amastigotes (4,452 structures), trypomastigotes (1635 structures) and epimastigotes (1,322 structures), respectively. From another perspective, a set of 1,306 SLs was obtained from the SistematX database (http://sistematx.ufpb.br). For predictions, SL structures included in the ChEMBL dataset were excluded for each parasitic form. For all structures, SMILES codes were used as input data in Marvin; ChemAxon (version 16.11.28 (2016), a calculation module developed by ChemAxon, http://www.chemaxon.com/). We used Standardizer software (Jchem, version 16.11.28 (2016), a calculation module developed by ChemAxon, http://www.chemaxon.com/); ChemAxon to canonize structures, add hydrogens, perform aromatic form conversions, and clean the molecular graph in three dimensions. The process uses a divide and conquer approach. The structure is split into small fragments that are organized into a tree using connectivity information. Conformers generated for the initial structure (represented by the root node in the tree) are optimized. The tree building process uses a proprietary extended version of the Dreiding force field [24]. Molecules that presented structural problems in the 3D generation were manually corrected using Marvin. (The same methodology was performed for a dataset with *T. cruzi* Cruzain as the target, which was also obtained from ChEMBL (483 structures). The number of structures in this dataset was 222 active compounds and 261 inactive compounds.

4.2. Volsurf+ descriptors

Three-dimensional (3-D) structures in SDF format were used as input data in the Volsurf+ program v. 1.0.7 [25] and were subjected to molecular interaction fields (MIFs) to generate descriptors using the following probes: N1 (amide nitrogen–hydrogen-bond donor probe), O (carbonyl oxygen–hydrogen-bond acceptor probe), OH2 (water probe) and DRY (hydrophobic probe). Additional non-MIF-derived descriptors were generated to create a total of 128 descriptors [25].

VolSurf descriptors are rarely influenced by conformational sampling and averaging, probably due to the peculiarity of the grid force field, which allows for the conformational flexibility of external groups, hydrogens and lone pairs [26].

4.3. RF models

Knime 3.1.0 software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org)[27] was used to perform all of the analyses described below. Initially, the descriptors calculated using the Volsurf+ program were imported in CSV format and subsequently divided using the "Partitioning" node in the stratified sampling option; 80% of the initial dataset was classified as the training set. Meanwhile, the remaining 20% composed the test set. Additionally, random splits were tested while maintaining the same proportions for the training and test sets. Although the compounds were selected randomly, the same proportion of active and inactive samples was maintained in both sets to maintain the proportion of the training series in the test series and to ensure the diversity of the chemical space for each activity class (active - inactive). The model was generated using the training set and the RF algorithm. For internal validation of the model, we employed cross-validation using 10 randomly selected stratified groups, and the distributions stratified according to activity class variables were maintained in all validation groups and in the training set. The parameters selected for RF included the following settings: number of trees to build = 100, 200 and 500 for epimastigote, amastigote and trypomastigote, respectively, seed for random number generator = 1, and Gini Index as a split criterion for training and internal cross-validation sets. From the confusion matrix, the internal and external performances of the selected models were analyzed using the parameters sensitivity (true positive rate), specificity (true negative rate) and accuracy (overall predictability). In addition, the ROC curve employing a "ROC curve" node, which uses the sensitivity and specificity parameters, was constructed to describe the true performance of the model with greater clarity than accuracy. The plotted ROC curve shows the true positive

(active) rate versus the false positive rates (1-specificity)[17]. In this representation, when a variable of interest cannot be distinguished between the two groups, the area under the ROC curve is 0.5, whereas in the opposite case where a perfect separation of values between two groups without overlap of distributions exists, the area under the ROC curve is 1. The MCC was also calculated using this method, where a value of 1 represents a perfect prediction, a value of 0 represents a random prediction, and a value of −1 represents total disagreement between the prediction and observation. Additionally, two data sets of four and 11 SLs were selected from the literature [21f, 28] as having biological activity against the parasitic forms of *T. cruzi* to analyze the predictive performance of the three RF models for SLs.

4.4.Molecular docking

The structure of eight *T. cruzi* proteins, Cruzain (PDB ID: 4XUI) [29], sterol 14α-demethylase (CYP51–PDB ID: 4COH) [30], glyceraldehyde-3-phosphate dehydrogenase (GAPDH–PDB ID: 1K3T) [31], Dihydrofolate Reductase-Thymidylate synthase (PDB ID: 2H2Q; E.C. 1.5.1.3 and 2.1.1.45) [32], trans-sialidase (PDB ID: 3B69) [33], spermidine synthase (PDB ID: 4YV1) [34], Histidyl-tRNA synthetase (PDB ID: 3HRK; E.C. 6.1.1.21) [35] and Hypoxanthine phosphoribosyltransferase (PDB ID: 1P19; E.C. 2.4.2.8)[36] in complex with the respective inhibitors: *N*-[(2*S*)-5-(carbamimidamidooxy)-1-oxo-1-{[(1*E*,3*S*)-5-phenyl-1-(pyrimidin-2-ylsulfonyl)pent-1-en-3-yl]amino}pentan-2-yl]-4-methylpiperazine-1-carboxamide (PDB ID: 2VC), 2-fluoranyl-*N*-[(2*R*)-3-(1H-indol-3-yl)-1-oxidanylidene-1-(pyridin-4-ylamino)propan-2-yl]-4-(4-thiophen-2-ylsulfonylpiperazin-1-yl)benzamide (PDB ID: T9H), 6-(1,1-dimethylallyl)-2-(1-hydroxy-1-methylethyl)-2,3-dihydro-7H-furo[3,2-g]chromen-7-one (PDB: BRZ), 2'-deoxyuridine-5'-monophosphate (PDB ID: du), 5-(acetylamino)-9-(benzoylamino)-3,5,9-trideoxy-3-fluoro-D-erythro-alpha-L-manno-non-2-ulopyranosonic acid (PDB ID: BFN), 5'-[(*S*)-(3-aminopropyl)(methyl)-λ-4-sulfanyl]-5'-deoxyadenosine monophosphate (PDB ID: S4M), histidyl-adenosine monophosphate (PDB ID: HAM) and inosinic acid (PDB ID: IMP) were downloaded from the Protein Data Bank—PDB. All water molecules were deleted from the enzyme structure, and the enzyme and compound structures were prepared using the same default parameter settings in the same software package (score function: MolDock Score; ligand evaluation: internal ES, internal H-bond, Sp2–Sp2 torsions, all checked; number of runs: 10 runs; algorithm: MolDock SE; maximum interactions: 1500; max. population size: 50; max. steps: 300; neighbor distance factor: 1.00; max. number of conformations returned: 5). The docking procedure was performed using a GRID with a radius

of 15 Å and a resolution of 0.30 Å to cover the ligand-binding site in the structures of the eight enzymes.

Using the PELE (Protein Energy Landscape Exploration) algorithm (https://pele.bsc.es/), a quick protein and ligand optimization procedure was performed and a Monte Carlo acceptance procedure was used to determine if the trial perturbation and optimization produce a likely minimum [20b, 37]. In the web server, the ligand binding refinement script was selected using three CPUs and a maximal wall clock of 24 hours. The PDB file containing the docking conformation and protein structure was uploaded. Afterwards, the file was prepared with the aim of providing more control over several aspects of the protein/ligand preparation. Finally, a maximum ratio of 10 angstroms with 100 steps were selected in Ligand Free Exploration Settings. The minimal ligand binding energy was selected and the respective PDB file was downloaded to analyze the similarity/dissimilarity of the initial docking conformation.

Acknowledgements

We would like to thank the CNPq and Capes for financial Support, and the Student Agreement Program of Graduate — PEC-PG of CNPq—Brazil.

References

[1] a) J. A. Pérez-Molina, I. Molina, *Lancet* **2017**; b) C. Herrera Acevedo, L. Scotti, M. Feitosa Alves, M. Formiga Melo Diniz, M. Scotti, *Molecules* **2017**, *22*, 79; c) GBD 2013 Mortality and Causes of Death Collaborators, *Lancet* **2015**, *385*, 117-171.

[2] Á. d. l. C. Pech-Canul, V. Monteón, R.-L. Solís-Oviedo, *J. Parasitol. Res* **2017**, *2017*, 3751403.

[3] a) C. J. Salomon, *J. Pharm. Sci.* **2012**, *101*, 888-894; b) P. M. Guedes, G. K. Silva, F. R. Gutierrez, J. S. Silva, *Expert Rev. Anti-Infect. Ther.* **2011**, *9*, 609-620.

[4] a) J. A. Castro, M. M. deMecca, L. C. Bartel, *Hum. Exp. Toxicol.* **2006**, *25*, 471-479; b) W. Apt, *Drug Des. Dev. Ther.* **2010**, *4*, 243.

[5] A. R. Teixeira, M. M. Hecht, M. C. Guimaro, A. O. Sousa, N. Nitz, *Clin. Microbiol. Rev.* **2011**, *24*, 592-630.

[6] a) I. A. Oliveira, A. S. Gonçalves, J. L. Neves, M. von Itzstein, A. R. Todeschini, *J. Biol. Chem.* **2014**, *289*, 423-436; b) J. Riley, S. Brand, M. Voice, I. Caballero, D. Calvo, K. D. Read, *PLoS Neglected Trop. Dis.* **2015**, *9*, e0004014.

[7] a) I. Pauli, L. G. Ferreira, M. L. de Souza, G. Oliva, R. S. Ferreira, M. A. Dessoy, B. W. Slafer, L. C. Dias, A. D. Andricopulo, *Future Med. Chem.* **2017**; b) K. E. Rogers, H. Keränen, J. D. Durrant, J. Ratnam, A. Doak, M. R. Arkin, J. A. McCammon, *Chem. Biol. Drug. Des.* **2012**, *80*, 398-405.

[8] H. J. Wiggers, J. R. Rocha, W. B. Fernandes, R. Sesti-Costa, Z. A. Carneiro, J. Cheleski, A. B. da Silva, L. Juliano, M. H. Cezari, J. S. Silva, *PLoS Neglected Trop. Dis.* **2013**, *7*, e2370.

[9] a) D. J. Newman, G. M. Cragg, *J. Nat. Prod* **2016**, *79*, 629-661; b) A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nat. Rev. Drug Discovery* **2015**, *14*, 111; c) B. B. Mishra, V. K. Tiwari, *Eur. J. Med. Chem.* **2011**, *46*, 4769-4807.

[10] M. Paul Gleeson, A. Hersey, S. Hannongbua, *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2011**, *11*, 358-381.

[11] a) T. J. Schmidt, R. Brun, G. Willuhn, S. A. Khalid, *Planta Med.* **2002**, *68*, 750-751; b) T. J. Schmidt, A. M. Nour, S. A. Khalid, M. Kaiser, R. Brun, *Molecules* **2009**, *14*, 2062-2076.

[12] Y. Tu, *Angew. Chem. Int. Ed.* **2016**.

[13] T. J. Schmidt, F. B. Da Costa, N. P. Lopes, M. Kaiser, R. Brun, *Antimicrob. Agents Chemother.* **2014**, *58*, 325-332.

[14] a) F. Broccatelli, N. Brown, *J. Chem. Inf. Model.* **2014**, 54, 1634; b) D. Fourches, E. Muratov, F. Ding, N. V. Dokholyan, A. Tropsha, *J. Chem. Inf. Model.* **2013**, 53, 1915; c) P. Joshi, G. J. McCann, v. Sonawane, R. A. Vishwakarma, B. Chaudhuri, S. B. Bharate, *J. Chem. Inf. Model.* **2017**; d) G. Fu, S. Liu, X. Nan, O. R. Dale, Z. Zhao, Y. Chen, D. E. Wilkins, S. P. Manly, S. J. Cutler, R. J. Doerksen, *Mol. Inf.* **2014**, 33, 627-640.

[15] a) D. Castillo-González, J.-L. Mergny, A. De Rache, G. Pérez-Machado, M. A. Cabrera-Pérez, O. Nicolotti, A. Introcaso, G. F. Mangiatordi, A. Guédin, A. Bourdoncle, *J. Chem. Inf. Model.* **2015**, 55, 2094-2110; b) M. Rueda, R. Abagyan, *bioRxiv, Pharmacol. Toxicol.* **2016**, 039446; cE. N. Muratov, E. V. Varlamova, A. G. Artemenko, P. G. Polishchuk, V. E. Kuz'Min, *Mol. Inf.* **2012**, 31, 202-221.

[16] W. S. Noble, *Nat. Biotechnol*. **2006,** 24, 1565-1567.

[17] J. A. Hanley, B. J. McNeil, *Radiology*.**1982**, 143, 29-36.

[18] B. W. Matthews, *Biochim. Biophys. Acta, Protein Struct.* **1975**, *405*, 442-451.

[19] K. Michalska, A. Beharav, W. Kisiel, *Phytochem. Lett.* **2014**, 10, 10-12.

[20] a) A. Madadkar-Sobhani, V. Guallar, *Nucleic Acids Res.* **2013**, 41, W322-W328; b) K. W. Borrelli, A. Vitalis, R. Alcantara, V. Guallar, *J. Chem. Theory Comput.* **2005**, 1, 1304-1311.

[21] a) V. P. Sülsen, F. M. Frank, S. I. Cazorla, P. Barrera, B. Freixa, R. Vila, M. A. Sosa, E. L. Malchiodi, L. V. Muschietti, V. S. Martino, *Int. J. Antimicrob. Agents*. **2011**, 37, 536-543; b) V. P. Sülsen, S. I. Cazorla, F. M. Frank, L. C. Laurella, L. V. Muschietti, C. A. Catalan, V. S. Martino, E. L. Malchiodi, *PLoS Neglected Trop. Dis.* **2013**, 7, e2494; c) F. M. Frank, J. Ulloa, S. I. Cazorla, G. Maravilla, E. L. Malchiodi, A. Grau, V. Martino, C. Catalán, L. Muschietti, *J. Evidence-Based Complementary Altern. Med.* **2013**, 2013; d) J. Cogo, A. de Oliveira Caleare, T. Ueda-Nakamura, B. P. Dias Filho, I. C. P. Ferreira, C. V. Nakamura, *Phytomedicine.* **2012**, 20, 59-66; e) T. C. Lima, R. d. J. Souza, M. H. d. Moraes, M. Steindel, M. W. Biavatti, *J. Braz. Chem. Soc.* **2017**, 28, 367-375.

[22] M. T. Scotti, L. Scotti, H. M. Ishiki, L. M. Peron, L. de Rezende, A. T. do Amaral, *Chemom. Intell. Lab. Syst.* **2016**, 154, 137-149.

[23] a) A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, *J. Med. Chem.* **2014**, 57, 4977; b) D. Fourches, E. Muratov, A. Tropsha, *Nat. Chem. Biol*. **2015,** 11, 535-535.

[24] A. D. Becke, *Phys. Rev. A.* **1988**, 38, 3098.

[25] G. Cruciani, P. Crivori, P.-A. Carrupt, B. Testa, *J. Mol. Struct.: THEOCHEM* **2000**, *503*, 17-30

[26] G. Cruciani, M. Pastor, W. Guba, *Eur. J. Pharm. Sci.* **2000**, 11, S29-S39.

[27] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, B. Wiswedel, *SIGKDD Explor* **2009**, *11*, 26-31.

[28] a) L. C. Laurella, N. Cerny, A. E. Bivona, A. S. Alberti, G. Giberti, E. L. Malchiodi, V. S. Martino, C. A. Catalan, M. R. Alonso, S. I. Cazorla, *PLoS Neglected Trop. Dis.* **2017**, 11, e0005929; b) L. Fabian, V. Sulsen, F. Frank, S. Cazorla, E. Malchiodi, V. Martino, E. Lizarraga, C. Catalán, A. Moglioni, L. Muschietti, *Mini-Rev. Med. Chem*. **2013**, 13, 1407-1414.

[29] B. D. Jones, A. Tochowicz, Y. Tang, M. D. Cameron, L.-I. McCall, K. Hirata, J. L. Siqueira-Neto, S. L. Reed, J. H. McKerrow, W. R. Roush, *ACS Med. Chem. Lett.* **2015**, *7*, 77-82.

[30] D. F. Vieira, J. Y. Choi, W. R. Roush, L. M. Podust, *ChemBioChem.* **2014**, *15*, 1111-1120.

[31] F. Pavao, M. Castilho, M. Pupo, R. Dias, A. Correa, J. Fernandes, M. Da Silva, J. Mafezoli, P. Vieira, G. Oliva, *FEBS Lett.* **2002**, *520*, 13-17.

[32] N. Schormann, O. Senkovich, K. Walker, D. Wright, A. Anderson, A. Rosowsky, S. Ananthan, B. Shinkre, S. Velu, D. Chattopadhyay, *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 889-901.

[33] S. Buchini, A. Buschiazzo, S. G. Withers, *Angew. Chem. Int. Ed.* **2008**, *47*, 2700-2703.

[34] Y. Amano, I. Namatame, Y. Tateishi, K. Honboh, E. Tanabe, T. Niimi, H. Sakashita, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 1879-1889.

[35] E. A. Merritt, T. L. Arakaki, J. R. Gillespie, E. T. Larson, A. Kelley, N. Mueller, A. J. Napuli, J. Kim, L. Zhang, C. L. Verlinde, *J. Mol. Biol.* **2010**, *397*, 481-494.

[36] B. Canyuk, F. J. Medrano, M. A. Wenck, P. J. Focia, A. E. Eakin, S. P. Craig, *J. Mol. Biol.* **2004**, *335*, 905-921.

[37] K. W. Borrelli, B. Cossins, V. Guallar, *J. Comput. Chem*. **2010**, 31, 1224-1235.

Anexos capítulo IV

Resumo dos acertos (para validação cruzada e teste) obtidos pelo algoritmo *Random forest*, usando valores de pIC_{50} = 4,0; 4,5 e 4,7 como ponto de corte atividade.

Amastigote							
		Cross-validation			Test		
Model		**Samples**	**Match**	**% Match**	**Samples**	**Match**	**% Match**
Cut-off pIC50 ≥ 4.0	Active	3312	3289	**99.3**	827	822	**99.4**
	Inactive	250	86	**34.4**	63	22	**34.9**
	Total	3562	3375	**94.4**	890	844	**94.8**
Cut-off pIC50 ≥ 4.5	Active	2687	2543	**94.6**	671	632	**94.2**
	Inactive	875	438	**50.0**	219	105	**47.9**
	Total	3562	2981	**83.7**	890	737	**82.8**
Cut-off pIC50 ≥ 4.7	Active	2272	1994	**87.7**	568	496	**87.3**
	Inactive	1290	706	**54.7**	322	185	**57.4**
	Total	3562	2700	**75.8**	890	681	**76.5**

Epimastigote							
		Cross-validation			Test		
Model		**Samples**	**Match**	**% Match**	**Samples**	**Match**	**% Match**
Cut-off pIC50 ≥ 4.0	Active	912	889	**0.97**	228	221	**0.97**
	Inactive	189	90	**0.48**	47	25	**0.53**
	Total	1101	979	**0.89**	275	246	**0.89**
Cut-off pIC50 ≥ 4.5	Active	552	472	**0.86**	138	118	**0.86**
	Inactive	493	395	**0.80**	123	95	**0.77**
	Total	1045	867	**0.83**	261	213	**0.82**
Cut-off pIC50 ≥ 4.7	Active	387	261	**0.67**	97	63	**0.65**
	Inactive	662	573	**0.87**	165	137	**0.83**
	Total	1049	834	**0.80**	262	200	**0.76**

Resumo dos acertos obtidos pelo algoritmo *Random forest,* utilizando dos diferentes métodos de partição (aleatório e estratificado) dos dados de treino e teste.

Random split							
		Cross-validation			Test		
Model		**Samples**	**Match**	**% Match**	**Samples**	**Match**	**% Match**
Trypomastigote	Active	599	415	**69.3**	144	101	**70.1**
	Inactive	709	573	**80.8**	183	151	**82.5**
	Total	1308	988	**75.5**	327	252	**77.1**
Amastigote	Active	1583	1468	**92.7**	471	355	**75.4**
	Inactive	1709	1290	**75.5**	419	315	**75.2**
	Total	3562	2758	**77.4**	890	670	**75.3**
Epimastigote	Active	378	246	**65.1**	104	71	**68.3**
	Inactive	680	596	**87.6**	160	138	**86.3**
	Total	1058	842	**79.6**	264	209	**79.2**

Stratified split							
		Cross-validation			Test		
Model		**Samples**	**Match**	**% Match**	**Samples**	**Match**	**% Match**
Amastigotes	Active	1859	1426	**76.7**	465	368	**79.1**
	Inactive	1703	1271	**74.6**	425	318	**74.8**
	Total	3562	2697	**75.7**	890	686	**77.1**
Trypomastigotes	Active	594	418	**70.4**	149	97	**65.1**
	Inactive	714	581	**81.4**	178	150	**84.3**
	Total	1308	999	**76.4**	327	247	**75.5**
Epimastigotes	Active	386	259	**67.1**	96	62	**64.6**
	Inactive	672	575	**85.6**	168	153	**91.1**
	Total	1058	834	**78.8**	264	215	**81.4**

Resultados das predições dos compostos selecionados da literatura utilizados na avaliação final dos modelos *random forest* construídos para as três formas parasitárias de *T. cruzi*

Epimastigote				
SMILES	**pIC$_{50}$**	**Activity**	**Predicted**	***p***
A	4.38	I	I	0.44
B	4.36	I	I	0.39
C	4.72	I	I	0.44
D	4.32	I	A	0.60
E	6.62	A	I	0.29
F	6.11	A	I	0.41
G	5.36	A	I	0.46
H	5.15	A	A	0.78
I	5.50	A	I	0.44
J	5.52	A	I	0.28
K	5.17	A	A	0.60

Amastigote				
SMILES	**pIC$_{50}$**	**Activity**	**Predicted**	***p***
L	4.81	I	I	0.27
M	4.54	I	I	0.28
N	4.64	I	I	0.26
O	3.74	I	I	0.24
P	5.28	A	I	0.13

Trypomastigote				
SMILES	**pIC$_{50}$**	**Activity**	**Predicted**	***p***
L	5.14	A	I	0.29
M	5.99	A	I	0.31
N	5.27	A	I	0.37
O	3.69	I	I	0.15

A [H][C@@]1(C)C2CCC(C)=C3C(=O)C=C(C)[C@]3([H])C2OC1=O

B [H][C@@]12C[C@H]3[C@@H](C[C@@]1(C)[C@@H](CCC2=C)O[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)OC(=O)C3=C

C [H][C@]1(C)C2CCC(C)=C3C(=O)C=C(C)[C@]3([H])C2OC1=O

D [H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@H](C[C@@]1(COC(C)=O)C=C)OC(C)=O)C(=C)C=O

E [H][C@@]12C[C@H]3O[C@@]3(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1CCC2=C

F [H][C@@]12CC=C(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)C(\CO)=C\CO

G C\C1=C/[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\C(C)=C\CC1)OC(=O)C(\CO)=C\CO

H CC(=O)OCC1=C2[C@H](OC1=O)[C@@H]1O[C@]1(C)CCC(=O)[C@@](C)(C[C@@H]2OC(=O)C(C)=C)OC(C)=O

I CC(=O)OC\C1=C/CC\C(CO)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](C1)OC(C)=O

J C\C1=C/CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2CC1

K COC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C/C)[C@H]\1O

L C[C@]12C[C@@H]3OC(=O)C(=C)[C@H]3[C@@H]3OC(=O)C(=C3)[C@H]3O[C@H]3[C@H]1O2

M C[C@@H]1[C@@H]2[C@H](C[C@]3(C)O[C@@H]3[C@@H]3O[C@@H]3C3=C[C@H]2OC3=O)OC1=O

N [H]C1([H])[C@H]2O[C@@]2(C)C[C@@H]2OC(=O)C(=C)[C@H]2[C@@H]2OC(=O)C(=C2)C1([H])[H]

O [H]C1([H])[C@H]2O[C@@]2(C)C[C@@H]2OC(=O)C(=C)[C@H]2[C@@H]2OC(=O)C(=C2)[C@]1([H])OC(C)=O

P [H][C@@]12[C@@H](OC(=O)[C@@]1(O)CO)\C=C(C)/C1=CC(=O)[C@@](C)(C[C@H]2OC(=O)C(C)=C)O1

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do ligante, para as três formas parasitárias de *T. cruzi*.

Tripomastigote	
ID	**p**
13106	0.64
13145	0.62
13134	0.62
13061	0.60
12668	0.57
13132	0.55
13139	0.54
13133	0.53
13027	0.52
12969	0.52
13060	0.52
13032	0.51
13108	0.51
13031	0.51
13407	0.50
12811	0.50
14536	0.50

Amastigote			
ID	**p**	**ID**	**p**
14412	0.58	**12123**	0.51
12716	0.56	**12124**	0.51
12862	0.56	**12173**	0.51
12098	0.55	**12406**	0.51
13016	0.54	**12696**	0.51
13024	0.54	**12811**	0.51
14508	0.53	**13049**	0.51
12126	0.52	**15516**	0.51
13139	0.52	**12094**	0.50
12615	0.52	**12097**	0.50
12697	0.52	**12298**	0.50
15517	0.52	**12460**	0.50
12449	0.51	**12497**	0.50
12646	0.51	**12709**	0.50
13398	0.51	**13005**	0.50
12001	0.51	**14546**	0.50
12119	0.51	**15208**	0.50

Epimastigote							
ID	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**
12399	0.82	**12454**	0.66	**13057**	0.61	**12861**	0.58
12447	0.82	**13107**	0.66	**13117**	0.61	**12883**	0.58
12118	0.81	**13396**	0.66	**13305**	0.61	**12892**	0.58
12449	0.77	**12415**	0.65	**13948**	0.61	**12950**	0.58
13017	0.77	**12498**	0.65	**12068**	0.6	**13047**	0.58
12426	0.76	**12720**	0.65	**12097**	0.6	**13105**	0.58
12446	0.76	**12774**	0.65	**12330**	0.6	**13170**	0.58
12451	0.76	**13059**	0.65	**12413**	0.6	**13186**	0.58
12482	0.76	**14376**	0.65	**12414**	0.6	**13302**	0.58
13061	0.76	**14412**	0.65	**12461**	0.6	**13308**	0.58
12636	0.74	**12037**	0.64	**12603**	0.6	**14268**	0.58
13024	0.74	**12155**	0.64	**12657**	0.6	**14320**	0.58
12098	0.73	**12226**	0.64	**12844**	0.6	**14420**	0.58
12424	0.73	**12331**	0.64	**12871**	0.6	**14529**	0.58
12634	0.73	**12560**	0.64	**12889**	0.6	**14563**	0.58
13006	0.73	**12812**	0.64	**13070**	0.6	**12242**	0.57
12406	0.72	**13037**	0.64	**13111**	0.6	**12365**	0.57
12549	0.72	**13085**	0.64	**13139**	0.6	**12398**	0.57
12896	0.72	**13398**	0.64	**14267**	0.6	**12401**	0.57
13005	0.72	**13399**	0.64	**12042**	0.59	**12428**	0.57
13056	0.72	**14419**	0.64	**12070**	0.59	**12510**	0.57
12119	0.71	**12126**	0.63	**12148**	0.59	**12618**	0.57
12123	0.71	**12146**	0.63	**12197**	0.59	**12648**	0.57
12210	0.71	**12660**	0.63	**12227**	0.59	**12651**	0.57
12716	0.71	**12815**	0.63	**12299**	0.59	**12659**	0.57
12862	0.71	**12880**	0.63	**12467**	0.59	**12721**	0.57
13045	0.71	**13034**	0.63	**12538**	0.59	**12817**	0.57
13088	0.71	**13400**	0.63	**12559**	0.59	**12837**	0.57
12416	0.7	**12020**	0.62	**12658**	0.59	**12860**	0.57
12897	0.7	**12174**	0.62	**12779**	0.59	**12955**	0.57
12907	0.7	**12255**	0.62	**12781**	0.59	**12992**	0.57
13165	0.7	**12262**	0.62	**12797**	0.59	**13007**	0.57
14396	0.7	**12403**	0.62	**13016**	0.59	**13011**	0.57
12918	0.69	**12652**	0.62	**13079**	0.59	**13012**	0.57
13129	0.69	**12874**	0.62	**13309**	0.59	**13058**	0.57
14564	0.69	**12899**	0.62	**14329**	0.59	**13149**	0.57
12099	0.68	**12901**	0.62	**14331**	0.59	**13168**	0.57
12277	0.68	**13062**	0.62	**14450**	0.59	**13947**	0.57
12418	0.68	**14448**	0.62	**14493**	0.59	**14273**	0.57
13082	0.68	**14453**	0.62	**14536**	0.59	**14418**	0.57
13106	0.68	**15516**	0.62	**15517**	0.59	**14500**	0.57
13330	0.68	**12019**	0.61	**12001**	0.58	**1036**	0.56
12086	0.67	**12125**	0.61	**12048**	0.58	**12009**	0.56
12175	0.67	**12147**	0.61	**12079**	0.58	**12010**	0.56
12186	0.67	**12206**	0.61	**12111**	0.58	**12067**	0.56
12196	0.67	**12276**	0.61	**12236**	0.58	**12076**	0.56
12608	0.67	**12366**	0.61	**12468**	0.58	**12163**	0.56
12921	0.67	**12558**	0.61	**12471**	0.58	**12270**	0.56
13027	0.67	**12600**	0.61	**12639**	0.58	**12273**	0.56
14528	0.67	**12625**	0.61	**12655**	0.58	**12338**	0.56
12124	0.66	**12626**	0.61	**12732**	0.58	**12353**	0.56
12170	0.66	**12811**	0.61	**12796**	0.58	**12400**	0.56
12257	0.66	**13049**	0.61				

Epimastigote							
ID	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**
12624	0.56	**12653**	0.54	**12145**	0.52	**12887**	0.51
12654	0.56	**12702**	0.54	**12168**	0.52	**12916**	0.51
12784	0.56	**12762**	0.54	**12198**	0.52	**13010**	0.51
12794	0.56	**12816**	0.54	**12342**	0.52	**13013**	0.51
12847	0.56	**12833**	0.54	**12352**	0.52	**13086**	0.51
12888	0.56	**12850**	0.54	**12370**	0.52	**13096**	0.51
12922	0.56	**12854**	0.54	**12440**	0.52	**13297**	0.51
12940	0.56	**13026**	0.54	**12477**	0.52	**13408**	0.51
12989	0.56	**13041**	0.54	**12521**	0.52	**13954**	0.51
13033	0.56	**13046**	0.54	**12554**	0.52	**14277**	0.51
13051	0.56	**13063**	0.54	**12597**	0.52	**14278**	0.51
13303	0.56	**13097**	0.54	**12606**	0.52	**14341**	0.51
13401	0.56	**14386**	0.54	**12616**	0.52	**14352**	0.51
14537	0.56	**14403**	0.54	**12635**	0.52	**14533**	0.51
12162	0.55	**14515**	0.54	**12641**	0.52	**15513**	0.51
12194	0.55	**14517**	0.54	**12775**	0.52	**12087**	0.5
12289	0.55	**14539**	0.54	**12807**	0.52	**12088**	0.5
12361	0.55	**14548**	0.54	**12848**	0.52	**12093**	0.5
12373	0.55	**12100**	0.53	**12851**	0.52	**12229**	0.5
12419	0.55	**12115**	0.53	**12852**	0.52	**12314**	0.5
12456	0.55	**12131**	0.53	**12858**	0.52	**12318**	0.5
12485	0.55	**12135**	0.53	**12915**	0.52	**12337**	0.5
12573	0.55	**12151**	0.53	**12925**	0.52	**12431**	0.5
12646	0.55	**12160**	0.53	**13000**	0.52	**12437**	0.5
12718	0.55	**12169**	0.53	**13028**	0.52	**12439**	0.5
12795	0.55	**12171**	0.53	**13066**	0.52	**12570**	0.5
12838	0.55	**12203**	0.53	**13102**	0.52	**12591**	0.5
12849	0.55	**12272**	0.53	**13109**	0.52	**12665**	0.5
12881	0.55	**12298**	0.53	**13329**	0.52	**12694**	0.5
13023	0.55	**12368**	0.53	**14218**	0.52	**12741**	0.5
13110	0.55	**12462**	0.53	**14354**	0.52	**12765**	0.5
13157	0.55	**12463**	0.53	**14421**	0.52	**12868**	0.5
13182	0.55	**12472**	0.53	**14518**	0.52	**12877**	0.5
13196	0.55	**12490**	0.53	**14538**	0.52	**12923**	0.5
13205	0.55	**12637**	0.53	**14540**	0.52	**12927**	0.5
13949	0.55	**12722**	0.53	**14542**	0.52	**13002**	0.5
12077	0.54	**12750**	0.53	**14551**	0.52	**13008**	0.5
12105	0.54	**12772**	0.53	**12054**	0.51	**13014**	0.5
12176	0.54	**12840**	0.53	**12085**	0.51	**13040**	0.5
12239	0.54	**12846**	0.53	**12204**	0.51	**13044**	0.5
12256	0.54	**12882**	0.53	**12208**	0.51	**13064**	0.5
12286	0.54	**12891**	0.53	**12263**	0.51	**13137**	0.5
12320	0.54	**12898**	0.53	**12297**	0.51	**13184**	0.5
12367	0.54	**12951**	0.53	**12389**	0.51	**13293**	0.5
12378	0.54	**13029**	0.53	**12397**	0.51	**13413**	0.5
12412	0.54	**13060**	0.53	**12407**	0.51	**14276**	0.5
12417	0.54	**13169**	0.53	**12408**	0.51	**14351**	0.5
12481	0.54	**13206**	0.53	**12420**	0.51	**14496**	0.5
12537	0.54	**13207**	0.53	**12435**	0.51	**14516**	0.5
12568	0.54	**13301**	0.53	**12771**	0.51	**14519**	0.5
12598	0.54	**13950**	0.53	**12845**	0.51	**14522**	0.5
12620	0.54	**14541**	0.53	**12885**	0.51	**15208**	0.5
12638	0.54	**15518**	0.53				

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para oito proteínas alvo de *T. cruzi*

Cruzain: 12 compostos

Energia do ligante: -79,97 KJ/mol

Cruzain		
ID	**E (KJ/mol)**	**p_s**
14466	-91.39	1.00
12881	-85.19	0.93
14467	-84.24	0.92
14465	-83.90	0.92
12068	-83.64	0.92
12860	-83.50	0.91
14296	-83.41	0.91
12634	-83.20	0.91
12139	-82.86	0.91
14414	-81.34	0.89
12884	-80.71	0.88

Sterol 14α-demethylase - CYP51: 3 compostos

Energia do ligante: -106,78 KJ/mol

Sterol 14α-demethylase - CYP51		
ID	**E (KJ/mol)**	**p_s**
12299	-113.58	1.00
12941	-112.52	0.99
12900	-107.51	0.95

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para oito proteínas alvo de *T. cruzi*

Dihydrofolate Reductase-Thymidylate synthase: 7 compostos

Energia do ligante: -73,96 KJ/mol

Dihydrofolate Reductase-Thymidylate synthase		
ID	**E (KJ/mol)**	**p_s**
12498	-82.08	1.00
12897	-81.68	1.00
14529	-79.10	0.96
12882	-77.91	0.95
13331	-76.11	0.93
14564	-74.77	0.91
13179	-74.45	0.91

Hypoxanthine phosphoribosyltransferase: 20 compostos

Energia do ligante: -88,45 KJ/mol

Hypoxanthine phosphoribosyltransferase					
ID	**E (KJ/mol)**	**p_s**	**ID**	**E (KJ/mol)**	**p_s**
12868	-107.54	1.00	**14351**	-93.41	0.87
14414	-105.34	0.98	**12608**	-92.31	0.86
12869	-105.17	0.98	**12626**	-92.23	0.86
12043	-105.01	0.98	**12413**	-91.38	0.85
12042	-101.88	0.95	**14468**	-91.28	0.85
12573	-98.23	0.91	**12867**	-89.82	0.84
12045	-97.45	0.91	**13334**	-89.30	0.83
12887	-95.95	0.89	**12701**	-89.23	0.83
12893	-94.95	0.88	**14453**	-89.00	0.83
14427	-94.19	0.88	**12700**	-88.92	0.83

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para oito proteínas alvo de *T. cruzi*

Glyceraldehyde-3-phosphate dehydrogenase: 379 compostos

Energia do ligante: -26,54 KJ/mol

Glyceraldehyde-3-phosphate dehydrogenase											
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
13399	-94.62	1.00	**13302**	-64.09	0.68	**12785**	-59.64	0.63	**13249**	-57.17	0.60
12570	-78.93	0.83	**13330**	-64.08	0.68	**12850**	-59.55	0.63	**13109**	-57.14	0.60
12705	-78.47	0.83	**12446**	-63.91	0.68	**14380**	-59.51	0.63	**14452**	-57.13	0.60
13401	-77.32	0.82	**12892**	-63.75	0.67	**12882**	-59.45	0.63	**15202**	-57.13	0.60
12365	-76.76	0.81	**13293**	-63.59	0.67	**13331**	-59.42	0.63	**12702**	-57.06	0.60
12901	-75.78	0.80	**12029**	-63.52	0.67	**12463**	-59.42	0.63	**12781**	-57.02	0.60
12861	-75.09	0.79	**12958**	-63.43	0.67	**12098**	-59.35	0.63	**12019**	-56.88	0.60
14564	-74.06	0.78	**12815**	-63.39	0.67	**12699**	-59.33	0.63	**14516**	-56.81	0.60
13308	-73.97	0.78	**14546**	-63.35	0.67	**12458**	-59.29	0.63	**12165**	-56.80	0.60
12900	-73.24	0.77	**13296**	-63.33	0.67	**13116**	-59.17	0.63	**14457**	-56.71	0.60
12899	-72.49	0.77	**12794**	-63.10	0.67	**12167**	-59.15	0.63	**14342**	-56.62	0.60
12898	-72.14	0.76	**12896**	-62.95	0.67	**12604**	-59.07	0.62	**13100**	-56.58	0.60
13024	-72.14	0.76	**14277**	-62.85	0.66	**12295**	-58.97	0.62	**12931**	-56.56	0.60
12860	-71.88	0.76	**12460**	-62.79	0.66	**12308**	-58.96	0.62	**14528**	-56.53	0.60
13298	-71.86	0.76	**14451**	-62.59	0.66	**14421**	-58.96	0.62	**13075**	-56.49	0.60
12307	-71.84	0.76	**12204**	-62.21	0.66	**13138**	-58.94	0.62	**12123**	-56.33	0.60
12893	-71.47	0.76	**12862**	-62.19	0.66	**12037**	-58.84	0.62	**13104**	-56.31	0.60
14418	-71.38	0.75	**12426**	-62.08	0.66	**13305**	-58.75	0.62	**12637**	-56.21	0.59
12927	-71.30	0.75	**14392**	-62.02	0.66	**12955**	-58.69	0.62	**12418**	-56.17	0.59
12874	-71.06	0.75	**12451**	-62.02	0.66	**13300**	-58.68	0.62	**12993**	-56.10	0.59
12212	-70.65	0.75	**12427**	-61.99	0.66	**14450**	-58.66	0.62	**12498**	-56.06	0.59
12125	-69.89	0.74	**13129**	-61.54	0.65	**12640**	-58.59	0.62	**12424**	-56.00	0.59
12897	-69.49	0.73	**12198**	-61.47	0.65	**12043**	-58.50	0.62	**14485**	-55.98	0.59
14412	-69.45	0.73	**15516**	-61.46	0.65	**12847**	-58.44	0.62	**12933**	-55.96	0.59
12859	-68.52	0.72	**12211**	-61.41	0.65	**13195**	-58.43	0.62	**13955**	-55.87	0.59
12431	-68.43	0.72	**12367**	-61.34	0.65	**12939**	-58.37	0.62	**15266**	-55.85	0.59
13309	-68.10	0.72	**13301**	-61.33	0.65	**13016**	-58.25	0.62	**12624**	-55.84	0.59
14352	-67.82	0.72	**15517**	-61.30	0.65	**14484**	-58.08	0.61	**12349**	-55.71	0.59
12200	-67.74	0.72	**12837**	-61.27	0.65	**14413**	-57.98	0.61	**14343**	-55.70	0.59
12366	-66.86	0.71	**12941**	-61.11	0.65	**12921**	-57.93	0.61	**14403**	-55.67	0.59
12848	-66.79	0.71	**14448**	-61.03	0.65	**14267**	-57.92	0.61	**12456**	-55.60	0.59
12077	-66.75	0.71	**12880**	-60.94	0.64	**12438**	-57.91	0.61	**12086**	-55.57	0.59
12793	-66.71	0.71	**14547**	-60.89	0.64	**12419**	-57.88	0.61	**12558**	-55.55	0.59
1036	-66.52	0.70	**14398**	-60.69	0.64	**12726**	-57.74	0.61	**14522**	-55.53	0.59
12797	-66.35	0.70	**12795**	-60.67	0.64	**12866**	-57.72	0.61	**13063**	-55.50	0.59
12940	-66.34	0.70	**12538**	-60.65	0.64	**12701**	-57.59	0.61	**12791**	-55.30	0.58
12858	-66.33	0.70	**12608**	-60.48	0.64	**12673**	-57.58	0.61	**14551**	-55.30	0.58
14563	-66.23	0.70	**12088**	-60.40	0.64	**12406**	-57.56	0.61	**13297**	-55.30	0.58
12889	-65.73	0.69	**12471**	-60.15	0.64	**13017**	-57.55	0.61	**13153**	-55.26	0.58
13294	-65.64	0.69	**12462**	-60.14	0.64	**12512**	-57.45	0.61	**12745**	-55.21	0.58
12572	-65.60	0.69	**13348**	-60.04	0.63	**12863**	-57.42	0.61	**12700**	-55.11	0.58
12796	-65.49	0.69	**12010**	-60.02	0.63	**12890**	-57.42	0.61	**12600**	-55.09	0.58
13311	-65.44	0.69	**13286**	-59.97	0.63	**12665**	-57.39	0.61	**12012**	-54.97	0.58
12884	-65.15	0.69	**12297**	-59.91	0.63	**13280**	-57.27	0.61	**13303**	-54.95	0.58
12303	-65.09	0.69	**12093**	-59.90	0.63	**12090**	-57.25	0.61	**12299**	-54.91	0.58
14270	-64.91	0.69	**13963**	-59.82	0.63	**12869**	-57.20	0.60	**13196**	-54.79	0.58
12399	-64.65	0.68	**12851**	-59.73	0.63	**12489**	-57.18	0.60	**12780**	-54.77	0.58
12117	-64.52	0.68	**12309**	-59.66	0.63						

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para oito proteínas alvo de *T. cruzi*

Glyceraldehyde-3-phosphate dehydrogenase: 379 compostos

Energia do ligante: -26,54 KJ/mol

Glyceraldehyde-3-phosphate dehydrogenase											
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
12580	-54.70	0.58	**14466**	-52.82	0.56	**12506**	-51.33	0.54	**12992**	-49.25	0.52
12618	-54.67	0.58	**12830**	-52.67	0.56	**13321**	-51.31	0.54	**12092**	-49.20	0.52
13299	-54.62	0.58	**13374**	-52.64	0.56	**12163**	-51.19	0.54	**12048**	-49.18	0.52
12301	-54.62	0.58	**12857**	-52.63	0.56	**13134**	-51.14	0.54	**13008**	-49.09	0.52
12397	-54.56	0.58	**14376**	-52.62	0.56	**12655**	-51.07	0.54	**12437**	-49.09	0.52
12750	-54.52	0.58	**14331**	-52.58	0.56	**12731**	-51.00	0.54	**13095**	-49.01	0.52
12296	-54.50	0.58	**14386**	-52.52	0.56	**12766**	-50.87	0.54	**12173**	-49.00	0.52
14329	-54.48	0.58	**12206**	-52.51	0.55	**12209**	-50.82	0.54	**13135**	-48.97	0.52
12639	-54.47	0.58	**12746**	-52.51	0.55	**12635**	-50.75	0.54	**13954**	-48.94	0.52
12710	-54.45	0.58	**12854**	-52.51	0.55	**12421**	-50.67	0.54	**12905**	-48.85	0.52
12799	-54.43	0.58	**14529**	-52.50	0.55	**15339**	-50.64	0.54	**12662**	-48.75	0.52
14517	-54.28	0.57	**14303**	-52.42	0.55	**12053**	-50.62	0.53	**14296**	-48.72	0.51
14477	-54.27	0.57	**12099**	-52.36	0.55	**13953**	-50.60	0.53	**12030**	-48.69	0.51
12517	-54.21	0.57	**12970**	-52.36	0.55	**12467**	-50.54	0.53	**14526**	-48.66	0.51
12008	-54.20	0.57	**13112**	-52.33	0.55	**13113**	-50.45	0.53	**12914**	-48.64	0.51
12468	-54.15	0.57	**13413**	-52.28	0.55	**12634**	-50.41	0.53	**12605**	-48.57	0.51
14218	-54.09	0.57	**12747**	-52.28	0.55	**12305**	-50.39	0.53	**12348**	-48.56	0.51
12433	-54.03	0.57	**13400**	-52.26	0.55	**13051**	-50.23	0.53	**13154**	-48.54	0.51
13205	-54.03	0.57	**13398**	-52.22	0.55	**13207**	-50.15	0.53	**13343**	-48.51	0.51
12716	-54.03	0.57	**12403**	-52.22	0.55	**13203**	-50.13	0.53	**13324**	-48.48	0.51
12449	-54.02	0.57	**13410**	-52.21	0.55	**13088**	-50.12	0.53	**12968**	-48.48	0.51
12528	-54.00	0.57	**12909**	-52.20	0.55	**12124**	-50.10	0.53	**12911**	-48.31	0.51
14279	-53.98	0.57	**12894**	-52.20	0.55	**13318**	-50.10	0.53	**12115**	-48.26	0.51
13319	-53.98	0.57	**14409**	-52.13	0.55	**12155**	-50.05	0.53	**12118**	-48.18	0.51
12481	-53.95	0.57	**12537**	-52.11	0.55	**12455**	-50.00	0.53	**12126**	-48.17	0.51
12636	-53.92	0.57	**14394**	-52.04	0.55	**13102**	-49.92	0.53	**14382**	-48.15	0.51
12461	-53.89	0.57	**12422**	-52.01	0.55	**14468**	-49.88	0.53	**12910**	-48.11	0.51
14552	-53.81	0.57	**12779**	-52.00	0.55	**13202**	-49.85	0.53	**12164**	-48.04	0.51
12162	-53.70	0.57	**12432**	-51.94	0.55	**13137**	-49.81	0.53	**12425**	-48.01	0.51
12306	-53.66	0.57	**12852**	-51.93	0.55	**12529**	-49.78	0.53	**14453**	-48.00	0.51
12515	-53.55	0.57	**14319**	-51.89	0.55	**12011**	-49.75	0.53	**12614**	-47.96	0.51
12071	-53.52	0.57	**12108**	-51.86	0.55	**13114**	-49.75	0.53	**12720**	-47.95	0.51
13053	-53.46	0.57	**12497**	-51.79	0.55	**13968**	-49.74	0.53	**12171**	-47.94	0.51
13408	-53.45	0.56	**12841**	-51.78	0.55	**12329**	-49.66	0.52	**12707**	-47.93	0.51
14268	-53.30	0.56	**12513**	-51.74	0.55	**12304**	-49.65	0.52	**13333**	-47.88	0.51
12095	-53.23	0.56	**12412**	-51.73	0.55	**12194**	-49.63	0.52	**12464**	-47.85	0.51
12929	-53.21	0.56	**12069**	-51.72	0.55	**12051**	-49.62	0.52	**13190**	-47.85	0.51
12496	-53.21	0.56	**12017**	-51.70	0.55	**12846**	-49.58	0.52	**12166**	-47.83	0.51
15518	-53.14	0.56	**14553**	-51.68	0.55	**12579**	-49.58	0.52	**13281**	-47.82	0.51
13011	-53.13	0.56	**12952**	-51.66	0.55	**12402**	-49.58	0.52	**13032**	-47.81	0.51
12704	-53.10	0.56	**13198**	-51.58	0.55	**13142**	-49.58	0.52	**12097**	-47.69	0.50
12871	-53.08	0.56	**13108**	-51.56	0.54	**13005**	-49.54	0.52	**12510**	-47.63	0.50
12368	-53.04	0.56	**12441**	-51.52	0.54	**12744**	-49.51	0.52	**12670**	-47.52	0.50
12812	-53.02	0.56	**14393**	-51.46	0.54	**13018**	-49.48	0.52	**14428**	-47.47	0.50
13204	-52.98	0.56	**12576**	-51.45	0.54	**12849**	-49.37	0.52	**12826**	-47.41	0.50
12786	-52.92	0.56	**12159**	-51.36	0.54	**12420**	-49.34	0.52	**13152**	-47.34	0.50
13285	-52.90	0.56	**12615**	-51.35	0.54	**12675**	-49.31	0.52	**12298**	-47.33	0.50
14533	-52.90	0.56									

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para oito proteínas alvo de *T. cruzi*

Trans-sialidase: 211 compostos

Energia do ligante: -26,90 KJ/mol

Trans-sialidase											
ID	**E (KJ/mol)**	**p_s**	**ID**	**E (KJ/mol)**	**p_s**	**ID**	**E (KJ/mol)**	**p_s**	**ID**	**E (KJ/mol)**	**p_s**
12882	-103.22	1.00	**12526**	-70.82	0.69	**12498**	-62.98	0.61	**12198**	-56.61	0.55
12456	-99.84	0.97	**14450**	-70.78	0.69	**13411**	-62.90	0.61	**12830**	-56.60	0.55
12197	-94.33	0.91	**12901**	-70.75	0.69	**14417**	-62.84	0.61	**13142**	-56.57	0.55
12204	-93.55	0.91	**13102**	-70.62	0.68	**15339**	-62.72	0.61	**13142**	-56.57	0.55
12076	-92.71	0.90	**13011**	-70.55	0.68	**12858**	-62.66	0.61	**12012**	-56.28	0.55
12893	-91.16	0.88	**12524**	-69.63	0.67	**12798**	-62.55	0.61	**13175**	-56.23	0.54
14453	-90.67	0.88	**13155**	-69.26	0.67	**12200**	-62.50	0.61	**13001**	-56.10	0.54
12797	-90.09	0.87	**12874**	-69.03	0.67	**13348**	-62.43	0.60	**12201**	-56.04	0.54
12489	-86.17	0.83	**12411**	-68.80	0.67	**12972**	-62.42	0.60	**12808**	-55.92	0.54
12207	-86.05	0.83	**12511**	-68.60	0.66	**12208**	-62.24	0.60	**14479**	-55.88	0.54
12417	-84.41	0.82	**12407**	-68.50	0.66	**14394**	-61.99	0.60	**12347**	-55.82	0.54
13104	-82.92	0.80	**14320**	-68.42	0.66	**12163**	-61.68	0.60	**12911**	-55.76	0.54
12665	-81.47	0.79	**13012**	-68.20	0.66	**13152**	-61.63	0.60	**14393**	-55.64	0.54
14386	-81.45	0.79	**12705**	-67.99	0.66	**14382**	-61.27	0.59	**12679**	-55.52	0.54
13070	-80.84	0.78	**12029**	-67.95	0.66	**13092**	-61.24	0.59	**14477**	-55.43	0.54
13063	-79.77	0.77	**12900**	-67.78	0.66	**14350**	-60.97	0.59	**12114**	-55.27	0.54
12042	-79.32	0.77	**13008**	-67.77	0.66	**12227**	-60.79	0.59	**14524**	-54.98	0.53
13335	-79.05	0.77	**14273**	-67.68	0.66	**14451**	-60.79	0.59	**12906**	-54.83	0.53
12418	-79.00	0.77	**12298**	-67.13	0.65	**12295**	-60.79	0.59	**15208**	-54.81	0.53
12572	-78.41	0.76	**12600**	-67.12	0.65	**12666**	-60.76	0.59	**13007**	-54.71	0.53
12856	-78.17	0.76	**14467**	-67.08	0.65	**12973**	-60.61	0.59	**13953**	-54.67	0.53
12897	-78.12	0.76	**12019**	-67.08	0.65	**12793**	-60.55	0.59	**13145**	-54.65	0.53
12167	-77.86	0.75	**12234**	-66.98	0.65	**13244**	-60.51	0.59	**12021**	-54.53	0.53
12795	-77.67	0.75	**12194**	-66.97	0.65	**14533**	-60.44	0.59	**13002**	-54.36	0.53
12075	-76.91	0.75	**12504**	-66.50	0.64	**13101**	-60.13	0.58	**12778**	-54.26	0.53
14329	-76.60	0.74	**12077**	-66.38	0.64	**14473**	-59.98	0.58	**13324**	-54.24	0.53
13107	-76.42	0.74	**15202**	-66.37	0.64	**12165**	-59.81	0.58	**13334**	-54.20	0.53
13095	-76.27	0.74	**12781**	-66.17	0.64	**13303**	-59.79	0.58	**12785**	-53.97	0.52
13103	-76.11	0.74	**13333**	-66.05	0.64	**12780**	-59.59	0.58	**13076**	-53.93	0.52
12529	-76.02	0.74	**14448**	-66.04	0.64	**12291**	-59.58	0.58	**12662**	-53.89	0.52
13112	-75.94	0.74	**12050**	-65.49	0.63	**13407**	-59.43	0.58	**12667**	-53.69	0.52
12124	-75.84	0.73	**12615**	-65.45	0.63	**12248**	-59.00	0.57	**13149**	-53.68	0.52
12349	-75.11	0.73	**14528**	-65.33	0.63	**12779**	-58.96	0.57	**14331**	-53.58	0.52
12860	-74.59	0.72	**12306**	-65.32	0.63	**12125**	-58.85	0.57	**12852**	-53.43	0.52
12861	-74.15	0.72	**12573**	-65.26	0.63	**12159**	-58.68	0.57	**12008**	-53.32	0.52
12933	-73.95	0.72	**12401**	-65.20	0.63	**12327**	-58.68	0.57	**12570**	-53.15	0.51
12398	-73.71	0.71	**12203**	-65.07	0.63	**14296**	-58.66	0.57	**12328**	-53.05	0.51
15516	-73.56	0.71	**13154**	-64.96	0.63	**15338**	-58.55	0.57	**12093**	-52.96	0.51
12909	-73.33	0.71	**12028**	-64.91	0.63	**12624**	-58.31	0.56	**12940**	-52.81	0.51
12348	-73.18	0.71	**13153**	-64.87	0.63	**12421**	-58.27	0.56	**12812**	-52.75	0.51
12604	-73.16	0.71	**12027**	-64.66	0.63	**14342**	-58.10	0.56	**14277**	-52.67	0.51
15517	-72.78	0.71	**12206**	-64.58	0.63	**12538**	-58.07	0.56	**12559**	-52.65	0.51
12002	-72.53	0.70	**12128**	-64.49	0.62	**12447**	-57.93	0.56	**14452**	-52.62	0.51
13129	-72.46	0.70	**12435**	-64.47	0.62	**12694**	-57.88	0.56	**12687**	-52.56	0.51
12528	-72.33	0.70	**12692**	-64.32	0.62	**12001**	-57.88	0.56	**15210**	-52.55	0.51
12891	-72.11	0.70	**12083**	-64.28	0.62	**12715**	-57.61	0.56	**12299**	-52.50	0.51
12164	-71.98	0.70	**12346**	-64.25	0.62	**12482**	-57.57	0.56	**14459**	-52.46	0.51
14352	-71.98	0.70	**12794**	-64.22	0.62	**12745**	-57.49	0.56	**12043**	-52.11	0.50
12957	-71.68	0.69	**12211**	-63.27	0.61	**13954**	-57.40	0.56	**13321**	-52.10	0.50
14418	-71.61	0.69	**14531**	-63.26	0.61	**13296**	-57.31	0.56	**12195**	-51.85	0.50
14469	-71.54	0.69	**14530**	-63.18	0.61	**12000**	-57.28	0.55	**12968**	-51.75	0.50
12894	-71.54	0.69	**12757**	-63.16	0.61	**12734**	-57.04	0.55	**13110**	-51.74	0.50
14319	-71.01	0.69	**12689**	-63.07	0.61	**12217**	-56.66	0.55			

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **tripomastigotas** de *T. cruzi*

Cruzain			
ID	p	p_s	p_c
13145	**0.62**	**0.64**	**0.63**
13080	0.47	0.85	0.61
13407	**0.50**	**0.74**	**0.59**
12948	0.47	0.79	0.59
12949	0.42	0.86	0.58
13134	0.62	0.48	0.57
13132	**0.55**	**0.59**	**0.56**
12833	0.43	0.79	0.56
13108	**0.51**	**0.63**	**0.55**
14278	0.43	0.78	0.55
12634	0.35	0.91	0.55
12668	**0.57**	**0.50**	**0.54**
13413	0.43	0.75	0.54
13106	0.64	0.36	0.54
13031	**0.51**	**0.57**	**0.53**
14319	0.42	0.71	0.52
12968	0.42	0.70	0.52
12947	0.45	0.65	0.52
14536	**0.50**	**0.54**	**0.51**
12835	0.40	0.71	0.51
13185	0.38	0.74	0.51
12979	0.48	0.54	0.50
13405	0.44	0.61	0.50

Sterol 14α-demethylase - CYP51			
ID	p	p_s	p_c
13134	**0.62**	**0.84**	**0.70**
13132	**0.55**	**0.61**	**0.57**
13031	**0.51**	**0.67**	**0.57**
13133	**0.53**	**0.63**	**0.57**
13032	**0.51**	**0.64**	**0.56**
12969	**0.52**	**0.61**	**0.55**
13139	**0.54**	**0.55**	**0.54**
13027	**0.52**	**0.59**	**0.54**
12915	0.46	0.69	0.54
13060	**0.52**	**0.59**	**0.54**
13407	**0.50**	**0.61**	**0.54**
12781	0.44	0.73	0.54
13061	0.60	0.43	0.54
13106	0.64	0.34	0.54
14536	**0.50**	**0.60**	**0.53**
13145	0.62	0.37	0.53
13080	0.47	0.65	0.53
13342	0.41	0.76	0.53
13044	0.44	0.66	0.52
12979	0.48	0.58	0.51
12968	0.42	0.68	0.51
12088	0.39	0.74	0.51
12668	0.57	0.39	0.50
12949	0.42	0.65	0.50
13396	0.47	0.56	0.50

Glyceraldehyde-3-phosphate dehydrogenase			
ID	p	p_s	p_c
13134	**0.62**	**0.54**	**0.59**
13145	0.62	0.49	0.57
13106	0.64	0.39	0.55
13108	**0.51**	**0.54**	**0.53**
13032	**0.51**	**0.51**	**0.51**
13061	0.60	0.35	0.51

Histidyl-tRNA synthetase			
ID	p	p_s	p_c
13106	**0.64**	**0.63**	**0.64**
13108	**0.51**	**0.57**	**0.53**
13146	0.46	0.65	0.53
13407	**0.50**	**0.56**	**0.53**
13092	0.39	0.75	0.52

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **tripomastigotas** de *T. cruzi*

Dihydrofolate Reductase-Thymidylate synthase			
ID	p	p_s	p_c
13048	0.44	0.87	0.59
13031	**0.51**	**0.72**	**0.59**
14536	**0.50**	**0.68**	**0.56**
13060	**0.52**	**0.64**	**0.56**
13049	0.40	0.84	0.56
13407	**0.50**	**0.66**	**0.56**
13044	0.44	0.75	0.55
12305	0.42	0.77	0.55
13183	0.39	0.81	0.54
13185	0.38	0.82	0.54
12757	0.42	0.72	0.53
12912	0.40	0.75	0.52
13047	0.31	0.89	0.52
13182	0.40	0.73	0.52
12599	0.42	0.70	0.52
12969	**0.52**	**0.50**	**0.51**
13179	0.29	0.91	0.51
12970	0.38	0.72	0.50

Hypoxanthine phosphoribosyltransferase			
ID	p	p_s	p_c
13134	**0.62**	**0.57**	**0.60**
13061	**0.60**	**0.52**	**0.57**
12979	0.48	0.72	0.56
13080	0.47	0.73	0.56
13407	**0.50**	**0.66**	**0.56**
13106	0.64	0.39	0.55
14301	0.40	0.79	0.54
12910	0.43	0.73	0.54
13133	**0.53**	**0.54**	**0.53**
13108	**0.51**	**0.56**	**0.53**
13145	0.62	0.36	0.53
13132	**0.55**	**0.50**	**0.53**
13405	0.44	0.68	0.53
13031	**0.51**	**0.55**	**0.53**
13146	0.46	0.63	0.52
13096	0.47	0.60	0.52
12668	0.57	0.41	0.51
14342	0.43	0.65	0.51
13032	**0.51**	**0.50**	**0.51**
13183	0.39	0.71	0.50
13139	0.54	0.44	0.50

Trans- sialidase			
ID	p	p_s	p_c
13145	**0.62**	**0.53**	**0.59**
12856	0.41	0.76	0.53
13407	**0.50**	**0.58**	**0.53**
13061	0.60	0.40	0.52
13095	0.39	0.74	0.52
12668	0.57	0.42	0.52
14319	0.42	0.69	0.52
13133	0.53	0.48	0.51
12781	0.44	0.64	0.51

Spermidine synthase			
ID	p	p_s	p_c
12969	**0.52**	**0.61**	**0.55**
14302	0.40	0.82	0.55
13407	**0.50**	**0.61**	**0.54**
13952	0.33	0.91	0.54
12968	0.42	0.73	0.53
13370	0.33	0.89	0.53
12967	0.31	0.92	0.52
12392	0.26	0.98	0.52

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *T. cruzi*

Cruzain			
ID	p	p_s	p_c
14412	0.58	0.83	0.67
12862	0.56	0.81	0.65
12126	0.52	0.73	0.60
13024	0.54	0.70	0.60
15208	0.50	0.75	0.59
12709	0.50	0.75	0.59
13005	0.50	0.73	0.58
12098	0.55	0.65	0.58
15516	0.51	0.71	0.58
12119	0.51	0.70	0.57
12449	0.51	0.69	0.57
14508	0.53	0.64	0.57
15517	0.52	0.67	0.57
12460	0.50	0.68	0.56
13016	0.54	0.61	0.56
12716	0.56	0.52	0.55
12097	0.50	0.62	0.54
12697	0.52	0.57	0.53
12696	0.51	0.58	0.53
12124	0.51	0.56	0.52
12298	0.50	0.54	0.52
12497	0.50	0.53	0.51
12173	0.51	0.52	0.51

Sterol 14α-demethylase - CYP51			
ID	p	p_s	p_c
15516	0.51	0.87	0.64
15517	0.52	0.82	0.63
12716	0.56	0.65	0.59
12123	0.51	0.74	0.59
12119	0.51	0.73	0.59
12126	0.52	0.66	0.57
12460	0.50	0.64	0.55
12124	0.51	0.63	0.55
14412	0.58	0.48	0.55
12001	0.51	0.61	0.54
12449	0.51	0.60	0.54
13024	0.54	0.55	0.54
13139	0.52	0.55	0.53
12497	0.50	0.58	0.53
13049	0.51	0.56	0.53
12173	0.51	0.56	0.52
13005	0.50	0.57	0.52
13016	0.54	0.50	0.52
12646	0.51	0.54	0.52
12697	0.52	0.53	0.52
12298	0.50	0.55	0.52
12098	0.55	0.46	0.51
14546	0.50	0.53	0.51
12709	0.50	0.53	0.51
15208	0.50	0.53	0.51
12862	0.56	0.40	0.50
12406	0.51	0.50	0.50
12094	0.50	0.50	0.50

Hypoxanthine phosphoribosyltransferase			
ID	p	p_s	p_c
12497	0.50	0.56	0.54

Histidyl-tRNA synthetase			
ID	p	p_s	p_c
12497	0.50	0.77	0.60
15208	0.50	0.68	0.57

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *T. cruzi*

Dihydrofolate Reductase-Thymidylate synthase			
ID	p	p_s	p_c
14412	0.58	0.71	0.63
13049	0.51	0.84	0.63
15208	0.50	0.71	0.58
12126	0.52	0.62	0.56
14546	0.50	0.63	0.55
14508	0.53	0.56	0.54
12001	0.51	0.60	0.54
12449	0.51	0.57	0.53
12696	0.51	0.55	0.52
12497	0.50	0.52	0.51

Trans-sialidase			
ID	p	p_s	p_c
12124	0.51	0.73	0.59
15517	0.52	0.71	0.58
15516	0.51	0.71	0.58
12615	0.52	0.63	0.56
12298	0.50	0.65	0.55
14412	0.58	0.50	0.55
12001	0.51	0.56	0.53
15208	0.50	0.53	0.51

Glyceraldehyde-3-phosphate dehydrogenase			
ID	p	p_s	p_c
14412	0.58	0.73	0.64
13024	0.54	0.76	0.62
12862	0.56	0.66	0.60
12098	0.55	0.63	0.57
13016	0.54	0.62	0.56
12716	0.56	0.57	0.56
15517	0.52	0.65	0.56
14546	0.50	0.67	0.56
12460	0.50	0.66	0.56
15516	0.51	0.65	0.56
12406	0.51	0.61	0.54
12123	0.51	0.60	0.54
12449	0.51	0.57	0.53
13398	0.51	0.55	0.53
12615	0.52	0.54	0.53
12497	0.50	0.55	0.52
12126	0.52	0.51	0.52
12124	0.51	0.53	0.51
12173	0.51	0.52	0.51
13005	0.50	0.52	0.51
12119	0.51	0.50	0.50
12097	0.50	0.50	0.50
12298	0.50	0.50	0.50

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **epimastigotas** de *T. cruzi*

Cruzain			
ID	p	p_s	p_c
12634	0.73	0.91	0.79
12897	0.70	0.83	0.75
12862	0.71	0.81	0.74
12449	0.77	0.69	0.74
12426	0.76	0.67	0.73
13024	0.74	0.70	0.73
13005	0.72	0.73	0.72
12636	0.74	0.69	0.72
14412	0.65	0.83	0.71
12068	0.60	0.92	0.71
12119	0.71	0.70	0.71
12098	0.73	0.65	0.70
12549	0.72	0.66	0.70
12880	0.63	0.83	0.70
13088	0.71	0.66	0.69
12860	0.57	0.91	0.69
12881	0.55	0.93	0.68
12403	0.62	0.80	0.68
12720	0.65	0.72	0.68
12424	0.73	0.58	0.68

Sterol 14α-demethylase - CYP51			
ID	p	p_s	p_c
12118	0.81	0.78	0.80
12447	0.82	0.69	0.78
12299	0.59	1.00	0.73
12415	0.65	0.88	0.73
12123	0.71	0.74	0.72
12119	0.71	0.73	0.72
12897	0.70	0.74	0.71
12449	0.77	0.60	0.71
15516	0.62	0.87	0.71
13056	0.72	0.66	0.70
12815	0.63	0.81	0.69
12716	0.71	0.65	0.69
12901	0.62	0.80	0.68
12907	0.70	0.63	0.68
14528	0.67	0.69	0.68
13024	0.74	0.55	0.67
12125	0.61	0.79	0.67
12874	0.62	0.77	0.67
12871	0.60	0.81	0.67
15517	0.59	0.82	0.67

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **epimastigotas** de *T. cruzi*

Dihydrofolate Reductase-Thymidylate synthase			
ID	p	p_s	p_c
12897	0.70	1.00	0.80
12498	0.65	1.00	0.77
14564	0.69	0.91	0.77
12549	0.72	0.76	0.73
13330	0.68	0.83	0.73
14529	0.58	0.96	0.71
13045	0.71	0.70	0.71
12449	0.77	0.57	0.70
13056	0.72	0.63	0.69
13049	0.61	0.84	0.69
13047	0.58	0.89	0.69
12482	0.76	0.55	0.69
12634	0.73	0.58	0.68
12882	0.53	0.95	0.68
14412	0.65	0.71	0.67
12603	0.60	0.81	0.67
12815	0.63	0.74	0.67
14267	0.60	0.79	0.66
12844	0.60	0.78	0.66
12901	0.62	0.74	0.66

Glyceraldehyde-3-phosphate dehydrogenase			
ID	p	p_s	p_c
12399	0.82	0.68	0.77
13399	0.64	1.00	0.77
13024	0.74	0.76	0.75
12446	0.76	0.68	0.73
12426	0.76	0.66	0.72
12451	0.76	0.66	0.72
14564	0.69	0.78	0.72
13017	0.77	0.61	0.71
12897	0.70	0.73	0.71
12118	0.81	0.51	0.70
12896	0.72	0.67	0.70
12449	0.77	0.57	0.70
12098	0.73	0.63	0.69
12862	0.71	0.66	0.69
12901	0.62	0.80	0.68
12424	0.73	0.59	0.68
12406	0.72	0.61	0.68
12636	0.74	0.57	0.68
14412	0.65	0.73	0.68
13330	0.68	0.68	0.68

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para epimastigotas de *T. cruzi*

Hypoxanthine phosphoribosyltransferase			
ID	p	p_s	p_c
12608	0.67	0.86	0.74
12416	0.70	0.76	0.72
12042	0.59	0.95	0.71
12449	0.77	0.61	0.71
12896	0.72	0.69	0.71
12098	0.73	0.66	0.71
12626	0.61	0.86	0.70
14453	0.62	0.83	0.69
13006	0.73	0.61	0.69
12413	0.60	0.85	0.69
12874	0.62	0.81	0.69
12482	0.76	0.55	0.69
12716	0.71	0.64	0.69
12257	0.66	0.73	0.68
12636	0.74	0.56	0.68
12573	0.55	0.91	0.68
13061	0.76	0.52	0.68
12868	0.50	1.00	0.67
12126	0.63	0.75	0.67
13005	0.72	0.58	0.67
12418	0.68	0.65	0.67

Trans-sialidase			
ID	p	p_s	p_c
12447	0.82	0.56	0.73
12897	0.70	0.76	0.72
14453	0.62	0.88	0.71
12418	0.68	0.77	0.71
12197	0.59	0.91	0.70
12456	0.55	0.97	0.70
12882	0.53	1.00	0.69
13129	0.69	0.70	0.69
12482	0.76	0.56	0.69
12797	0.59	0.87	0.69
13107	0.66	0.74	0.69
12124	0.66	0.73	0.69
12076	0.56	0.90	0.68
13070	0.60	0.78	0.66
14528	0.67	0.63	0.66
15516	0.62	0.71	0.65
12042	0.59	0.77	0.65
12204	0.51	0.91	0.65
14329	0.59	0.74	0.64
12901	0.62	0.69	0.64

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para epimastigotas de *T. cruzi*

Histidyl-tRNA synthetase			
ID	p	p_s	p_c
13106	0.68	0.63	0.66
12634	0.73	0.50	0.65
12019	0.61	0.59	0.60
14352	0.51	0.76	0.60
12020	0.62	0.52	0.59
12242	0.57	0.59	0.58
13408	0.51	0.68	0.57
13186	0.58	0.55	0.57
12521	0.52	0.65	0.57
15208	0.50	0.68	0.56
12992	0.57	0.52	0.55
12606	0.52	0.59	0.55
12472	0.53	0.56	0.54
13033	0.56	0.50	0.54
13413	0.50	0.59	0.53
13028	0.52	0.50	0.51

Spermidine synthase			
ID	p	p_s	p_c
13949	0.55	0.72	0.61
14267	0.60	0.63	0.61
12352	0.52	0.55	0.53
13413	0.50	0.57	0.52

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de duas triagens virtuais baseada na estrutura do ligante. p_z = probabilidade de ser ativo no modelo RF para cruzaina. p = probabilidade de ser ativo no modelo RF para a forma parasitária.

Epimastigote					
ID	**Activity z**	**Pz**	**Activity**	**P**	**PLB2**
12416	A	0.61	A	0.70	0.66
12415	A	0.59	A	0.65	0.62
13399	A	0.59	A	0.64	0.62
13400	A	0.56	A	0.63	0.60
14419	A	0.55	A	0.64	0.60
12720	A	0.53	A	0.65	0.59
12603	A	0.56	A	0.60	0.58
12883	A	0.57	A	0.58	0.58
12226	A	0.50	A	0.64	0.58
12874	A	0.52	A	0.62	0.57
12880	A	0.51	A	0.63	0.57
12111	A	0.56	A	0.58	0.57
12413	A	0.53	A	0.60	0.57
14418	A	0.56	A	0.57	0.57
13948	A	0.52	A	0.61	0.57
14420	A	0.55	A	0.58	0.56
12950	A	0.54	A	0.58	0.56
13401	A	0.56	A	0.56	0.56
12299	A	0.53	A	0.59	0.56
13947	A	0.54	A	0.57	0.56
14563	A	0.53	A	0.58	0.55
12570	A	0.59	A	0.50	0.54
12860	A	0.50	A	0.57	0.54
13205	A	0.53	A	0.55	0.54
13206	A	0.55	A	0.53	0.54
12848	A	0.54	A	0.52	0.53
14386	A	0.52	A	0.54	0.53
12868	A	0.51	A	0.50	0.51

Amastigote					
ID	**Activity**	**Pz**	**Activity**	**P**	**PLB2**
12615	I	0.49	A	0.52	0.50

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de três metodologias de triagem virtual (2 baseadas na estrutura do ligante, uma baseada na estrutura do receptor). p_z = probabilidade de ser ativo no modelo RF para cruzaina.

Amastigote								
ID	**Activity**	**Pz**	**Erank**	**Activity**	**Ps**	**Activity**	**P**	**Pc**
14412	I	0.38	-75.46	A	0.83	A	0.58	**0.56**
12862	I	0.40	-73.89	A	0.81	A	0.56	**0.55**
12460	I	0.42	-62.03	A	0.68	A	0.50	**0.51**
13024	I	0.36	-64.38	A	0.70	A	0.54	**0.50**

Epimastigotae								
ID	**Activity**	**Pz**	**Erank**	**Activity**	**Ps**	**Activity**	**P**	**Pc**
12880	**A**	**0.57**	**-75.86**	**A**	**0.83**	**A**	**0.58**	**0.63**
12897	**A**	**0.56**	**-76.01**	**A**	**0.83**	**A**	**0.63**	**0.65**
12720	**A**	**0.51**	**-66.22**	**A**	**0.72**	**A**	**0.50**	**0.55**
12860	**A**	**0.53**	**-83.50**	**A**	**0.91**	**A**	**0.55**	**0.62**
12874	**A**	**0.59**	**-70.38**	**A**	**0.77**	**A**	**0.50**	**0.59**
12862	**A**	**0.54**	**-73.89**	**A**	**0.81**	**A**	**0.58**	**0.62**
12634	A	0.63	-83.20	A	0.91	I	0.42	0.61
12883	A	0.54	-63.91	A	0.70	I	0.48	0.55
14414	A	0.55	-81.34	A	0.89	I	0.43	0.58
13309	A	0.53	-72.86	A	0.80	I	0.47	0.56
14529	A	0.51	-74.27	A	0.81	I	0.39	0.53
12449	A	0.53	-62.61	A	0.69	I	0.37	0.50

Valores de RMSD determinados entre a pose do Docking e a pose minimizada por dinâmica molecular para os sesquiterpenos lactonizados melhor ranqueados na aproximação combinada de metodologias de triagem virtual.

	Amastigote		Trypomastigote		Epimastigote	
Protein	**Structure**	**RMSD**	**Structure**	**RMSD**	**Structure**	**RMSD**
Cruzain	**1**	6.84	**7**	3.57	**32**	3.19
Sterol 14α-demethylase CYP51	**26**	5.07	**8**	-	**13**	5.98
Dihydrofolate Reductase-Thymidylate synthase	**1**	-	**30**	5.93	**33**	3.51
glyceraldehyde-3-phosphate dehydrogenase	**1**	3.85	**8**	-	**11**	3.83
Histidyl-tRNA synthetase	**27**	1.97	**6**	-	**6**	-
Hypoxanthine phosphoribosyltransferases	**27**	-	**8**	7.05	**34**	4.58
Spermidine synthase	**28**	1.79	**31**	-	**35**	1.38
Trans-sialidase	**29**	5.57	**7**	6.42	**12**	4.53

CAPÍTULO V

A leishmaniose é uma doença antroponótica, e zoonótica, causada por cerca de 20 espécies de parasitas protozoários do gênero *Leishmania*, que são transmitidas aos seres humanos por mais de 30 espécies diferentes de flebotomíneos (PEARSON, 1996; CHAPPUIS, 2007). De acordo com as manifestações clínicas, as três principais formas de leishmaniose são: cutânea causada por *L. mexicana, L. tropical* e *L. major*, mucocutânea causada geralmente por *L braziliensis* (estas dois destroem a mucosa nasal e oral) e visceral causada pela *L. donovani, L. infantum* e *L. chagasi* (PEARSON, 1996; MCGWIRE, 2014). A leishmaniose visceral ou calazar (kala-azar), é a forma mais grave da doença, sendo fatal na ausência de tratamento (SUNDAR, 2002), alguns estudos mesmo relatando que 5 a 10% dos pacientes sob medicação morreram (BELO, 2014).

Desde há 50 anos os tratamentos contra a Leishmaniose têm sido compostos por antimoniais pentavalentes (estibogluconato de sódio e antimoniato de meglumina), devido ao baixo custo de fabricação e distribuição, assim como a eficácia que pode alcançar (até 95%) (ASILIAN, 2003; VASCONCELLOS, 2012; BELO 2014; YESILOVA, 2016), porém diversos efeitos colaterais têm sido observados, como resistência do parasita, toxicidade entre outros (LIRA, 1999; SUNDAR, 2001; SUNDAR, 2016)

Além dos medicamentos antimoniais outros compostos usados atualmente são: pentamidina (ASILIAN, 2003), anfotericina B (CROFT, 2003; BURROWS, 2014), miltefosina e paromomicina (CLEMENTI, 2011; SINGH, 2012). No entanto para todos estes compostos têm sido determinados altos níveis de toxicidade e outros problemas, como alto custo de produção, resistência das cepas de *Leishmania* e a prolongada duração dos tratamentos.

Diversos estudos avaliando a atividade antileishmanial *in vitro* de sesquiterpenos lactonizados têm sido feitos, sendo encontradas algumas moléculas com potencial atividade contra as diferentes formas parasitárias do parasita responsável da doença. Especificamente para *L. donovani*, um ensaio *in vitro* de seis germacranolídeos de *Calea zacatechichi*, identificou a calealactona C e Caleína D como moléculas promissórias leishmanicidas, sendo que os valores de IC_{50} alcançados foram nos dois casos menores a 2,9 μM, valor calculado para pentamidina (WU, 2011). Por outra parte, Avolio e colaboradores realizaram um estudo avaliando a atividade de 12 compostos isolados de *Inula viscosa* contra promastigotas de *L. donovani*, sendo encontrados três sesquiterpenos lactonizados, inuloxina A, C e D como os mais ativos contra esta forma parasitaria, inuloxina A (IC_{50} 6,9 uM) apresentou um valor de

IC_{50} perto ao observado para pentamidina (4,8 uM), um tratamento de uso comum contra *L. donovani* (AVOLIO, 2014).

Estudos *in silico* nessa mesma direção também têm sido realizados, acrescentando o conhecimento respeito a atividade antiprotozoária dos sesquiterpenos lactonizados. Um modelo QSAR permitiu identificar numa série de 40 sesquiterpenos lactonizados que a atividade contra *L. donovani* é claramente relacionada com parâmetros de quiralidade assim como a presença de donores e aceptores de ligações de hidrogênio (TROSSINI, 2014). Uma aproximação diferente foi apresentada por Bernal e colaboradores, que realizando uma triagem virtual baseada na estrutura do receptor para uma serie de 123 sesquiterpenoides (principalmente, sesquiterpenos lactonizados), encontrando que dois xantanolideos, pungiolideo A e pungeolideo B, apresentam os menores valores de energia para PTR1, uma proteína alvo de *L. donovani* (BERNAL, 2014).

Nessa mesma direção, no presente capítulo é apresentado um estudo de triagem virtual do banco de dados de sesquiterpenos lactonizados cadastrado no SistematX, o qual está composto por 1.306 estruturas, com a finalidade de propor estruturas com potencial atividade contra *L. donovani*. Na primeira parte, uma triagem virtual baseada na estrutura do ligante foi realizada, gerando modelos *Random forest*, para amastigotas e promastigotas, as duas formas parasitarias de *Leishmania.* A validação interna e externa dos modelos foi realizada, obtendo percentagens de acertos acima do 70%. Mediante esta metodologia de triagem foram classificadas como potencialmente ativas, 741 estruturas para amastigotas e 11 moléculas para promastigotas.

Em paralelo, usando a estrutura cristalina de três enzimas alvo de *L. donovani* assim como um modelo homologo da proteína PTR1, o *docking* molecular do mesmo set de 1.306 moléculas foi feita, sendo classificadas de acordo ao valor de probabilidade determinado, o qual foi calculado usando os valores de energias obtidos por cada estrutura. Finalmente, uma aproximação combinada das duas metodologias de triagem virtual foi desenvolvida para determinar os sesquiterpenos lactonizados com promissória atividade leishmanicida.

REFERENCES

ASILIAN, A.; et al. The efficacy of treatment with intralesional meglumine antimoniate alone, compared with that of cryotherapy combined with the meglumine antimoniate or intralesional sodium stibogluconate, in the treatment of cutaneous leishmaniasis. **Annals of Tropical Medicine & parasitology**, v. 97, n. 5, p. 493-498, 2003.

AVOLIO, F.; RIMANDO, A.M.; CIMMINO, A.; ANDOLFI, A.; JAIN, S.; TEKWANI, B.L.; EVIDENTE, A. Inuloxins A-D and derivatives as antileishmanial agents: structure-activity relationship study. **Journal of Antibiotics,** v. 67, n. 8, p. 597, 2014.

BELO, V. S.; et al. Risk factors for adverse prognosis and death in American visceral leishmaniasis: a meta-analysis. **PLoS neglected tropical diseases**, v. 8, n. 7, p. e2982, 2014.

BERNAL, F.A.; COY-BARRERA, E. In-silico analyses of sesquiterpene-related compounds on selected leishmania enzyme-based targets. **Molecules,** v. 19, p. 5550–5569, 2014.

BURROWS, J.N.; et al. The role of modern drug discovery in the fight against neglected and tropical diseases. **MedChemComm**, v. 5, n. 6, p. 688-700, 2014.

CHAPPUIS, F.; SUNDAR, S.; HAILU, A.; GHALIB, H.; RIJAL, S.; PEELING, R.; ALVAR, J.; BOELAERT. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?. **Nature reviews microbiology**, v. 5, n. 11, p. 873-882, 2007.

CLEMENTI, A.; et al. Renal involvement in leishmaniasis: a review of the literature. **Nephrology Dialysis Transplantation Plus**, v. 4, n. 3, p. 147-152, 2011.

CROFT, S.L.; COOMBS, G. H. Leishmaniasis–current chemotherapy and recent advances in the search for novel drugs. **Trends in parasitology**, v. 19, n. 11, p. 502-508, 2003

LIRA, R.; et al. Evidence that the high incidence of treatment failures in Indian kala-azar is due to the emergence of antimony-resistant strains of Leishmania donovani. **The Journal of infectious diseases**, v. 180, n. 2, p. 564-567, 1999.

MCGWIRE, B. S.; SATOSKAR, A. R. Leishmaniasis: clinical syndromes and treatment. **QJM: An International Journal of Medicine,** v. 107, n. 1, p. 7-14, 201

PEARSON, R. D.; DE QUEIROZ SOUSA, A. Clinical spectrum of leishmaniasis. **Clinical Infectious Diseases**, p. 1-11, 1996.

SINGH, N.; KUMAR, M.; SINGH, R. K. Leishmaniasis: current status of available drugs and new potential drug targets. **Asian Pacific Journal of Tropical Medicine**, v. 5, n. 6, p. 485-497, 2012.

SUNDAR, S. Drug resistance in Indian visceral leishmaniasis. **Tropical Medicine & International Health**, v. 6, n. 11, p. 849-854, 2001.

SUNDAR, S.; RAI, M. Laboratory diagnosis of visceral leishmaniasis. **Clinical and diagnostic laboratory immunology**, v. 9, n. 5, p. 951-958, 2002.

SUNDAR, S.; SINGH, A. Recent developments and future prospects in the treatment of visceral leishmaniasis. **Therapeutic advances in infectious disease**, v. 3, n. 3-4, p. 98-109, 2016.

TROSSINI, G.H.; MALTAROLLO, V.G.; SCHMIDT, T.J. Hologram QSAR studies of antiprotozoal activities of sesquiterpene lactones. **Molecules**, v. 19, p. 10546–10562, 2014.

VASCONCELLOS, É.C.F.; et al. Intralesional meglumine antimoniate for treatment of cutaneous leishmaniasis patients with contraindication to systemic therapy from Rio de Janeiro (2000 to 2006). **The American journal of tropical medicine and hygiene**, v. 87, n. 2, p. 257-260, 2012.

WU, H.K.; FRONCZEK, F.R.; BURANDT, C.L.; ZJAWIONY, J.K. Antileishmanial germacranolides from Calea zacatechichi. **Planta medica**, v. 77, n. 07, p. 749-753, 2011.

YESILOVA, Y.; et al. Meglumine antimoniate is more effective than sodium stibogluconate in the treatment of cutaneous leishmaniasis. **Journal of Dermatological Treatment**, v. 27, n. 1, p. 83-87, 2016.

Ligand-based and Structure-based Virtual Screening to Select Compounds with Potential Leishmanicidal Activity from an Asteraceae Sesquiterpene Lactones Database

Chonny Herrera-Acevedo [a], Mayara Dos Santos Maia [a], Élida Batista Vieira de Sousa [a], Luciana Scotti [a] and Marcus Tullius Scotti [a]*

[a] Post-Graduate Program in Natural and Synthetic Bioactive Products
Federal University of Paraíba
Cidade Universitária- Castelo Branco III, João Pessoa, PB, Brazil

E-mail: mtscotti@gmail.com

Artigo será submetido na revista: ***Medicinal Chemistry***
Fator de Impacto: 2.331

Abstract: Leishmaniasis is a complex of diseases caused by intracellular parasitic protozoans of the genus *Leishmania*. Among the three main manifestations of the disease, visceral leishmaniasis (VL) is the most severe, causing more than 300,000 cases annually worldwide. Several problems have been observed with the current treatments used against this disease, being necessary the search of new drugs. In this study, using a ChEMBL dataset of 3,152 and 1,569 structures tested *in vitro* against amastigotes and promastigotes of *L. donovani*, two random forest (RF) models were performed with accuracy over 74% for cross-validation and test set. Subsequently, a ligand-based virtual screening (VS) of 1,306 structures registered in an in-house database SistematX was performed. In parallel, molecular modeling by homology of *L. donovani* Pteridine reductase 1 was developed using *L. major* PTR1 as template. From this homologue model and together with the crystal structure of three target proteins, a structure-based virtual screening (molecular docking) was also carried out for the SistematX SLs dataset. Finally, a consensus analysis of the two VS approaches allow the normalization of scores of probabilities to propose potentially active SLs against *L. donovani*, establishing a possible mechanism of action of them.

Keywords: *Leishmania donovani* • Sesquiterpene lactones• Ligand-based virtual screening • Structure-based virtual screening • Asteraceae • Machine learning •database

1. Introduction

Leishmaniasis is a complex of diseases caused by intracellular parasitic protozoans of the genus *Leishmania*, a representative of the order Kinetoplastida and family Trypanosomatidae [1]. They are heteroxenous parasites that require two hosts to complete their life cycle, namely a vertebrate and an invertebrate, particularly hematophagus flies of the genus *Phlebotomus*, order Diptera and family Psychodidae [2]. When the sandfly bites an infected individual or a reservoir draws host infected macrophages or free amastigotes in blood or even in tissues. The amastigotes, upon reaching the insect's midgut, develop into promastigotes. These flagella forms, after rapid multiplication, become infective and migratory promastigotes. Of the previous intestine are regurgitated or introduced into the skin of the next host when the insect takes a new blood meal [3].

The clinical manifestation of the disease depends on the complexity of the interaction between the host's immune system and the protozoan type, and there are four forms of presentation of the disease: cutaneous leishmaniasis, cutaneous mucosal leishmaniasis, disseminated cutaneous leishmaniasis and the most severe, visceral leishmaniasis [4]. Estimated of visceral leishmaniasis (VL) have reached 300,000 of 500,000 global cases annually [5], caused by specie *Leishmania donovani*, is fatal in absence of treatment [6]. VL is a chronic, generalized infectious disease characterized by clinically due to the manifestation of irregular fever, splenomegaly and anemia [7]. In addition, is observed a high incidence of canine visceral-cutaneous leishmaniasis develop a clinical form similar to that of humans [8].

Current treatment by drugs is complicated by severe adverse effects and high toxicity, there is thus an urgent need the search for new drugs or lead structures for such disease, considered “neglected tropical disease” due to the lack of research and development of new drugs over many decades [9]. Many classes of chemicals, among them, natural products, found to provide interesting leads for such parasites, particularly, sesquiterpene lactones (STLs) have shown interesting activities [9a, 10]. Several studies have been performed evaluating SL leishmanicidal activity, aiming to found potential anti-*Leishmania* drugs [11]. Schmidt et al., reported the bioactivity of 40 sesquiterpene lactones tested *in vitro* for activity *against Leishmania donovani* and other parasites as well as cytotoxicity against L6 rat skeletal myoblasts, 4,15-dinor-1,11(13)-xanthadiene-3,5β:12,8β-diolide and 8-epixanthatin 1β,5β-epoxide, two SLs isolated from *Xanthium brasilicum* were the most selective against *L. donovani* [12]. In another study, three irregular linear SLs present in *Anthemis auriculata*: 4-

hydroxyanthecotulide, anthecotulide and 4-acetoxyanthecotulide showed activity against amastigotes of *L. donovani* with IC_{50} values of 3.27, 8.18 and 12.5 mg / mL, respectively [13].

Enzymes or metabolites present in the parasite but absent from their mammalian host are considered as ideal targets for rational drug design [14]. Some of the main enzymatic targets studied in the development of new drugs for the treatment of leishmaniasis, using natural products as inhibitors are: Ornithine decarboxylase (ODC, E.C. 4.1.1.17), a key enzyme in polyamine biosynthesis in *Leishmania donovani,* catalyzes the conversion of ornithine to putrescine that is a precursor of other polyamines. Inhibition of ornithine decarboxylase deplete the parasite trypanothione, this imbalance in the redox metabolism causes increase in reactive oxygen species [15]. N- myristoyltransferase (NMT, E.C. 2.3.1.97), a ubiquitous enzyme catalyzes the attachment of the myristic acid (14-carbon saturated fatty acid), to the amino-terminal glycine residue of a subset of eukaryotic proteins, being a suitable, potential, and valid target for anti-*Leishmania* drug development [16]. Mitogen-activated protein kinase 3 (MPK3, E.C. 2.7.11.25), a mediator of signal transduction and important regulator of cell differentiation and cell proliferation in eukaryotic cells. So far, ten MAP kinases have been identified in *L. mexicana* [17]. Pteridine reductase 1 (PTR1, E.C. 1.5.1.33), a NADPH dependent short-chain reductase responsible for the unusual salvage of pterin in the protozoan parasite *Leishmania*, acts as a metabolic bypass for drugs targeting dihydrofolate reductase [18].

In this study, two virtual screening (VS) approaches were performed to select from 1,306 sesquiterpene lactones (SLs) of an in-house database, structures potentially active against *L. donovani*. Initially, ligand-based VS was developed using random forest (RF) models constructed for the two parasitic forms, amastigote and trypomastigote. Subsequently, using a homologue model of *L. donovani* Pteridine reductase 1 and the crystal structure of three enzymes of this species of *Leishmania,* a structure-based virtual screening was also carried out for the SistematX SLs dataset. Finally, a combined-approach of the two VS was performed to propose potentially active SLs against *L. donovani*, establishing a possible mechanism of action of them.

2. Results and Discussion

Following the good practices of QSAR and molecular docking [19], a combinatory study of ligand-based and structure based VS was performed using the databank of 1,306 SLs stored in SistematX. Ligand-based VS was performed starting of a ChEMBL databank of structures with

in vitro activity against *L. donovani*. After of process of data curation, Volsurf+ descriptors were calculated from the 3D structures that were generated in ChemAxon JChem. Using KNIME software two RF models were generated and after validated for determining the active probability of the entire SLs dataset (Figure 1).

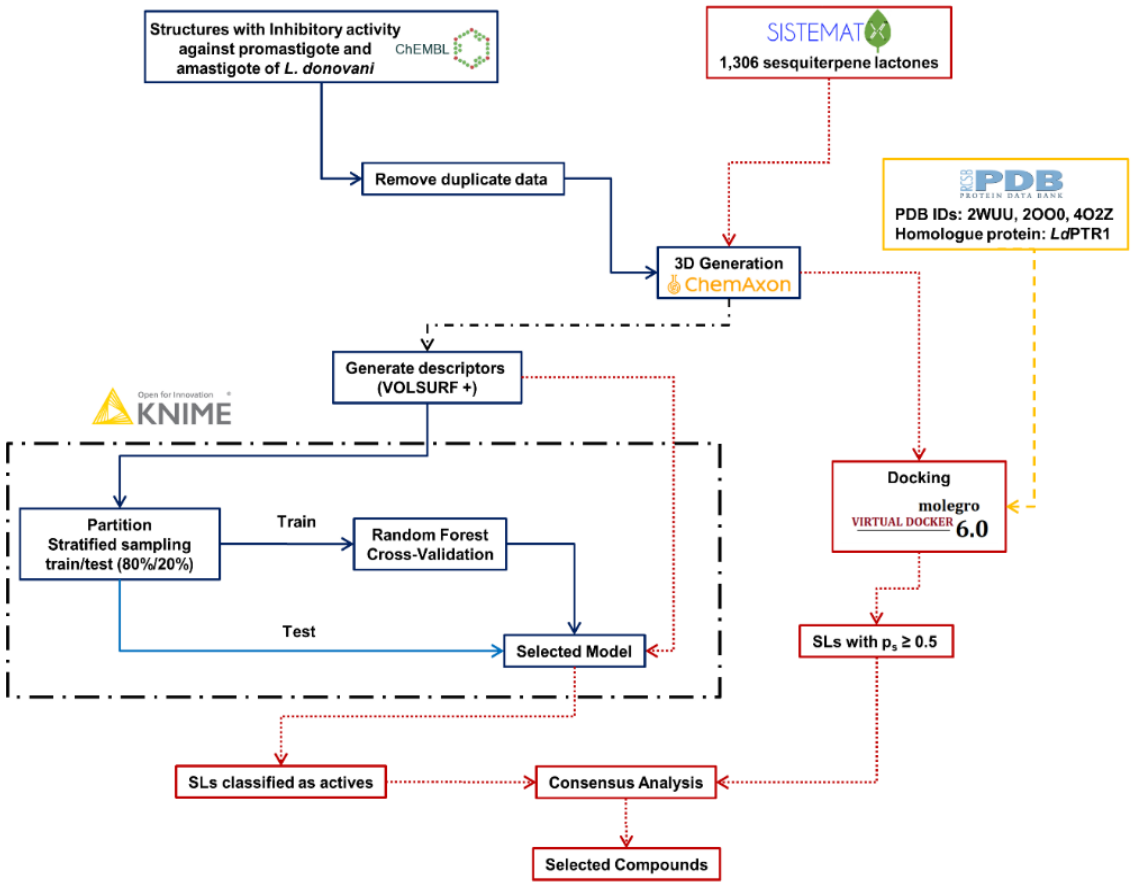


Figure 1: VS methodology used in this study. Solid blue lines represent the two sets of compounds used to generate the RF model for *L. donovani* amastigotes and promastigotes, and to validate them (Clear blue line – external test set). The red dotted lines represent the SLs of Asteraceae obtained from SistematX (in-house database). The black dash-dot line represents both datasets (CHEMBL and SistematX). The yellow dash line represents the three *L. donovani* protein structures extracted from the PDB databank (PDB ID: 2WUU, 2OO0 and 4O2Z) as well as the homologue protein *Ld*PTR1. The dash-dot border delimits the process performed in the KNIME software.

In parallel, using the homologue protein *Ld*PTR1, which was obtain from the crystal structure of *Leishmania major* Pteridine reductase 1 (*Lm*PTR1) [20] and three crystal structures of target enzymes for *L. donovani*: ornithine decarboxylase (ODC) [21], N-

Myristoyltransferase (NMT) [22] and Mitogen-activated protein kinase 3 (MPK3), molecular docking for the SLs dataset were performed in Molegro virtual docker 6.0.1. Finally, using active probability values obtained by the two methodologies, a consensus analysis was performed for selecting those molecules with the best values of combined scores (Figure 1).

2.1. Ligand-based virtual screening

Initially using a computer with an i7-4790 processor, running at 3.6 GHz, and equipped with 16 GB of RAM memory, 128 Volsurf+ descriptors were calculed for the two ChEMBL dataset: promastigote (1,628 structures) and amastigote (3,259 structures). This process took approximately 1 hour and 30 minutes. All structures were classified as active or inactive (binary classification) using as cut-off a $pIC_{50} \geq 4.7$, this value allowed the maximization of the representativeness of the chemical space for each class of structure (active – inactive). Subsequently, VolSurf+ descriptors values together the respective binary classification was used as input data for generating two Random Forest (RF) model in Knime program. Several models were evaluating trying to as minimize the false positive rate of the model. Finally, for both models 200 trees and the Gini Index as split criterion were selected. Structures with pIC_{50} values between 4.6 and 4.7 (range of 0.1 units) were excluded, allowing to avoid edge effect and improving the predictive capacity of the models since it minimizes the differences in the activity values resulting from the errors and different experimental protocol.

Table 1. Summary of internal cross-validation and test results obtained using the RF algorithm on the total set of 3152 compounds (2522–training set and 630–test set) for amastigote and 1569 compounds (1256–training set and 313–test set) for promastigote.

		Cross-validation			Test		
Model		Samples	Match	%	Samples	Match	%
Amastigote	Active	1240	937	**75.6**	321	230	**71.7**
	Inactive	1282	1003	**78.2**	309	241	**78.0**
	Total	2522	1940	**76.9**	630	471	**74.8**
Promastigote	Active	608	454	**74.7**	148	107	**72.3**
	Inactive	648	515	**79.5**	165	129	**78.2**
	Total	1256	969	**77.1**	313	236	**75.4**

For the training set of both RF models, the match percentage values reached were close to 100%. Meanwhile, for cross-validation and test set, values above of 71.7% were obtained. In all cases, inactive value rates (from 78.0 to 79.5%) were greater respect the active rates (from 71.7 to 75.6%), showing that the obtained models were highly restrictive, a fundamental

concept for this type of studies in order to minimize the probability of selection of false positives structures (molecules inactives predicted as actives).

Sensitivity and specificity are two key parameters for the selection of the RF models. Specificity that is defined as the true negative rate is bigger in Promastigote model (79.5 and 78.2%) respect the amastigote model (78.2 and 78.0%) both for internal validation and for the test set. In the same way, sensitivity (defined as the true positive rate) also was evaluated. Results show that the percentage of predicting true positive compounds is better in the amastigote model (75.6%) with respect to the cross-validation. For test set, promastigote model reached a percentage slightly higher (72.3%) than the amastigote model.

ROC curve is a quality parameter that shows the true positive rate (sensitivity) versus the false positive rates (1-specificity), the values of area under curve (AUC) of a ROC plot are can take values between 0 to 1 [23]. The maximum value is reached when exist a perfect separation of values between two groups, meanwhile a ROC area equal to 0.5 means that the variable of interest cannot be distinguished between the two groups (Figure 2, gray line).

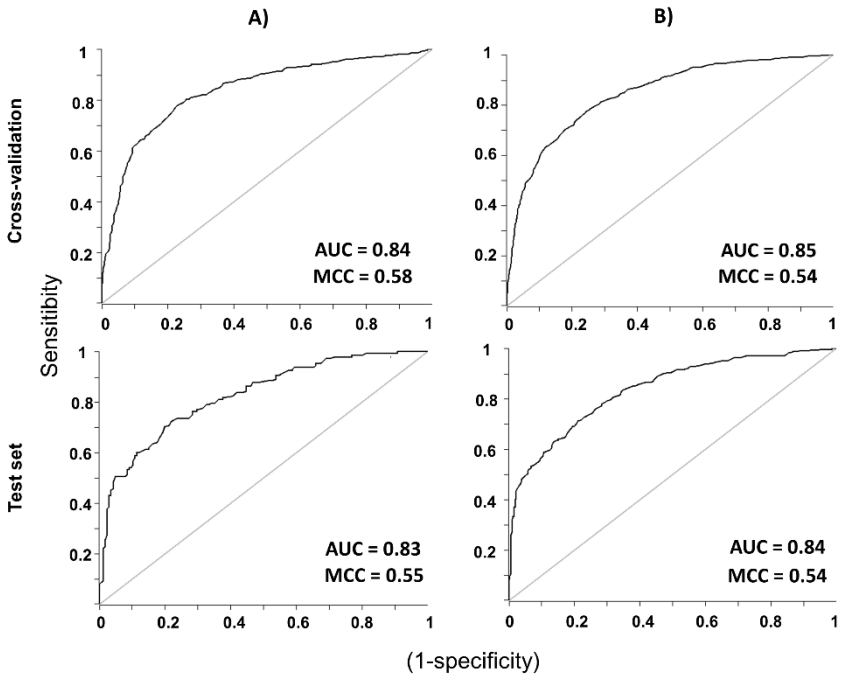


Figure 2. ROC plot, sensitivity versus (1-specificity), generated by the selected RF model for cross-validation and test set: A) Promastigote and B) Amastigote. AUC = value of the Area Under the Curve; MCC = Matthews's Correlation Coefficient

For ROC curves of *L. donovani* RF models, AUC above of 0.83 were obtained showing a high differentiation between the active compound and inactive compounds of the ChEMBL data set. AUC values for amastigote model (Figure 2 B), 0.85 (CV,) and 0.84 (test set) were minimally greater respect the values reached for the promastigote models (Figure 2A), 0.84 (CV) and 0.83 (test set).

Additionally, Matthews's Correlation Coefficient (MCC) a quality parameter was determined for the two models. MCC relates all values of the confusion matrix also were calculated (Equation 1). MCC value equal to 1 indicates a perfect correlation, 0 a random prediction and –1, total disagreement between prediction and observation [24]. Here, promastigote model showed the highest values, 0.58 and 0.55, for cross-validation and test set, respectively. Meanwhile amastigote model reached in both cases a MCC value of 0.54.

Equation 1:

$$\mathrm{MCC} = \frac{\mathrm{TP\ x\ TN} - \mathrm{FP\ x\ FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Here, TP = true positive rate, TN = true negative rate, FP = false positive rate and FN = False negative rate [24].

Likewise, aiming to identify compounds in the test set (for both models) as well as SLs for which predictions may be unreliable, the domain of applicability was calculated. Only one molecules of 630 structures that composed the test set of the amastigote model were classified as unreliable. For promastigote model as well as for the SLs dataset used for VS, the totality of the structures was classified as reliable.

Ligand-based virtual screening for predicting structures potentially active against the two parasitic forms of *Leishmania* was performed using the entire set of 1,306 SLs registered in SistematX. For amastigote model 741 molecules were classified as active, with *p* values between 0.5 and 0.87. Two types of skeleton are observed among the five best-ranked structures: three pseudoguaianolides (figure 3, structures **1, 3** and **4**) and two germacranolides (**2** and **5**) present the highest active probability values (*p*). Some structural features are clearly observed in this group of structures such as the present of two acetyl moieties for the most active germacranolide as well as for the most active SL, cumanin-diacetate (**1**), a secondary metabolite that is present in plants of *Ambrosia* genus (*A. psilostachya* and *A. acanthicarpa*)

[25]. In the same way, a system α-methylene-γ-lactone is observed only for the two best-ranked molecules (p=0.87), the presence of this moiety has been related with the interaction of this type of metabolites with sulfhydryl group of cysteine (Michael addition) [26].

Interestingly, arnicolide C (**3**, Figure 3) and arnicolide D (**4**), two characteristic SLs of *Arnica montana L.*, also known as mountain arnica [27], presented high p values by ligand-based VS, being two of the best-ranked molecules for amastigote parasitic form. Structurally, these two molecules are pseudoguaianolides with similar molecular weight, being only differenced by the substitution in the carbon-6, arnicolide C (**3**) present a isobutyrate group while arnicolide D (**4**) have a methacrylate group.

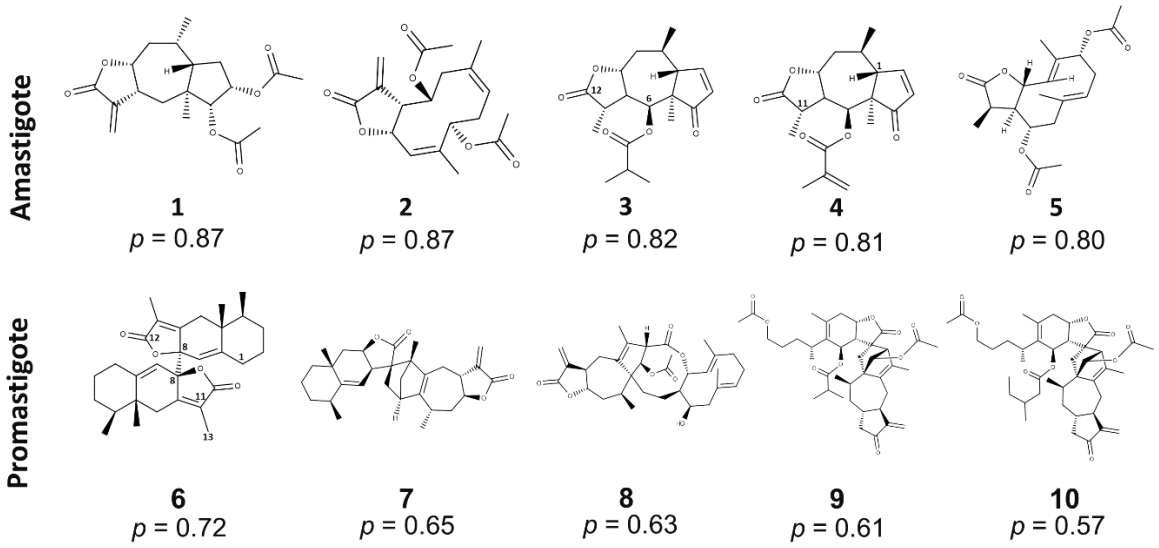


Figure 3. Potentially active sesquiterpene lactones (Five best-ranked) using ligand-based virtual screening for the two-parasitic form of L. donovani; p= active probability value.

For promastigote model, only 11 of the 1,306 SLs screened were classified as active. A different pattern structural was observed for the five best-ranked SLs respect the previously observed for amastigote parasitic form. In all five cases SL dimers presented the highest p values. Bedfordia symmetric dimeric lactone (**6**), a secondary metabolite isolated of *Bedfrodia salicina*, a specie located commonly in Tasmania, reached a p value equal to 0.72. This SL is composed by two units of eremophilanolide linked through a covalent bond between of carbon-8 of both structures (**6**, figure 3).

In the same way, carpedilactone G (**7**) also presented a high active probability value against promastigote, being a common feature with **6**, the presence of eremophilanolide skeleton in its structure. For this parasitic form are highlighted two dimers isolated of *Inula japonica*, japonicone V (**9**) and japonicone U (**10**), that reached *p* values of 0.61 and 0.57 respectively, and were classified by ligand-based VS as potential antileishmanial structures. Previously, have been reported that dimeric sesquiterpene lactones of this plant have anti-inflammatory and cytotoxic activity [28].

After, looking for false positive molecules, the entire set of 742 (amastigote) and 11 (promastigote) SLs classified as active in ligand-based VS, were tested using PAINS-remover, a web tool that remove the Pan Assay Interference Compounds (PAINS) from screening libraries [29]. The results show that all molecules classified as active for promastigote parasitic form passed by the filter, while for amastigote, among the 741 molecules with *p* values above of 0.5 by RF model, two-structures, Neurolobatin B ($p = 0.508$) and 3-chlorodehydroleucodin ($p = 0.544$) did not pass the filter being classified as false positives.

2.2. Structure-based virtual screening

Homology model of LdPTR1

Molecular Modeling by Homology of *L. donovani* Pteridine reductase 1 (PTR1) protein, was performed using *L. major* PTR1 as template [20]. A good template for *Ld*PTR1 was obtained because the high level of similarity between the target sequence (*Ld*PTR1) and the template sequence (*Lm*PTR1). To verify and validate the reliability and stereochemical quality of the modeled protein, data from the Ramachandran, VERIFY 3D and WHAT IF graph were considered. Analysis of the Ramachandran plot shows that the main chain conformations are 88.3% of the residues in the most favored regions, 11.7% allowed and 0% outliers, (supporting information). In this analysis, since there was no residue in the outlier region, the generated model was considered satisfactory. The G factors, which indicate the quality of the covalent distance and the bond angle, were 0.15 for dihedrons and 0.09 for phi / psi. Positive or non-negative values indicate a model of good stereochemistry. According of VERIFY 3D results; 85.07% of the residues had a mean 3D-1D score ≥ 0.2, which indicates a reliable model, since it is superior to 80% of the amino acids that marked = 0.2 in the 3D / 1D profile. The quality of the atomic contacts involving the atoms of each residue was analyzed using the Fine Packing Quality Control module of WHAT IF, which compares the distribution of atom positions

around each residue. The mean score of all wastes is -0.687. A score of less than -5.0 for a residue means bad or unusual atomic contacts. The graphical visualization of the modeled protein is observed in supporting information.

Molecular docking

In addition to homologue *Ld*PTR1, three *L. donovani* target proteins whose crystal structures were obtained in PDB databank: ornithine decarboxylase (ODC, E.C. 4.1.1.17), N-Myristoyltransferase (NMT, E.C. 2.3.1.97) and Mitogen-activated protein kinase 3 (MPK3, E.C. 2.7.11.25), were analyzed by molecular docking. Initially, the methodology was validated performing redocking of the ligand reported in PDB crystal structure for each one of three *L. donovani* proteins used in this study, for *Ld*PTR1 these values were not obtained. Docking scores and their respective RMSD values are listed in Table 2.

Using the same parameters, a virtual screening of the entire set of 1306 SLs was performed. The structures were ranked according to binding energy values as well as the calculated probability as active compound using equation 2.

Equation 2:

$$p_s = \frac{\mathrm{E_i}}{\mathrm{E_{min}}} \quad \text{if} \quad \mathrm{E_i} < \mathrm{E_{ligand}}$$

Here, p_s = structure-based probability; E_i = docking energy of compound i, and i varies from 1 to 1306 (SLs dataset); E_{min} = the lowest energy value of the dataset; E_{ligand} = the ligand energy from protein crystallography.

This equation proposed for structure-based VS, aims to normalize scores obtained by molecular docking to have values comparable to the active probability values obtained by Ligand-base VS. Those structures that presented structure-based probability values (p_s) above of 0.5 were classified as active. Additionally, a principle of selection is that the structures must have an energy lower than that obtained by the ligand of crystallography study, for homologue *Ld*PTR1, this principle of selection was not applied, being SLs only ranked respect the minimum value of docking energy.

NMT protein presented 528 compounds with probability values above 0.5, however, only 490 of these had binding energy values lower than the PDB ligand, 2-oxopentadecyl-CoA (PDB

ID: NHW), -50.4 KJ/mol. For MPK3 and ODC, the number of SLs that have p_s values > 0.5, was 243 and eight molecules respectively. For both enzymes, it was observed only two structures (**13-14** for MPK3 and **14-15** for ODC) with lower energy respect the inhibitors reported in crystal structures. Finally, for *Ld*PTR1, 346 molecules showed p_s values greater than cut-off, nevertheless no reference ligand was used for this homologue protein. The best-ranked molecule for each enzyme are listed in Table 2 and Figure 4. All binding energy values are found in the supporting information.

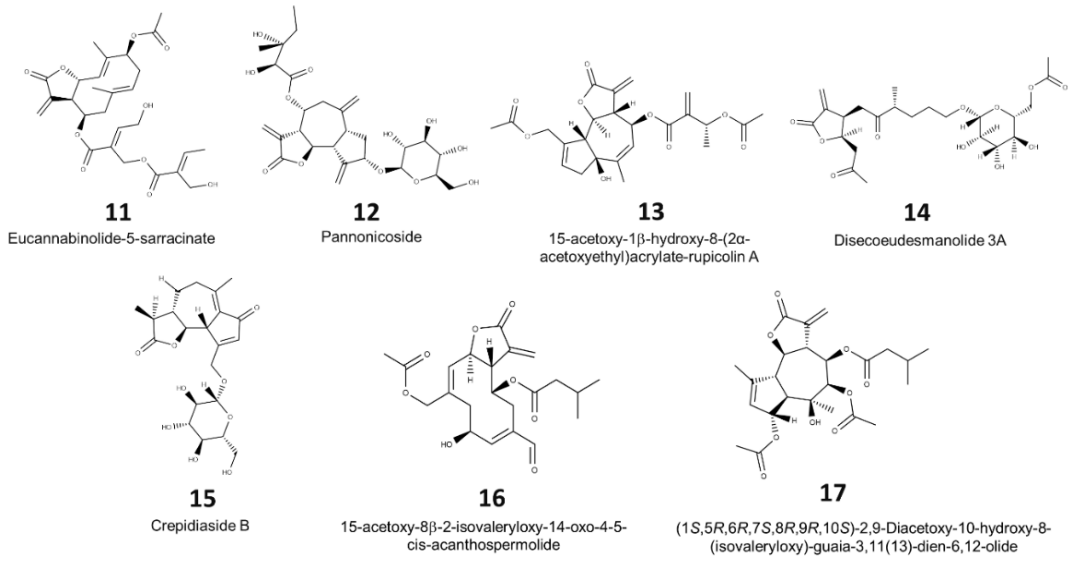


Figure 4. Antileishmanial SLs (best-ranked) from the structure-based virtual screening.

Table 2. The energy of docking (kJ/mol) of the best-ranked SLs by the structure-based approach for each of the four *L. donovani* protein studied. Ligand = Energy of docking (kJ/mol) for the PDB ligand.

***L. donovani* protein**	**SL-1**	**SL-2**	**Ligand**	**Redocking RMSD**
NMT	**11** (-98.8)	**12** (-98.1)	(-50.4)	1.77
ODC	**13** (-92.1)	**14** (-83.7)	(-72.0)	0.14
MPK3	**14** (-128.5)	**15** (-114.4)	(-112.6)	0.29
*Ld*PTR1	**16** (-76.7)	**17** (-76.3)	-	-

Guaianolide skeleton is predominant among the seven SLs that presented the highest probability values (p_s) for the four enzymes studied, **12, 13, 15** and **17** (figure 4) belong to this

group of substructures. In the same way, two germacranolides, **11** and **16**, have very low docking energy values for NMT and *Ld*PTR1, respectively. Structure **14**, a SL extracted from *Picradeniopsis woodhousei*, is classified as a disecoeudesmanolide and is the only compound by structure-based VS that presented high scores for two *L. donovani* proteins (table 2), being the best-ranked SL for MPK3 (-128.5 kJ/mol), at the same time that it has one of the lowest energy values for ODC (-83.7 kJ/mol).

Hydrogen-bonding (H-bond) interactions of disecoeudesmanolide 3A (**14**) with residues of ODC are observed in figure 5B. Six residues of the active site of ODC interacting with this SL: Lys69, Asp 88, Glu274, Ser200, Gly276 and Tyr389, these last three are highlighted since that also interact with the inhibitor reported for PDB crystal structure, pyridoxal-5'-phosphate (PDB ID: PLP, Figure 5A). In addition to the H-bond interactions of the ester group of disecoeudesmanolide 3A with Ser200, Gly276 and Tyr389 of ODC, an interaction among oxygen of the pyranose ring and Tyr389 also is observed.

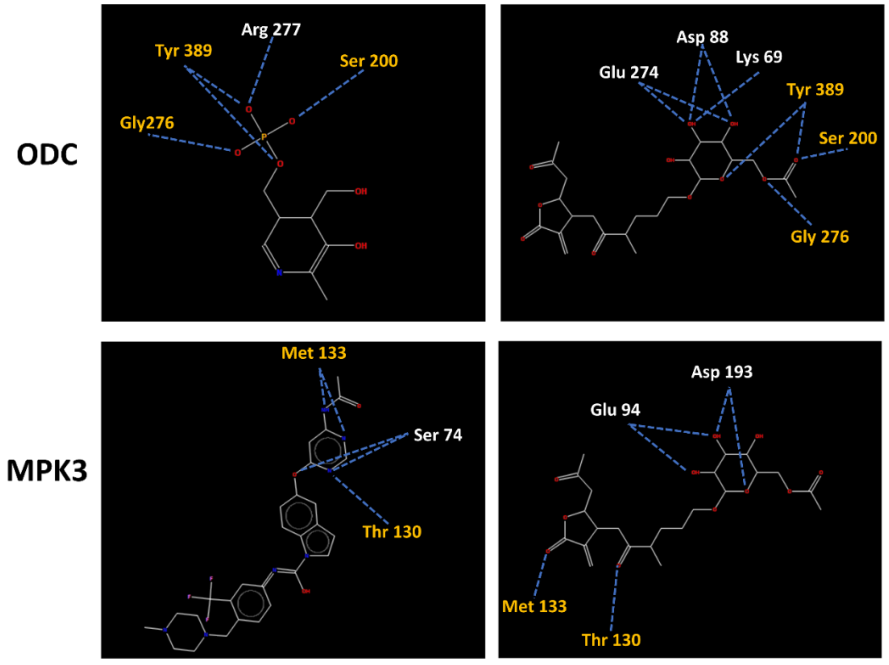


Figure 5. Hydrogen-bonding interactions (blue dot lines) of A) pyridoxal-5'-phosphate (PDB ID: PLP) and B) disecoeudesmanolide 3A (**14**) in the active site of *L. donovani* ODC protein. C) Inhibitor reported in the PDB for MPK3 (PDB ID: 046) and D) disecoeudesmanolide 3A (**14**), in the active site of MPK3 protein Common interactions are highlighted in yellow.

Alike, structure **14** also interacts with MPK3 through H-bonds with specific residues of the active site. Four H-bond interactions are formed between Glu94 and Asp193 mainly with

hydroxyl groups of the monosaccharide and the oxygen of the pyranose ring. Interestingly, two critical residues Met 133 and Thr130 are identified because these form H-bonds both with disecoeudesmanolide 3A (**14**) as well as with the inhibitor reported in the PDB for this protein (PDB ID: 046, Figure 5C). The interaction of this compound with these two residues of MPK3 is with the carbonyl groups. In this case and unlike that was observed previously for ODC, carbonyl group of the lactone ring of disecoeudesmanolide 3A interacts with MPK3, specifically with Met133.

The two best-ranked molecules for NMT were 5-sarracinate-eucannabinolide (**11**, Figure 4) and pannonicoside (**12**). In Figure 6 are represented H-bond interactions for these two SLs. Structures **11** and **12** have a similar number of H-bonds which may be related with the obtained docking energy values -98.8 and -98.1 kJ/mol, respectively. Likewise, some NMT residues are highlighted in these kinds of interactions such as: Phe14, Glu177 and Arg179, which are present in **11** and **12** as well as in PDB inhibitor, 2-oxopentadecyl-CoA (PDB ID: NHW).

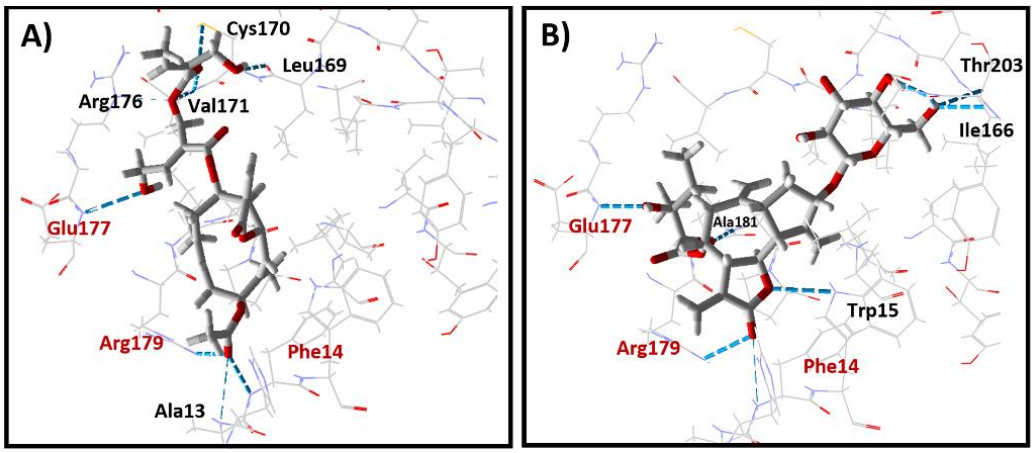


Figure 6. Docking poses of A) 5-sarracinate-eucannabinolide (**11**) and B) pannonicoside (**12**) in the active site of *L. donovani* NMT

2.3. Combined approach Structure-Based VS and Ligand-Based VS

A consensus analysis of the two methodologies used in this study (structure-based and ligand-based VS) was performed aiming to verify potentially active molecules as well as their possible mechanism of action, facilitating the identification of potential multitarget compounds. A new score (p_c, equation 3) was determined combining the probability score of both VS approaches as well as the true negative rate of RF model tries to minimize the probability of selecting false positive compounds. (equation 3)

Equation 3:

$$p_c = \frac{p_s + (1 + TN) \times p}{2 + TN}$$

where p_c = combined probability p_s = structure-based probability; TN = true negative rate; p = ligand-based probability.

Different weights are assigned to Structure-based (p_s) and Ligand-based VS (p) probabilities, since that in the first one, only are related interactions between protein and ligand, while for Ligand-based VS, pIC_{50} experimental values and molecular descriptors are used to generate the RF model. Therefore, looking for diminish the false positive rate (increment of TN), equation 3 relates the TN rate of the internal cross-validation with the p values reached by the SLs in each parasitic form model. The minimization of the probability of selection of molecules inactives as actives (FP) is very important in this kind of studies, since the selection of this type of molecules can be generated a significant increment in time and cost, something that goes against the aims of the VS studies.

Some structures that had not presented the highest scores for either of the two VS approximations appear in the consensus analysis performed as interesting structures in the search for new leishmanicidal drugs (**18**-**23**, Figure 7). Disecoeudesmanolide 3A (**14**) emerges as a potential multitarget structure because acts in three target enzymes of *L. donovani* promastigotes (table 3), and is the only SL of the data set that have high p_c score for the two parasitic forms in a specific protein, ODC.

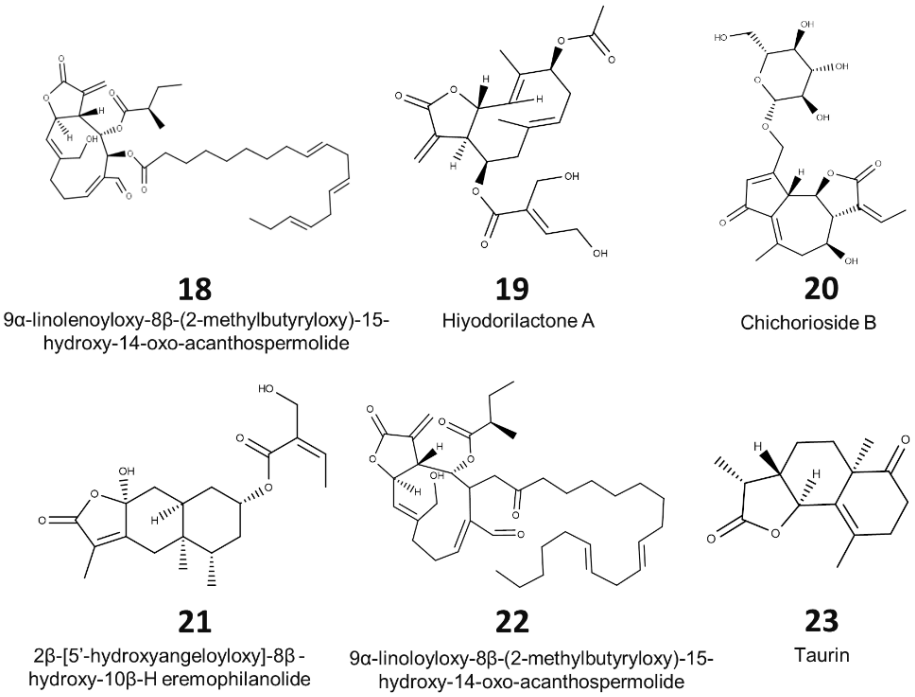


Figure 7. Representation of best-ranked structures through a combined approach of ligand-based and structure-based virtual screening for each parasitic form

Table 3. Summary of the best-ranked structures through a combined approach of ligand-based and structure-based virtual screening; p = active probability value in ligand-based VS; ps = active probability value in structure-based VS. pc = combined probability value

	Amastigote				Promastigote			
Protein	**Structure**	***p***	p_s	p_c	**Structure**	***p***	p_s	p_c
NMT	**18**	0.75	0.79	0.76	**14**	0.51	0.86	0.63
	5	0.80	0.68	0.76	**22**	0.51	0.79	0.61
ODC	**14**	0.56	0.91	0.69	**14**	0.51	0.91	0.65
	19	0.65	0.74	0.68		-	-	-
MPK3	**15**	0.63	0.89	0.73	**14**	0.51	1.00	0.69
	20	0.65	0.86	0.73	**23**	0.51	0.58	0.54
LdPTR1	**2**	0.87	0.67	0.80	**6**	0.72	0.66	0.70
	21	0.66	0.96	0.77	**22**	0.51	0.56	0.52

Besides,9α-linoloyloxy-8β-(2-methylbutyryloxy)-15-hydroxy-14-oxo-acanthospermolide (**22**, Figure 7), just like SL 14, present high combined probability values for two targets of *L. donovani* promastigotes (NMT and the homologue *Ld*PTR1), being classified as active through

the two VS methodologies of this study. Structure **22** together **18** (that appear as interesting compound against NMT in amastigote parasitic form), are characteristic secondary metabolite of *Acanthospermun hispidium*, a native plant of Central and South America [30], having a germacranolide skeleton in common bound to linoleic and linolenic fat acid ester, respectively.

For both parasitic forms in *Ld*PTR1, the best-ranked structures previously had presented high probability values in Ligand-based VS; cis,cis-3α-acetoxy-8β-acetoxy-costunolide (**2**) for amastigote and Bedfordia symmetric dimeric lactone (**6**) for promastigote, reached p_c values above of 0.70. Nine compounds were classified as active by this combined-approach for promastigote parasitic form, however, only four SLs also were classified as active by the two VS methodologies used. This behavior is observed in the four targets studied for promastigote parasitic form, since that the RF model was highly restrictive.

3. Conclusions

In this study, two different VS approaches were performed looking for structures with promising antileishmanial activity, from a dataset of 1,306 SLs that were obtained from SistematX. For Ligand-based VS, two RF models with accuracy above of 71% for both parasitic forms of *L. donovani* were constructed. Some structural features were identified, being germacranolide and guaianolide skeletons the most frequents among the most active structures, meanwhile for promastigote dimeric molecules have the highest values of *p* Structure-based VS using three PDB crystal structures and one homologue model of *L. donovani* proteins showed SLs with lower docking scores respect the reported inhibitors. Finally, through a consensus analysis the probability scores of the two VS were normalized to propose promising leishmanicidal SLs, 11 structures were proposed, since were classified as active in both VS. Disecoeudesmanolide 3A (**22)** and 9α-linoloyloxy-15-hydroxy-8β-(2-methylbutyryloxy)-14-oxo-acanthospermolide (**14**) appear as promising multitarget antileishmanial SLs.

Therefore, using two VS approaches was possible a preliminary identification of potentially active SLs against the two parasitic forms of *L. donovani*, from an in-house dataset. In the same way, the combination of the scores of probabilities of Ligand-based and Structure-based VS, emerges as interesting methodology for these types of studies being that allows the identification of molecules promising molecules as well as their possible mechanism of action

4. Experimental Section

4.1. Database

From the ChEMBL database (https://www.ebi.ac.uk/chembl/), we selected a diverse set that was initially classified according to the two *L. donovani* parasitic form: amastigote (5,500 structures), and promastigote (2,045 structures). The compounds were classified using values of $-\log IC_{50}$ (mol/L) = pIC_{50}, which led us to divide them into active ($pIC_{50} \geq 4.7$) and inactive ($pIC_{50} < 4.7$) structures; due to the variability of the experimental protocols of the ChEMBL database, a qualitative pattern was used allowing that the difference of activity values related to experimental protocols and even different strains are partially minimized. Here, IC_{50} represented the concentration required for 50% inhibition of parasite growth. For each ChEMBL dataset, 35 (amastigote set) and six (promastigote set) SLs structures, that are not included in the dataset using for ligand-base VS, with their pIC_{50} values were added in order to increase the representativeness of the generated models regarding the chemical space of this class of secondary metabolites.

Applicability domain (APD) based on the Euclidean distances was used to identify compounds in the test set for which predictions may be unreliable if the values are higher than $APD = d + Z\sigma$, where d and σ are average Euclidian distance and standard deviation of the set of samples in the training set that have lower Euclidian distance average values of all samples in the training set. The parameter Z is an empirical cut off value, 0.5 was used as the default [31]. Structures with pIC_{50} values from 4.6 to 4.7 (range of 0.1 units) were excluded, aiming to avoid edge effect and improving the predictive capacity of the models since it minimizes the differences in the activity values resulting from the errors and different experimental protocols [32].

Data curation of datasets was performed according with the suggested procedures in the literature [33]. Standardizer software (Jchem, version 16.11.28 (2016), calculation module developed by ChemAxon, http://www.chemaxon.com/) was used to canonize all SMILES codes and after duplicates structures were removed, being eliminated those have the higher values of pIC_{50}, trying to build more restrictive models since only lower activities values are used.

The final number of structures for each dataset was: 3,152 structures for amastigotes (1,561 actives - 1,591 inactives) and 1,569 structures for promastigotes (756 actives – 803 inactives). SLs dataset was obtained from the SistematX database (http://sistematx.ufpb.br), a total of 1,306 molecules were used in this study. For predictions, those SL structures that were added in the ChEMBL dataset were excluded for each parasitic form. For all structures, SMILES codes were used as input data in Marvin; ChemAxon (version 16.11.28 (2016), calculation module developed by ChemAxon, http://www.chemaxon.com/). We used Standardizer software (Jchem, version 16.11.28 (2016), calculation module developed by ChemAxon, http://www.chemaxon.com/); ChemAxon to canonize structures, add hydrogens, perform aromatic form conversions, clean the molecular graph in three dimensions. Conformers generated for the initial structure (represented by the root node in the tree) are optimized in this software. Those molecules that presented structural problems in the 3D generation were manually corrected using Marvin.

4.2. Volsurf+ descriptors

Three-dimensional (3-D) structures in SDF format were used as input data in the Volsurf+ program v. 1.0.7 [34] and were subjected to molecular interaction fields (MIFs) to generate descriptors using the following probes: N1 (amide nitrogen–hydrogen-bond donor probe), O (carbonyl oxygen–hydrogen-bond acceptor probe), OH2 (water probe) and DRY (hydrophobic probe). Additional non-MIF-derived descriptors were generated to create a total of 128 descriptors [34]. One of the main advantages to use VolSurf descriptors is the low influence of these by the conformational sampling and averaging [35].

4.3. RF models

Knime 3.1.0 software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org) [36] was used to perform all of the following analyses. Initially, the descriptors calculated in the Volsurf+ program were imported in CSV format and subsequently divided using the "Partitioning" node in the stratified sampling option, 80% of the initial dataset was classified as the training set, meanwhile, the remaining 20% composed the test set. The model was generated employing the training set and the RF algorithm. For internal validation of the model, we employed cross-validation using 10 randomly selected stratified groups, and the distributions according to activity class variables were found to be maintained in all validation groups and in the training set. The parameters selected for RF models included the

following settings: number of trees to build = 200, seed for random number generator = 1, and Gini Index as a split criterion, for training and internal cross-validation sets. From the confusion matrix the internal and external performances of the selected models were analyzed using the parameters sensitivity (true positive rate), specificity (true negative rate) and accuracy (overall predictability). In addition, for describing the true performance of the model with more clarity than accuracy, the ROC curve employing a "ROC curve" node, which uses the sensitivity and specificity parameters, was determined. The plotted ROC curve shows the true positive (active) rate versus the false positive rates (1-specificity) [23]. In this representation, when a variable of interest cannot be distinguished between the two groups, the area under the ROC curve is 0.5, while in an opposite case where a perfect separation of values between two groups without overlapping of distributions exists, the area under the ROC curve is 1. In this way, also were calculated the MCC, wherein a value of 1 represents a perfect prediction, a value of 0 represents a random prediction, and a value of −1 represents total disagreement between prediction and observation [24].

4.4. False positive remover

For detecting false positive structures among the SLs that was classified as active for the two RF models, the Substructure Filter for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays were used [29]. All SMILES codes of the SLs classified as active were submitted in PAINS removal (http://www.cbligand.org/PAINS/), those structures that the filter classified as false positive were excluded of the final analysis.

4.5. Homology model of LdPTR1

Homolog Identification of target sequences and selection of protein template

The target protein sequence was obtained from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/pubmed) [37]. The selection of the template protein was done through a search in the Basic Local Alignment Search Tool - BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [31] and to obtain the structure was used the RCSB Protein Data Bank (https://www.rcsb.org/pdb/home/home.do) [38] . The template protein selected was *Lm*Ptr1 (PDB ID: 1E7W) [20].

Alignment of template and target sequence

Alignment of multiple sequences was performed using FASTA (http://www.ebi.ac.uk/Tools/sss/fasta/) and obtained the following values for Leishmania donovani (*Ld*Ptr1): 91.0% identity and 97,2% similarity to *Leishmania major* Ptr1 (*Lm*Ptr1) (supporting information). This was not the best score alignment in the FASTA, but it is the target organism of the study.

Construction and validation of the model

The *Ld*Ptr1 model was constructed using the by homology molecular modeling method through MODELLER 9.18 software [39], which is based on the spatial constraints satisfaction modeling. Five models were generated and the lowest energy model was chosen. The stereochemical quality of the model was evaluated at the PROCHECK [40], which evaluates several stereochemical parameters, such as torsional angles of the main chain, torsional angles of the side chain, bad contacts or steric impediments, planarity, among others. The PROCHECK generates the Ramachandran graph [41] that verifies allowed and unallowed regions of the main amino acid chain. The structural quality was evaluated in VERIFY 3D software (http://services.mbi.ucla.edu/SAVES/) that analyzes the compatibility of the protein sequence with its 3D structure according to its chemical environments and WHAT IF (http: //swift.cmbi.ru.nl/servers/html/index.html) that analyzes various structural parameters, such as the atomic contacts between the residuals. The software Discovery Studio Visualizer [42] was used for the graphic visualization of the modeled protein.

4.6. Molecular docking

In addition to homologue model of PTR1, the structure of three *L. donovani* proteins, ornithine decarboxylase (PDB ID: 2OO0) [21], N-Myristoyltransferase (PDB ID: 2WUU) [22], Mitogen-activated protein kinase 3 (MPK3–PDB ID: 402Z), in complex with the respective inhibitors: pyridoxal-5'-phosphate (PDB ID: PLP), 2-oxopentadecyl-CoA (PDB ID: NHW) and 5-{[6-(acetylamino)pyrimidin-4-yl]oxy}-N-{4-[(4-methylpiperazin-1-yl)methyl]-3(trifluoromethyl)phenyl}-2,3-dihydro-1H-indole-1-carboxamide (PDB ID: 046), were downloaded from the Protein Data Bank—PDB. All the water compounds were deleted from the enzyme structure, and the enzyme and compound structures were prepared using the same default parameter settings in the same software package (Score function: MolDock Score; Ligand evaluation: Internal ES, Internal HBond, Sp2–Sp2 Torsions, all checked; Number of runs: 10 runs; Algorithm: MolDock SE; Maximum Interactions: 1500; Max. population size:

50; Max. steps: 300; Neighbor distance factor: 1.00; Max. number of poses returned: 5). The docking procedure was performed using a GRID of 15 Å in radius and 0.30 Å in resolution to cover the ligand-binding site of the four enzymes' structures.

Acknowledgements

We would like to thank the CNPq and Capes for financial Support, and the Student Agreement Program of Graduate — PEC-PG of CNPq—Brazil.

References

1. Babuadze, G.; Farlow, J.; De Koning, H. P.; Carrillo, E.; Chakhunashvili, G.; Murskvaladze, M.; Kekelidze, M.; Karseladze, I.; Kokaia, N.; Kalandadze, I., Seroepidemiology and molecular diversity of Leishmania donovani complex in Georgia. Parasites & vectors 2016, 9 (1), 279.

2. (a) Travi, B. L.; Montoya, J.; Gallego, J.; Jaramillo, C.; Llano, R.; Velez, I. D., Bionomics of Lutzomyia evansi (Diptera: Psychodidae) vector of visceral leishmaniasis in northern Colombia. Journal of medical entomology 1996, 33 (3), 278-285; (b) Basano, S. d. A.; Camargo, L. M. A., Leishmaniose tegumentar americana: histórico, epidemiologia e perspectivas de controle. Rev bras epidemiol 2004, 328-337.

3. Fiocruz As Leishmanioses. http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/ciclo_biol_gico.htm. (accessed 30 October 2017).

4. Kevric, I.; Cappel, M. A.; Keeling, J. H., New World and Old World Leishmania Infections. Dermatologic clinics 2015, 33 (3), 579-593.

5. (a) Desjeux, P., Leishmaniasis: current situation and new perspectives. Comparative immunology, microbiology and infectious diseases 2004, 27 (5), 305-318; (b) WHO Technical Report Series 949: Control of the leishmaniases: report of a
meeting of the WHO Expert Committee on the Control of Leishmaniases; World Health Organization: Geneva
2010.

6. Bhattacharya, P.; Dey, R.; Dagur, P. K.; Joshi, A. B.; Ismail, N.; Gannavaram, S.; Debrabant, A.; Akue, A. D.; KuKuruga, M. A.; Selvapandiyan, A., Live Attenuated Leishmania donovani Centrin Knock Out Parasites Generate Non-inferior Protective Immune Response in Aged Mice against Visceral Leishmaniasis. PLoS neglected tropical diseases 2016, 10 (8), e0004963.

7. Alvarenga, D. G. d.; Escalda, P. M. F.; Costa, A. S. V. d.; Monreal, M. T. F. D., Leishmaniose visceral: estudo retrospectivo de fatores associados à letalidade. 2010.

8. Alvar, J.; Vélez, I. D.; Bern, C.; Herrero, M.; Desjeux, P.; Cano, J.; Jannin, J.; den Boer, M.; Team, W. L. C., Leishmaniasis worldwide and global estimates of its incidence. PloS one 2012, 7 (5), e35671.

9. (a) Hoet, S.; Opperdoes, F.; Brun, R.; Quetin-Leclercq, J., Natural products active against African trypanosomes: a step towards new drugs. Natural product reports 2004, 21 (3), 353-364; (b) Gilbert, I. H., Drug discovery for neglected diseases: molecular target-based and phenotypic approaches: miniperspectives series on phenotypic screening for antiinfective targets. Journal of medicinal chemistry 2013, 56 (20), 7719-7726; (c) Utzinger, J.; Becker, S. L.; Knopp, S.; Blum, J.; Neumayr, A. L.; Keiser, J.; Hatz, C. F., Neglected tropical diseases: diagnosis, clinical management, treatment and control. Swiss medical weekly: official journal of the Swiss Society of Infectious Diseases, the Swiss Society of Internal Medicine, the Swiss Society of Pneumology 2012, 142.

10. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs from 1981 to 2014. J. Nat. Prod 2016, 79 (3), 629-661.

11. Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M.; Scotti, M., Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. Molecules 2017, 22 (1), 79.

12. Schmidt, T. J.; Nour, A. M.; Khalid, S. A.; Kaiser, M.; Brun, R., Quantitative structure–antiprotozoal activity relationships of sesquiterpene lactones. Molecules 2009, 14 (6), 2062-2076.

13. Karioti, A.; Skaltsa, H.; Kaiser, M.; Tasdemir, D., Trypanocidal, leishmanicidal and cytotoxic effects of anthecotulide-type linear sesquiterpene lactones from Anthemis auriculata. Phytomedicine 2009, 16 (8), 783-787.

14. Kumar, P.; Kumar, A.; Verma, S. S.; Dwivedi, N.; Singh, N.; Siddiqi, M. I.; Tripathi, R. P.; Dube, A.; Singh, N., Leishmania donovani pteridine reductase 1: biochemical properties and structure-modeling studies. Experimental parasitology 2008, 120 (1), 73-79.

15. (a) Saudagar, P.; Saha, P.; Saikia, A. K.; Dubey, V. K., Molecular mechanism underlying antileishmanial effect of oxabicyclo [3.3. 1] nonanones: inhibition of key redox enzymes of the pathogen. European Journal of Pharmaceutics and Biopharmaceutics 2013, 85 (3), 569-577; (b) Shukla, A. K.; Patra, S.; Dubey, V. K., Evaluation of selected antitumor agents as subversive substrate and potential inhibitor of trypanothione reductase: an alternative approach for chemotherapy of Leishmaniasis. Molecular and cellular biochemistry 2011, 352

(1-2), 261-270; (c) Saudagar, P.; Dubey, V. K., Cloning, expression, characterization and inhibition studies on trypanothione synthetase, a drug target enzyme, from Leishmania donovani. Biological chemistry 2011, 392 (12), 1113-1122; (d) Shukla, A. K.; Patra, S.; Dubey, V. K., Iridoid glucosides from Nyctanthes arbortristis result in increased reactive oxygen species and cellular redox homeostasis imbalance in Leishmania parasite. European journal of medicinal chemistry 2012, 54, 49-58.

16. (a) Resh, M. D., Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1999, 1451 (1), 1-16; (b) Tate, E. W.; Bell, A. S.; Rackham, M. D.; Wright, M. H., N-Myristoyltransferase as a potential drug target in malaria and leishmaniasis. Parasitology 2014, 141 (1), 37-49.

17. Chawla, B.; Madhubala, R., Drug targets in Leishmania. Journal of Parasitic Diseases 2010, 34 (1), 1-13.

18. Nichol, C. A.; Smith, G. K.; Duch, D. S., Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. Annual review of biochemistry 1985, 54 (1), 729-764.

19. (a) Castillo-González, D.; Mergny, J.-L.; De Rache, A.; Pérez-Machado, G.; Cabrera-Pérez, M. A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G. F.; Guédin, A.; Bourdoncle, A., Harmonization of QSAR best practices and molecular docking provides an efficient virtual screening tool for discovering new G-quadruplex ligands. Journal of chemical information and modeling 2015, 55 (10), 2094-2110; (b) Rueda, M.; Abagyan, R., Best Practices in Docking and Activity Prediction. bioRxiv 2016, 039446; (c) Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Kuz'Min, V. E., Existing and developing approaches for QSAR analysis of mixtures. Molecular informatics 2012, 31 (3-4), 202-221.

20. Gourley, D. G.; Schüttelkopf, A. W.; Leonard, G. A.; Luba, J.; Hardy, L. W.; Beverley, S. M.; Hunter, W. N., Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. Nature Structural & Molecular Biology 2001, 8 (6), 521-525.

21. Dufe, V. T.; Ingner, D.; Heby, O.; Khomutov, A. R.; Persson, L.; Al-Karadaghi, S., A structural insight into the inhibition of human and Leishmania donovani ornithine decarboxylases by 1-amino-oxy-3-aminopropane. Biochemical Journal 2007, 405 (2), 261-268.

22. Brannigan, J. A.; Smith, B. A.; Yu, Z.; Brzozowski, A. M.; Hodgkinson, M. R.; Maroof, A.; Price, H. P.; Meier, F.; Leatherbarrow, R. J.; Tate, E. W., N-myristoyltransferase from Leishmania donovani: structural and functional characterisation of a potential drug target for visceral leishmaniasis. Journal of molecular biology 2010, 396 (4), 985-999.

23. Hanley, J. A.; McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982, 143 (1), 29-36.

24. Matthews, B. W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure 1975, 405 (2), 442-451.

25. Geissman, T.; Griffin, S.; Waddell, T. G.; Chen, H. H., Sesquiterpene lactones. Some new constituents of Ambrosia species: A. psilostachya and A. acanthicarpa. Phytochemistry 1969, 8 (1), 145-150.

26. Schmidt, T. J., Toxic activities of sesquiterpene lactones: structural and biochemical aspects. Curr. Org. Chem 1999, 3 (577-608), 4.

27. Popławski, J.; Holub, M.; Samek, Z.; Herout, V., On terpenes. CCIX. Arnicolides-sesquiterpenic lactones from the leaves of Arnica montana L. Collection of Czechoslovak Chemical Communications 1971, 36 (6), 2189-2199.

28. Xie, C.; Wang, H.; Sun, X.; Meng, L.; Wang, M.; Bartlam, M.; Guo, Y., Isolation, characterization, and antiproliferative activities of eudesmanolide derivatives from the flowers of Inula japonica. Journal of agricultural and food chemistry 2015, 63 (41), 9006-9011.

29. Baell, J. B.; Holloway, G. A., New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. Journal of medicinal chemistry 2010, 53 (7), 2719-2740.

30. Walker, R. H.; Wells, L. W.; McGUIRE, J. A., Bristly starbur (Acanthospermum hispidum) interference in peanuts (Arachis hypogaea). Weed Science 1989, 37 (2), 196-200.

31. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. Journal of molecular biology 1990, 215 (3), 403-410.

32. Noble, W. S., What is a support vector machine? Nature biotechnology 2006, 24 (12), 1565-1567.

33. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R., QSAR modeling: where have you been? Where are you going to? Journal of medicinal chemistry 2014, 57 (12), 4977.

34. Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B., Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. Journal of Molecular Structure: THEOCHEM 2000, 503 (1), 17-30.

35. Cruciani, G.; Pastor, M.; Guba, W., VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. European Journal of Pharmaceutical Sciences 2000, 11, S29-S39.

36. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B., KNIME-the Konstanz information miner: version 2.0 and beyond. AcM SIGKDD explorations Newsletter 2009, 11 (1), 26-31.

37. Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank. Nucleic acids research 2012, 41 (D1), D36-D42.

38. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer Jr, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., The Protein Data Bank: a computer-based archival file for macromolecular structures. Archives of biochemistry and biophysics 1978, 185 (2), 584-591.

39. Šali, A.; Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology 1993, 234 (3), 779-815.

40. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography 1993, 26 (2), 283-291.

41. Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics 2003, 50 (3), 437-450.

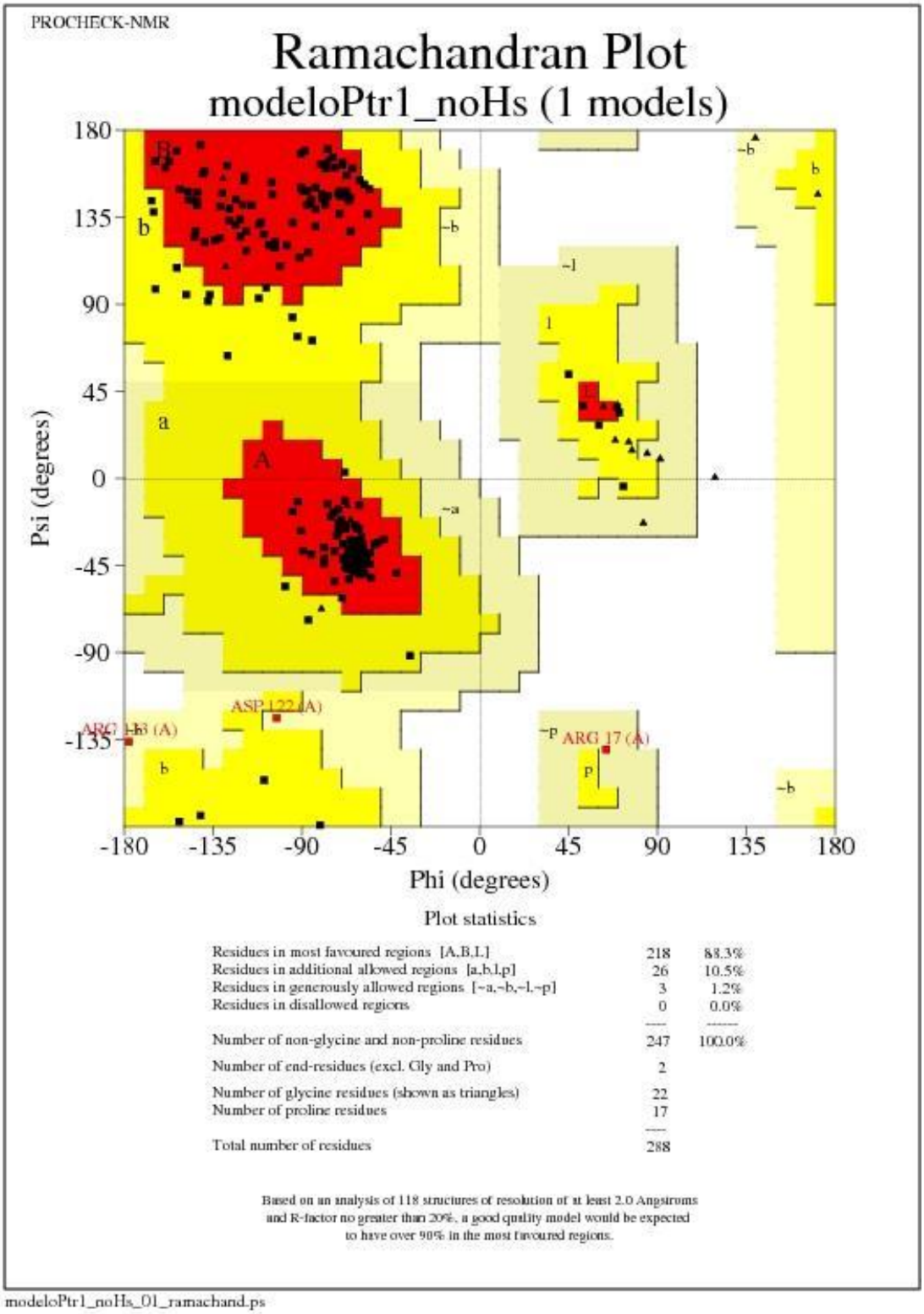
42. Diego, S., Discovery Studio Modeling Environment Release 2.1. Accelerys Software Inc 2007.

Anexos capítulo IV

Alinhamento entre a sequência molde (*Lm*Ptr1) e a sequência alvo (*Ld*Ptr1). As regiões destacadas em cinza correspondem aos aminoácidos idênticos nas duas sequências, em amarelo são os resíduos similares e em vermelho os resíduos não idênticos e não similares.

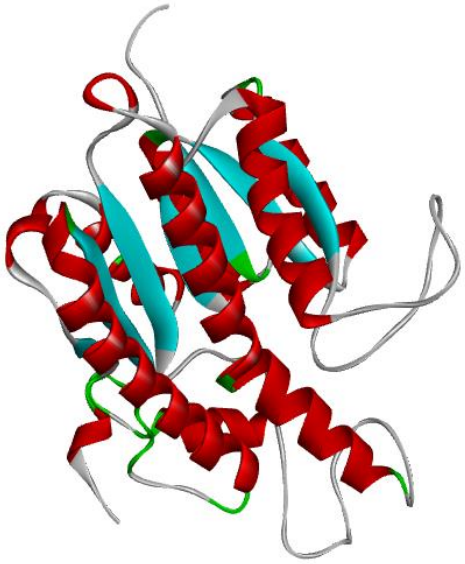
```
Query:
1>>>AAA29249.2 pteridine reductase 1 [Leishmania major] - 288 aa
Library: UniProt Knowledgebase
30331636978 residues in 89991387 sequences

Statistics:  Expectation_n fit: rho(ln(x)) = 7.5383+/-0.00016; mu= 6.3890+/- 0.009
mean_var=66.7192+/-14.011, 0's: 224 Z-trim (116.0): 780 B-trim: 5447 in 2/59
Lambda= 0.157018
statistics sampled from 60000 (103160) to 12357685 sequences
Algorithm: FASTA (3.8 Nov 2011) [optimized]
Parameters: BL50 matrix (15: -5), open/ext: -10/-2
ktup: 2, E-join: 1 (0.463), E-opt: 0.2 (0.141), width:  16
Scan time: 1647.220

>>TR:E9BG51_LEIDB E9BG51 Pteridine reductase 1 OS=Leishmania
donovani (strain BPK282A1) GN=LDBPK_230310 PE=3 SV=1 (288 aa)
initn: 1725 init1: 1725 opt: 1725 Z-score: 2117.1 bits: 399.6 E (89991387): 3.7e-108
Smith-Waterman score: 1725; 90.6% identity (97.2% similar) in 288 aa overlap (1-288:1-288)


               10        20        30        40        50        60
LmPtr1 MTAPTVPVALVTGAAKRLGRSIAEGLHAEGYAVCLHYHRSAAEANALSATLNARRPNSAI
       :.:::::::::::::::::: .:::::::::::::::::::::::.:.::::::::::::
LdPtr1 MAAPTVPVALVTGAAKRLGSGIAEGLHAEGYAVCLHYHRSAAEANTLAATLNARRPNSAI
               10        20        30        40        50        60

               70        80        90       100       110       120
LmPtr1 TVQADLSNVATAPVSGADGSAPVTLFTRCAELVAACYTHWGRCDVLVNNASSFYPTPLLR
       :::::::::: ::..:::::::::: :::.:::::::::::::::::::::::::::::
LdPtr1 TVQADLSNVAKAPAGGADGAAPVTLFKRCADLVAACYTHWGRCDVLVNNASSFYPTPLLR
               70        80        90       100       110       120

              130       140       150       160       170       180
LmPtr1 NDEDGHEPCVGDREAMETATADLFGSNAIAPYFLIKAFAHRFAGTPAKHRGTNYSIINMV
       .:::::.::::::::::.:.:::::::.:::::::::::: : ::::.:::::::.:::
LdPtr1 KDEDGHVPCVGDREAMEAAAADLFGSNAMAPYFLIKAFAHRVADTPAEQRGTNYSIVNMV
              130       140       150       160       170       180

              190       200       210       220       230       240
LmPtr1 DAMTNQPLLGYTIYTMAKGALEGLTRSAALELAPLQIRVNGVGPGLSVLVDDMPPAVWEG
       ::::.:::::::::::::::::::::::::::::::::::::::::.::::::: :
LdPtr1 DAMTSQPLLGYTIYTMAKGALEGLTRSAALELAPLQIRVNGVGPGLSVLADDMPPAVRED
              190       200       210       220       230       240

              250       260       270       280
LmPtr1 HRSKVPLYQRDSSAAEVSDVVIFLCSSKAKYITGTCVKVDGGYSLTRA
       .:::::::::::::::::::::::::::.:::::::::::::::::
LdPtr1 YRSKVPLYQRDSSAAEVSDVVIFLCSSKAKYVTGTCVKVDGGYSLTRA
              250       260       270       280
```

Gráfico de Ramachandran da proteína homologa *L. donovani* Pteridine reductase 1 *Ld*Ptr1 usando o Procheck

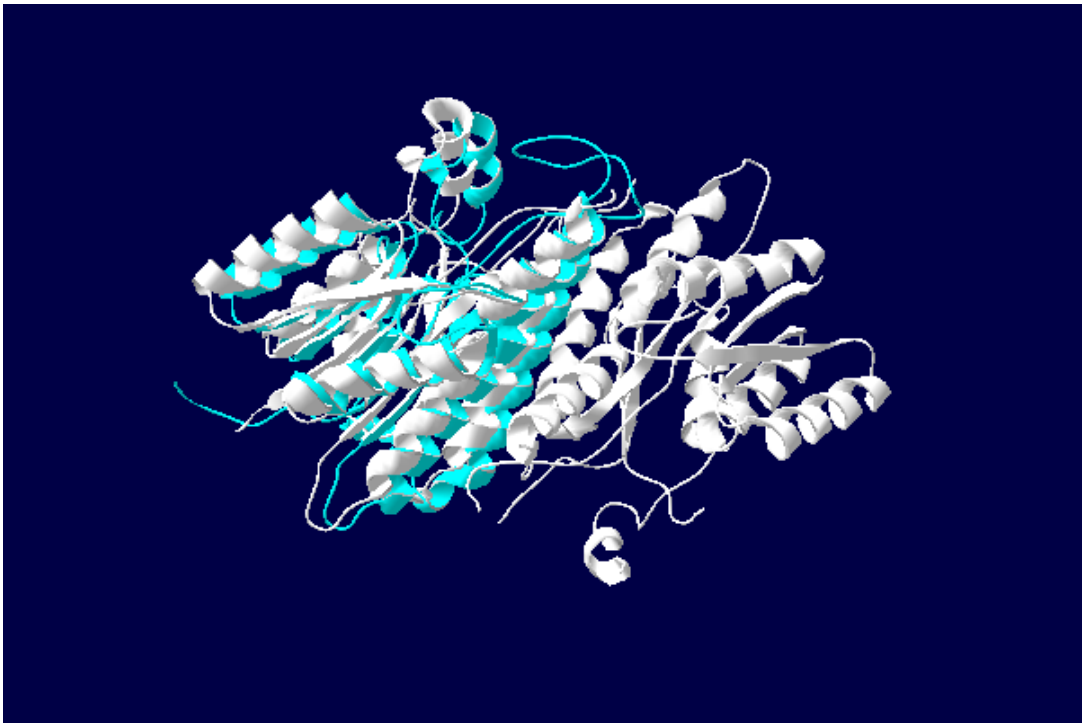


modeloPtr1_noHs_01_ramachand.ps

Visualização da proteína homologa *L. donovani* Pteridine reductase 1 *Ld*Ptr1 no Discovery Studio Visualizer.



Visualização do alinhamento da proteína homologa *L. donovani* Pteridine reductase 1 *Ld*Ptr1 com o template *L. major* Pteridine reductase 1.



Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do ligante, para **amastigotas** de *L. donovani*.

Amastigote									
ID	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**
13070	0.87	**13032**	0.72	**12642**	0.68	**12216**	0.66	**12077**	0.64
13348	0.87	**13167**	0.72	**12060**	0.68	**12658**	0.66	**12464**	0.64
13133	0.82	**12491**	0.72	**13170**	0.68	**13184**	0.66	**13294**	0.64
13132	0.81	**12906**	0.72	**12674**	0.68	**14500**	0.66	**12031**	0.64
12131	0.80	**13110**	0.72	**13137**	0.68	**12111**	0.66	**12134**	0.64
12779	0.79	**13129**	0.72	**12253**	0.68	**12213**	0.66	**12745**	0.64
13135	0.79	**12007**	0.72	**12466**	0.68	**14436**	0.66	**12758**	0.64
12560	0.78	**12234**	0.72	**12473**	0.68	**15275**	0.66	**12822**	0.64
12521	0.78	**12115**	0.71	**12561**	0.68	**12366**	0.66	**13199**	0.64
12150	0.77	**12162**	0.71	**12655**	0.68	**12724**	0.66	**15516**	0.64
13134	0.77	**12372**	0.71	**13114**	0.68	**13349**	0.66	**12457**	0.64
14277	0.77	**13146**	0.71	**12240**	0.68	**14504**	0.66	**12772**	0.64
14491	0.77	**14444**	0.71	**12320**	0.68	**12142**	0.65	**12791**	0.64
12155	0.77	**12015**	0.71	**12352**	0.68	**12230**	0.65	**14523**	0.64
13108	0.77	**12656**	0.71	**12370**	0.68	**12381**	0.65	**12856**	0.64
14408	0.76	**12377**	0.71	**14320**	0.68	**12554**	0.65	**13168**	0.64
13095	0.76	**12765**	0.71	**12329**	0.67	**12672**	0.65	**12622**	0.64
11998	0.76	**12953**	0.71	**12401**	0.67	**13019**	0.65	**12635**	0.64
12781	0.75	**13092**	0.71	**12577**	0.67	**13026**	0.65	**13297**	0.64
12416	0.75	**12518**	0.71	**12171**	0.67	**14537**	0.65	**12360**	0.63
12461	0.75	**12667**	0.71	**12553**	0.67	**12807**	0.65	**12460**	0.63
12835	0.75	**13046**	0.71	**14492**	0.67	**13103**	0.65	**12636**	0.63
14440	0.75	**14443**	0.71	**12214**	0.67	**13107**	0.65	**12712**	0.63
12472	0.74	**12654**	0.70	**12330**	0.67	**14282**	0.65	**12738**	0.63
12626	0.74	**12752**	0.70	**12647**	0.67	**14349**	0.65	**14296**	0.63
13317	0.74	**12750**	0.70	**12833**	0.67	**12024**	0.65	**14516**	0.63
12593	0.74	**12908**	0.70	**12899**	0.67	**12625**	0.65	**12898**	0.63
12853	0.74	**14555**	0.70	**12937**	0.67	**12671**	0.65	**13102**	0.63
13212	0.74	**12088**	0.70	**13195**	0.67	**14294**	0.65	**13109**	0.63
12087	0.74	**12270**	0.70	**14403**	0.67	**14442**	0.65	**13163**	0.63
12755	0.74	**12501**	0.70	**12774**	0.67	**14468**	0.65	**13319**	0.63
13211	0.74	**12784**	0.70	**12365**	0.67	**14522**	0.65	**14268**	0.63
12480	0.74	**14303**	0.70	**12865**	0.67	**12140**	0.65	**14557**	0.63
13091	0.74	**12584**	0.70	**13142**	0.67	**12511**	0.65	**12085**	0.63
12373	0.73	**15517**	0.70	**14376**	0.67	**12617**	0.65	**12375**	0.63
12943	0.73	**12549**	0.69	**12532**	0.66	**12645**	0.65	**12855**	0.63
13171	0.73	**13161**	0.69	**12576**	0.66	**12678**	0.65	**12976**	0.63
1036	0.73	**13067**	0.69	**12826**	0.66	**13063**	0.65	**14319**	0.63
12108	0.73	**14498**	0.69	**13064**	0.66	**13085**	0.65	**12559**	0.63
12666	0.73	**12812**	0.69	**14470**	0.66	**13100**	0.65	**12851**	0.63
12367	0.73	**12921**	0.69	**12110**	0.66	**14556**	0.65	**13208**	0.63
12653	0.73	**13346**	0.69	**12594**	0.66	**12043**	0.65	**14467**	0.63
12778	0.73	**14445**	0.69	**12965**	0.66	**12338**	0.65	**15518**	0.63
13145	0.73	**12262**	0.69	**13396**	0.66	**12565**	0.65	**12413**	0.63
12515	0.72	**12415**	0.69	**14414**	0.66	**12900**	0.65	**13038**	0.63
13246	0.72	**12578**	0.69	**14439**	0.66	**13074**	0.65	**13362**	0.63
13293	0.72	**13153**	0.69	**14459**	0.66	**13176**	0.65	**14487**	0.63
12754	0.72	**13154**	0.69	**14518**	0.66	**14378**	0.65	**14519**	0.63
13174	0.72	**12369**	0.68	**12143**	0.66	**12065**	0.64	**12257**	0.62

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do ligante, para **amastigotas** de *L. donovani*

Amastigote									
ID	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**
12520	0.62	**14552**	0.61	**12896**	0.59	**14554**	0.58	**12948**	0.57
12793	0.62	**12000**	0.61	**14485**	0.59	**12265**	0.58	**13005**	0.57
12831	0.62	**12053**	0.61	**12050**	0.59	**12368**	0.58	**14475**	0.57
13088	0.62	**12481**	0.61	**12893**	0.59	**12389**	0.58	**14531**	0.57
12018	0.62	**12568**	0.61	**13045**	0.59	**12918**	0.58	**14549**	0.57
12611	0.62	**12129**	0.61	**13069**	0.59	**13318**	0.58	**12109**	0.57
12707	0.62	**12575**	0.61	**14273**	0.59	**14218**	0.58	**12435**	0.57
12888	0.62	**12739**	0.61	**14421**	0.59	**14473**	0.58	**12451**	0.57
13040	0.62	**12780**	0.61	**12147**	0.59	**12075**	0.58	**12471**	0.57
12218	0.62	**12901**	0.61	**12187**	0.59	**12841**	0.58	**12740**	0.57
12458	0.62	**13196**	0.61	**12468**	0.59	**12859**	0.58	**12986**	0.57
12624	0.62	**15276**	0.61	**12797**	0.59	**13139**	0.58	**13001**	0.57
12949	0.62	**12037**	0.60	**13104**	0.59	**14458**	0.58	**13055**	0.57
13204	0.62	**12353**	0.60	**14539**	0.59	**12020**	0.57	**14404**	0.57
14514	0.62	**12405**	0.60	**12070**	0.59	**12030**	0.57	**12086**	0.56
12052	0.62	**12432**	0.60	**12099**	0.59	**12350**	0.57	**12201**	0.56
12095	0.62	**12988**	0.60	**12260**	0.59	**12474**	0.57	**12395**	0.56
12151	0.62	**14402**	0.60	**12436**	0.59	**12751**	0.57	**13096**	0.56
12718	0.62	**14483**	0.60	**12448**	0.59	**13078**	0.57	**13116**	0.56
13058	0.62	**12323**	0.60	**12613**	0.59	**13190**	0.57	**14351**	0.56
12026	0.62	**12762**	0.60	**12646**	0.59	**13320**	0.57	**12090**	0.56
12112	0.62	**14517**	0.60	**12730**	0.59	**12246**	0.57	**12189**	0.56
12249	0.62	**15266**	0.60	**12914**	0.59	**12483**	0.57	**12245**	0.56
12336	0.62	**15515**	0.60	**12947**	0.59	**12558**	0.57	**12441**	0.56
12387	0.62	**12156**	0.60	**14401**	0.59	**12633**	0.57	**12638**	0.56
12608	0.62	**12280**	0.60	**14526**	0.59	**12705**	0.57	**13314**	0.56
12644	0.62	**12827**	0.60	**15210**	0.59	**12866**	0.57	**14382**	0.56
12801	0.62	**12839**	0.60	**12694**	0.58	**13280**	0.57	**12599**	0.56
14348	0.62	**13090**	0.60	**12994**	0.58	**13954**	0.57	**12661**	0.56
14434	0.62	**13111**	0.60	**13079**	0.58	**14460**	0.57	**12929**	0.56
14474	0.62	**14336**	0.60	**13156**	0.58	**14495**	0.57	**14384**	0.56
12355	0.61	**14550**	0.60	**13307**	0.58	**12126**	0.57	**12045**	0.56
12414	0.61	**15208**	0.60	**13333**	0.58	**12852**	0.57	**12467**	0.56
12563	0.61	**15514**	0.60	**14507**	0.58	**13203**	0.57	**12716**	0.56
14377	0.61	**12118**	0.60	**12427**	0.58	**14270**	0.57	**12722**	0.56
14380	0.61	**12657**	0.60	**12736**	0.58	**14329**	0.57	**12757**	0.56
14490	0.61	**13279**	0.60	**13053**	0.58	**14520**	0.57	**13106**	0.56
12433	0.61	**14481**	0.60	**14267**	0.58	**15204**	0.57	**13342**	0.56
12455	0.61	**14482**	0.60	**14350**	0.58	**12019**	0.57	**13369**	0.56
13166	0.61	**12328**	0.60	**14374**	0.58	**12145**	0.57	**12008**	0.56
14331	0.61	**12868**	0.60	**12066**	0.58	**12165**	0.57	**12051**	0.56
12347	0.61	**12977**	0.60	**12069**	0.58	**12225**	0.57	**12148**	0.56
12903	0.61	**14412**	0.60	**12074**	0.58	**12286**	0.57	**12912**	0.56
14278	0.61	**14493**	0.60	**12119**	0.58	**12308**	0.57	**13031**	0.56
14353	0.61	**12702**	0.59	**12731**	0.58	**12590**	0.57	**14281**	0.56
14430	0.61	**13016**	0.59	**12743**	0.58	**12616**	0.57	**12425**	0.55
14452	0.61	**12409**	0.59	**12870**	0.58	**12663**	0.57	**12639**	0.55
14461	0.61	**12660**	0.59	**12886**	0.58	**12680**	0.57	**12643**	0.55
14471	0.61	**12700**	0.59	**14494**	0.58	**12720**	0.57	**12882**	0.55

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do ligante, para **amastigotas** de *L. donovani*

Amastigote									
ID	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**	**ID**	**p**
13034	0.55	**13169**	0.54	**12076**	0.53	**12327**	0.52	**12670**	0.51
13037	0.55	**14302**	0.54	**12164**	0.53	**12333**	0.52	**12810**	0.51
13048	0.55	***14447***	0.54	**12316**	0.53	**12504**	0.52	**12881**	0.51
13398	0.55	**12021**	0.54	**12858**	0.53	**12514**	0.52	**12897**	0.51
12042	0.55	**12104**	0.54	**13075**	0.53	**12529**	0.52	**12922**	0.51
12741	0.55	**12256**	0.54	**13292**	0.53	**12634**	0.52	**12968**	0.51
12767	0.55	**12305**	0.54	**14488**	0.53	**13406**	0.52	**13056**	0.51
13172	0.55	**12317**	0.54	**14529**	0.53	**14279**	0.52	**13065**	0.51
14521	0.55	**12510**	0.54	**12049**	0.53	**14293**	0.52	**13202**	0.51
12193	0.55	**12734**	0.54	**12062**	0.53	**12291**	0.52	**14462**	0.51
12391	0.55	**12785**	0.54	**12255**	0.53	**12420**	0.52	**12159**	0.51
12406	0.55	**12951**	0.54	**12348**	0.53	**12449**	0.52	**12173**	0.51
12592	0.55	**12996**	0.54	**12384**	0.53	**12507**	0.52	**12403**	0.51
12618	0.55	**13191**	0.54	**12475**	0.53	**12573**	0.52	**12591**	0.51
12860	0.55	**14538**	0.54	**12887**	0.53	**12952**	0.52	**12641**	0.51
12915	0.55	**14544**	0.54	**13324**	0.53	**12973**	0.52	**12811**	0.51
12927	0.55	**12010**	0.54	**13378**	0.53	**13049**	0.52	**13334**	0.51
13185	0.55	**12028**	0.54	**14527**	0.53	**14415**	0.52	***14450***	0.51
13209	0.55	**12166**	0.54	**14565**	0.53	**12029**	0.52	**12016**	0.51
14515	0.55	**12512**	0.54	**12229**	0.53	**12125**	0.52	**12310**	0.51
12124	0.55	**12612**	0.54	**12273**	0.53	**12340**	0.52	**12484**	0.51
12137	0.55	**12687**	0.54	**12476**	0.53	**12869**	0.52	**13003**	0.51
12139	0.55	**12787**	0.54	**12768**	0.53	**13200**	0.52	**13011**	0.51
12222	0.55	**14386**	0.54	**12871**	0.53	**14419**	0.52	**12130**	0.50
12364	0.55	**14395**	0.54	**14352**	0.53	**12103**	0.52	**12295**	0.50
12374	0.55	**14463**	0.54	**14453**	0.53	**12459**	0.52	**12349**	0.50
12446	0.55	**12304**	0.54	**14548**	0.53	**12513**	0.52	**12503**	0.50
12537	0.55	**12382**	0.54	**12138**	0.53	**12721**	0.52	**12662**	0.50
12703	0.55	**12399**	0.54	**12210**	0.53	**13013**	0.52	**12837**	0.50
12744	0.55	**12482**	0.54	**12272**	0.53	**14472**	0.52	**12884**	0.50
13044	0.55	**12676**	0.54	**12462**	0.53	**14524**	0.52	**13023**	0.50
13057	0.55	**12867**	0.54	**12878**	0.53	**12519**	0.51	**13024**	0.50
14479	0.55	**12944**	0.54	**12966**	0.53	**13329**	0.51	**12277**	0.50
15209	0.55	**13201**	0.54	**13298**	0.53	**14457**	0.51	**12620**	0.50
12123	0.55	**12354**	0.54	**14427**	0.53	**12408**	0.51	**12862**	0.50
12346	0.55	**12747**	0.54	**14543**	0.53	**12517**	0.51	**13068**	0.50
12438	0.55	**12902**	0.54	**12067**	0.52	**12555**	0.51	**13340**	0.50
12669	0.55	**12925**	0.54	**12100**	0.52	**12567**	0.51	**14301**	0.50
12704	0.55	**12964**	0.54	**12442**	0.52	**12838**	0.51	**14341**	0.50
13098	0.55	**13002**	0.54	**12756**	0.52	**12857**	0.51	**14343**	0.50
13183	0.55	**13006**	0.54	**12863**	0.52	**12877**	0.51	**14425**	0.50
14489	0.55	**13399**	0.54	**12909**	0.52	**12933**	0.51	**12047**	0.50
12258	0.54	**14532**	0.54	**12916**	0.52	**14398**	0.51	**12048**	0.50
12490	0.54	**12268**	0.53	**13027**	0.52	**14508**	0.51	**12105**	0.50
12673	0.54	**12890**	0.53	**13051**	0.52	**15211**	0.51	**12114**	0.50
12759	0.54	**12894**	0.53	**13327**	0.52	**12002**	0.51	**12212**	0.50
12766	0.54	**13216**	0.53	**14276**	0.52	**12091**	0.51	**12742**	0.50
12861	0.54	**13323**	0.53	**14297**	0.52	**12096**	0.51	**12830**	0.50
13061	0.54	**14536**	0.53	**14342**	0.52	**12386**	0.51		

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do ligante, para **promastigotas** de *L. donovani*

Promastigote	
ID	**p**
13191	0.72
14565	0.65
14419	0.63
14549	0.61
14548	0.57
12915	0.55
12771	0.54
12413	0.52
12270	0.51
12395	0.51
12415	0.51

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para quatro proteínas alvo de *L. donovani*

N-Myristoyltransferase: 490 compostos

Energia do ligante: -50,4 KJ/mol

N- myristoyltransferase														
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
12607	-98.83	1.00	**12304**	-74.20	0.75	**12656**	-68.80	0.70	**12441**	-65.40	0.66	**13298**	-62.52	0.63
14473	-98.07	0.99	**12209**	-74.03	0.75	**12898**	-68.78	0.70	**12899**	-65.40	0.66	**12504**	-62.51	0.63
12567	-95.07	0.96	**13401**	-73.92	0.75	**12067**	-68.76	0.70	**12855**	-65.36	0.66	**13320**	-62.48	0.63
12307	-94.18	0.95	**12397**	-73.89	0.75	**12234**	-68.71	0.70	**12301**	-65.33	0.66	**13112**	-62.46	0.63
14484	-89.77	0.91	**12068**	-73.86	0.75	**12854**	-68.69	0.70	**12227**	-65.31	0.66	**14278**	-62.41	0.63
12489	-88.42	0.89	**14281**	-73.85	0.75	**12392**	-68.46	0.69	**14482**	-65.02	0.66	**12303**	-62.36	0.63
14485	-86.46	0.87	**12946**	-73.80	0.75	**12391**	-68.40	0.69	**13138**	-64.93	0.66	**13411**	-62.33	0.63
12395	-84.93	0.86	**14393**	-73.76	0.75	**12011**	-68.31	0.69	**12616**	-64.87	0.66	**14404**	-62.32	0.63
12139	-84.64	0.86	**12223**	-73.48	0.74	**12658**	-68.23	0.69	**14479**	-64.84	0.66	**12702**	-62.31	0.63
13399	-84.28	0.85	**12053**	-73.47	0.74	**13953**	-68.22	0.69	**12510**	-64.84	0.66	**14431**	-62.30	0.63
12125	-83.71	0.85	**14489**	-73.30	0.74	**13303**	-68.09	0.69	**12924**	-64.81	0.66	**12513**	-62.30	0.63
12424	-83.24	0.84	**12463**	-72.88	0.74	**14480**	-68.05	0.69	**12098**	-64.80	0.66	**14488**	-62.25	0.63
12634	-82.68	0.84	**12839**	-72.63	0.73	**12992**	-67.77	0.69	**12951**	-64.75	0.66	**12799**	-62.22	0.63
12815	-82.61	0.84	**14546**	-72.50	0.73	**12435**	-67.66	0.68	**12117**	-64.73	0.65	**12433**	-62.20	0.63
12124	-81.75	0.83	**12615**	-72.46	0.73	**13116**	-67.62	0.68	**12467**	-64.60	0.65	**12795**	-62.19	0.63
12126	-80.88	0.82	**12701**	-72.27	0.73	**12729**	-67.54	0.68	**12012**	-64.58	0.65	**12648**	-62.18	0.63
12572	-80.75	0.82	**12225**	-71.82	0.73	**13216**	-67.51	0.68	**14296**	-64.48	0.65	**12732**	-62.18	0.63
12212	-79.87	0.81	**12861**	-71.82	0.73	**12531**	-67.34	0.68	**14293**	-64.42	0.65	**14294**	-62.14	0.63
12027	-79.68	0.81	**13249**	-71.74	0.73	**12641**	-67.34	0.68	**12217**	-64.37	0.65	**13142**	-62.12	0.63
12416	-78.21	0.79	**12408**	-71.72	0.73	**12490**	-67.22	0.68	**13177**	-64.33	0.65	**12071**	-62.12	0.63
12210	-78.15	0.79	**12440**	-71.70	0.73	**12866**	-67.11	0.68	**12432**	-64.32	0.65	**12515**	-62.10	0.63
12570	-78.13	0.79	**12914**	-71.67	0.73	**12887**	-67.09	0.68	**12016**	-64.16	0.65	**12296**	-62.10	0.63
12201	-78.05	0.79	**13300**	-71.62	0.72	**12909**	-67.03	0.68	**15338**	-64.13	0.65	**12670**	-62.07	0.63
12229	-77.88	0.79	**12168**	-71.59	0.72	**15209**	-67.02	0.68	**12766**	-64.03	0.65	**15518**	-62.06	0.63
12009	-77.83	0.79	**12020**	-71.49	0.72	**12131**	-66.96	0.68	**13286**	-63.94	0.65	**12657**	-62.00	0.63
12415	-77.67	0.79	**12010**	-71.47	0.72	**13302**	-66.93	0.68	**14412**	-63.93	0.65	**12913**	-61.97	0.63
12419	-77.65	0.79	**14267**	-71.43	0.72	**12707**	-66.85	0.68	**12605**	-63.79	0.65	**12827**	-61.97	0.63
12705	-77.46	0.78	**13184**	-71.35	0.72	**12309**	-66.65	0.67	**12297**	-63.73	0.64	**13194**	-61.92	0.63
12498	-77.38	0.78	**12438**	-71.28	0.72	**13309**	-66.57	0.67	**13183**	-63.63	0.64	**12401**	-61.89	0.63
12958	-77.26	0.78	**13205**	-71.24	0.72	**12123**	-66.48	0.67	**13281**	-63.62	0.64	**13318**	-61.84	0.63
14547	-76.79	0.78	**12568**	-70.94	0.72	**12029**	-66.46	0.67	**12524**	-63.46	0.64	**12680**	-61.81	0.63
12076	-76.71	0.78	**14453**	-70.87	0.72	**14277**	-66.41	0.67	**13075**	-63.36	0.64	**14301**	-61.74	0.62
12205	-76.45	0.77	**12460**	-70.86	0.72	**13206**	-66.39	0.67	**12715**	-63.20	0.64	**14468**	-61.73	0.62
12796	-76.45	0.77	**12703**	-70.78	0.72	**12973**	-66.34	0.67	**1037**	-63.16	0.64	**12045**	-61.67	0.62
12901	-76.32	0.77	**14465**	-70.60	0.71	**12198**	-66.31	0.67	**12412**	-63.15	0.64	**12403**	-61.58	0.62
12604	-76.17	0.77	**12226**	-70.57	0.71	**12699**	-66.29	0.67	**12218**	-63.14	0.64	**13009**	-61.39	0.62
12204	-76.06	0.77	**12538**	-70.42	0.71	**14424**	-66.23	0.67	**12878**	-63.05	0.64	**14457**	-61.35	0.62
12793	-75.74	0.77	**14297**	-70.24	0.71	**12906**	-66.05	0.67	**13407**	-63.04	0.64	**13153**	-61.34	0.62
12891	-75.67	0.77	**12077**	-70.09	0.71	**12090**	-65.99	0.67	**14474**	-62.97	0.64	**14341**	-61.34	0.62
14466	-75.64	0.77	**12437**	-70.00	0.71	**13408**	-65.94	0.67	**12704**	-62.97	0.64	**12449**	-61.33	0.62
12299	-75.38	0.76	**12222**	-69.85	0.71	**12871**	-65.79	0.67	**14380**	-62.92	0.64	**12000**	-61.30	0.62
12323	-75.31	0.76	**12426**	-69.85	0.71	**14472**	-65.79	0.67	**13374**	-62.88	0.64	**13199**	-61.26	0.62
12528	-75.15	0.76	**12730**	-69.59	0.70	**12900**	-65.78	0.67	**12110**	-62.87	0.64	**12047**	-61.26	0.62
12860	-74.80	0.76	**13369**	-69.36	0.70	**13080**	-65.74	0.67	**13101**	-62.84	0.64	**12610**	-61.25	0.62
12624	-74.49	0.75	**12052**	-69.35	0.70	**13030**	-65.70	0.66	**14270**	-62.76	0.64	**12736**	-61.18	0.62
12167	-74.42	0.75	**12496**	-69.26	0.70	**13137**	-65.64	0.66	**12042**	-62.69	0.63	**12700**	-61.09	0.62
13308	-74.40	0.75	**15207**	-69.15	0.70	**12722**	-65.63	0.66	**12826**	-62.68	0.63	**14409**	-61.05	0.62
12600	-74.39	0.75	**12731**	-69.13	0.70	**12206**	-65.55	0.66	**12406**	-62.58	0.63	**12517**	-61.04	0.62
12950	-74.28	0.75	**13063**	-68.84	0.70	**14268**	-65.55	0.66	**14429**	-62.57	0.63	**13284**	-60.99	0.62
14417	-74.21	0.75												

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para quatro proteínas alvo de *L. donovani*

N-Myristoyltransferase: 490 compostos

Energia do ligante: -50,4 KJ/mol

N- myristoyltransferase														
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
12050	-60.86	0.62	**12455**	-58.44	0.59	**12246**	-55.99	0.57	**12786**	-54.11	0.55	**14403**	-52.22	0.53
13043	-60.83	0.62	**13322**	-58.30	0.59	**12679**	-55.98	0.57	**12921**	-54.11	0.55	**14427**	-52.19	0.53
12650	-60.75	0.61	**14451**	-58.28	0.59	**14481**	-55.95	0.57	**12520**	-54.06	0.55	**13052**	-52.18	0.53
12195	-60.72	0.61	**12390**	-58.22	0.59	**14343**	-55.82	0.56	**12931**	-54.04	0.55	**12694**	-52.15	0.53
13370	-60.70	0.61	**12111**	-58.19	0.59	**12080**	-55.82	0.56	**12933**	-54.01	0.55	**13003**	-52.14	0.53
13103	-60.61	0.61	**12208**	-58.17	0.59	**12889**	-55.75	0.56	**13301**	-53.99	0.55	**14562**	-52.05	0.53
14381	-60.58	0.61	**13398**	-58.14	0.59	**12977**	-55.71	0.56	**12051**	-53.96	0.55	**12248**	-51.99	0.53
12420	-60.56	0.61	**12008**	-58.12	0.59	**12797**	-55.71	0.56	**12659**	-53.88	0.55	**12710**	-51.98	0.53
12070	-60.54	0.61	**12431**	-58.04	0.59	**12744**	-55.70	0.56	**12497**	-53.84	0.54	**14444**	-51.98	0.53
12915	-60.46	0.61	**12955**	-58.01	0.59	**12794**	-55.67	0.56	**12150**	-53.78	0.54	**12903**	-51.86	0.52
12526	-60.43	0.61	**12306**	-57.95	0.59	**13185**	-55.66	0.56	**14392**	-53.76	0.54	**12687**	-51.85	0.52
13413	-60.17	0.61	**13018**	-57.92	0.59	**12447**	-55.63	0.56	**13299**	-53.75	0.54	**13285**	-51.81	0.52
12400	-60.10	0.61	**12310**	-57.91	0.59	**12733**	-55.62	0.56	**14285**	-53.69	0.54	**13055**	-51.81	0.52
14329	-60.09	0.61	**12159**	-57.82	0.59	**12295**	-55.62	0.56	**12726**	-53.63	0.54	**12697**	-51.78	0.52
12199	-60.08	0.61	**14448**	-57.77	0.58	**13335**	-55.60	0.56	**13010**	-53.53	0.54	**12945**	-51.66	0.52
12639	-59.78	0.60	**12368**	-57.74	0.58	**14390**	-55.58	0.56	**12957**	-53.50	0.54	**12577**	-51.65	0.52
14439	-59.71	0.60	**15339**	-57.73	0.58	**14443**	-55.57	0.56	**13297**	-53.47	0.54	**12638**	-51.54	0.52
12002	-59.71	0.60	**14408**	-57.67	0.58	**12511**	-55.55	0.56	**12259**	-53.46	0.54	**12847**	-51.51	0.52
13294	-59.64	0.60	**14430**	-57.63	0.58	**12346**	-55.53	0.56	**12468**	-53.43	0.54	**12652**	-51.48	0.52
13113	-59.61	0.60	**12389**	-57.57	0.58	**12885**	-55.48	0.56	**15203**	-53.43	0.54	**12092**	-51.47	0.52
13952	-59.58	0.60	**13412**	-57.54	0.58	**14319**	-55.46	0.56	**12970**	-53.41	0.54	**12907**	-51.39	0.52
12886	-59.57	0.60	**13011**	-57.52	0.58	**12923**	-55.43	0.56	**13044**	-53.38	0.54	**13129**	-51.34	0.52
12245	-59.50	0.60	**12305**	-57.52	0.58	**12409**	-55.43	0.56	**12480**	-53.34	0.54	**12626**	-51.34	0.52
12745	-59.46	0.60	**12166**	-57.48	0.58	**13108**	-55.42	0.56	**12673**	-53.32	0.54	**12982**	-51.30	0.52
12840	-59.45	0.60	**14459**	-57.35	0.58	**14507**	-55.33	0.56	**12462**	-53.30	0.54	**12333**	-51.29	0.52
13104	-59.45	0.60	**14302**	-57.22	0.58	**14284**	-55.33	0.56	**13048**	-53.28	0.54	**12867**	-51.27	0.52
12018	-59.26	0.60	**12465**	-57.19	0.58	**12814**	-55.18	0.56	**12370**	-53.23	0.54	**12857**	-51.25	0.52
12654	-59.25	0.60	**12709**	-57.16	0.58	**15517**	-55.06	0.56	**14452**	-53.17	0.54	**12148**	-51.24	0.52
12972	-59.20	0.60	**13179**	-57.12	0.58	**14467**	-55.03	0.56	**13198**	-53.07	0.54	**12087**	-51.23	0.52
12883	-59.16	0.60	**12466**	-57.07	0.58	**13054**	-55.01	0.56	**12808**	-53.06	0.54	**12622**	-51.13	0.52
13114	-59.12	0.60	**12049**	-57.02	0.58	**12327**	-54.98	0.56	**12881**	-53.06	0.54	**13008**	-51.11	0.52
12188	-59.09	0.60	**12163**	-56.94	0.58	**12573**	-54.95	0.56	**12910**	-53.03	0.54	**14463**	-51.07	0.52
15516	-59.08	0.60	**14279**	-56.84	0.58	**13033**	-54.95	0.56	**12939**	-52.99	0.54	**12314**	-51.05	0.52
13162	-59.04	0.60	**13246**	-56.84	0.58	**12031**	-54.93	0.56	**12593**	-52.98	0.54	**14348**	-51.04	0.52
12798	-59.04	0.60	**14040**	-56.80	0.57	**12947**	-54.92	0.56	**12996**	-52.95	0.54	**12830**	-50.94	0.52
12618	-59.03	0.60	**12993**	-56.68	0.57	**13134**	-54.89	0.56	**12894**	-52.92	0.54	**12203**	-50.86	0.51
13190	-58.91	0.60	**12989**	-56.67	0.57	**12417**	-54.87	0.56	**12155**	-52.78	0.53	**13135**	-50.82	0.51
14273	-58.84	0.60	**14523**	-56.64	0.57	**14533**	-54.67	0.55	**13406**	-52.78	0.53	**12720**	-50.77	0.51
14395	-58.84	0.60	**12912**	-56.55	0.57	**12882**	-54.67	0.55	**12651**	-52.75	0.53	**12506**	-50.67	0.51
13110	-58.81	0.60	**12028**	-56.48	0.57	**13963**	-54.54	0.55	**12442**	-52.74	0.53	**14498**	-50.66	0.51
12688	-58.71	0.59	**12611**	-56.44	0.57	**12869**	-54.52	0.55	**13038**	-52.74	0.53	**14483**	-50.61	0.51
13005	-58.71	0.59	**12791**	-56.37	0.57	**14491**	-54.40	0.55	**12037**	-52.73	0.53	**13100**	-50.59	0.51
12689	-58.65	0.59	**15266**	-56.33	0.57	**12640**	-54.39	0.55	**12262**	-52.72	0.53	**12069**	-50.58	0.51
12461	-58.65	0.59	**12757**	-56.17	0.57	**13074**	-54.34	0.55	**12655**	-52.69	0.53	**12436**	-50.56	0.51
13013	-58.64	0.59	**12407**	-56.13	0.57	**12662**	-54.32	0.55	**12603**	-52.64	0.53	**13279**	-50.55	0.51
12868	-58.59	0.59	**12801**	-56.11	0.57	**14378**	-54.28	0.55	**12197**	-52.62	0.53	**13378**	-50.54	0.51
12428	-58.56	0.59	**13410**	-56.11	0.57	**12743**	-54.27	0.55	**14471**	-52.54	0.53	**12948**	-50.48	0.51
13032	-58.51	0.59	**12988**	-56.09	0.57	**13247**	-54.26	0.55	**12521**	-52.45	0.53	**13067**	-50.46	0.51
12941	-58.48	0.59	**12249**	-56.06	0.57	**12494**	-54.24	0.55						

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para quatro proteínas alvo de *L. donovani*

Ornithine decarboxylase: 2 compostos

Energia do ligante: -72,0 KJ/mol

Ornithine decarboxylase		
ID	**E (KJ/mol)**	**p_s**
12489	-92.13	1.00
12395	-83.69	0.91

Mitogen-activated protein kinase 3 (MPK3): 2 compostos

Energia do ligante: -112,6 KJ/mol

MPK3		
ID	**E (KJ/mol)**	**p_s**
12395	-128.47	1.00
14296	-114.35	0.89

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para quatro proteínas alvo de *L. donovani*

Pteridine reductase 1: 346 compostos

MPK3											
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
12793	-76.71	1.00	**12412**	-59.60	0.78	**12408**	-54.29	0.71	**13191**	-50.65	0.66
13309	-76.28	0.99	**14331**	-59.27	0.77	**12071**	-54.27	0.71	**12636**	-50.56	0.66
13184	-73.35	0.96	**12880**	-59.27	0.77	**13285**	-54.27	0.71	**12476**	-50.49	0.66
12462	-72.57	0.95	**13286**	-59.26	0.77	**12883**	-54.15	0.71	**12870**	-50.48	0.66
12796	-70.27	0.92	**13005**	-59.22	0.77	**12947**	-54.03	0.70	**12967**	-50.42	0.66
14546	-69.49	0.91	**13300**	-59.17	0.77	**12885**	-53.84	0.70	**14517**	-50.34	0.66
14451	-68.27	0.89	**12881**	-58.78	0.77	**12744**	-53.67	0.70	**12437**	-50.31	0.66
12463	-67.78	0.88	**12886**	-58.64	0.76	**12849**	-53.59	0.70	**12948**	-50.31	0.66
12889	-67.19	0.88	**13308**	-58.61	0.76	**12568**	-53.58	0.70	**12835**	-50.11	0.65
12093	-66.70	0.87	**12887**	-58.20	0.76	**13317**	-53.56	0.70	**13314**	-50.03	0.65
12731	-66.66	0.87	**12833**	-58.11	0.76	**13311**	-53.36	0.70	**12625**	-49.99	0.65
13006	-66.40	0.87	**12827**	-58.11	0.76	**14395**	-53.30	0.69	**12925**	-49.81	0.65
14352	-66.07	0.86	**14303**	-57.99	0.76	**12489**	-53.17	0.69	**12409**	-49.81	0.65
12907	-65.97	0.86	**13002**	-57.96	0.76	**12456**	-53.17	0.69	**12513**	-49.80	0.65
12838	-65.54	0.85	**14404**	-57.11	0.74	**14401**	-53.15	0.69	**12468**	-49.77	0.65
12634	-65.17	0.85	**12830**	-57.06	0.74	**12008**	-52.97	0.69	**12734**	-49.74	0.65
12639	-64.82	0.84	**12921**	-57.04	0.74	**12857**	-52.92	0.69	**12716**	-49.68	0.65
13319	-64.69	0.84	**13281**	-57.03	0.74	**13284**	-52.83	0.69	**14366**	-49.21	0.64
12616	-64.66	0.84	**13302**	-56.98	0.74	**12163**	-52.75	0.69	**12781**	-49.08	0.64
13011	-64.63	0.84	**12641**	-56.96	0.74	**13008**	-52.72	0.69	**13195**	-49.07	0.64
12877	-64.50	0.84	**15210**	-56.91	0.74	**12724**	-52.72	0.69	**13134**	-49.06	0.64
14402	-64.18	0.84	**12067**	-56.81	0.74	**12482**	-52.70	0.69	**12801**	-49.04	0.64
12859	-64.07	0.84	**12868**	-56.78	0.74	**14406**	-52.67	0.69	**12164**	-49.04	0.64
13301	-64.03	0.83	**14553**	-56.54	0.74	**12198**	-52.59	0.69	**13279**	-48.97	0.64
12194	-63.47	0.83	**12688**	-56.42	0.74	**13007**	-52.48	0.68	**13287**	-48.95	0.64
12737	-63.41	0.83	**12863**	-56.24	0.73	**13307**	-52.29	0.68	**12529**	-48.83	0.64
12854	-63.38	0.83	**12908**	-56.02	0.73	**12779**	-52.21	0.68	**12205**	-48.76	0.64
13183	-63.27	0.82	**14448**	-55.97	0.73	**12390**	-52.18	0.68	**13321**	-48.49	0.63
13406	-63.22	0.82	**14458**	-55.77	0.73	**12635**	-52.09	0.68	**12949**	-48.47	0.63
14516	-63.20	0.82	**12906**	-55.76	0.73	**12573**	-52.08	0.68	**12165**	-48.42	0.63
14403	-63.01	0.82	**12567**	-55.72	0.73	**13023**	-52.06	0.68	**14381**	-48.41	0.63
14329	-62.64	0.82	**13280**	-55.69	0.73	**12826**	-52.03	0.68	**12091**	-48.37	0.63
12933	-62.55	0.82	**12702**	-55.42	0.72	**12467**	-52.02	0.68	**12018**	-48.36	0.63
13001	-61.95	0.81	**12068**	-55.42	0.72	**14405**	-51.96	0.68	**13137**	-48.34	0.63
12618	-61.94	0.81	**12791**	-55.40	0.72	**12879**	-51.93	0.68	**14382**	-48.19	0.63
14413	-61.84	0.81	**12420**	-55.23	0.72	**12030**	-51.90	0.68	**12988**	-48.15	0.63
12204	-61.39	0.80	**12945**	-55.05	0.72	**12867**	-51.84	0.68	**12957**	-48.11	0.63
13012	-61.30	0.80	**12295**	-55.02	0.72	**15266**	-51.71	0.67	**12119**	-48.07	0.63
12939	-61.28	0.80	**14450**	-54.79	0.71	**13348**	-51.68	0.67	**12392**	-48.00	0.63
13185	-61.17	0.80	**12743**	-54.68	0.71	**15203**	-51.25	0.67	**12622**	-47.97	0.63
12400	-60.57	0.79	**12613**	-54.62	0.71	**12640**	-51.09	0.67	**12115**	-47.90	0.62
12099	-60.40	0.79	**12451**	-54.61	0.71	**12009**	-51.04	0.67	**12475**	-47.79	0.62
12831	-59.79	0.78	**12799**	-54.54	0.71	**12912**	-50.98	0.66	**12940**	-47.74	0.62
12878	-59.66	0.78	**12077**	-54.33	0.71	**12662**	-50.92	0.66	**14547**	-47.53	0.62

Sesquiterpenos lactonizados classificados como ativos por uma triagem virtual baseada na estrutura do receptor, para quatro proteínas alvo de *L. donovani*

Pteridine reductase 1: 346 compostos

MPK3											
ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s	ID	E (KJ/mol)	p_s
12638	-47.51	0.62	**14427**	-44.28	0.58	**12521**	-42.36	0.55	**13054**	-40.21	0.52
12411	-47.33	0.62	**13310**	-44.26	0.58	**13405**	-42.32	0.55	**12996**	-40.13	0.52
12065	-47.33	0.62	**13009**	-44.19	0.58	**14391**	-42.30	0.55	**12955**	-40.09	0.52
14520	-47.23	0.62	**12207**	-44.14	0.58	**12074**	-42.26	0.55	**12490**	-40.06	0.52
12045	-47.18	0.62	**12787**	-44.13	0.58	**12086**	-42.10	0.55	**12700**	-40.06	0.52
12471	-47.12	0.61	**12644**	-44.11	0.58	**12746**	-42.09	0.55	**13396**	-39.96	0.52
14407	-46.89	0.61	**13050**	-44.07	0.57	**13398**	-41.98	0.55	**12941**	-39.90	0.52
14348	-46.86	0.61	**12112**	-44.03	0.57	**13295**	-41.93	0.55	**14550**	-39.89	0.52
12929	-46.82	0.61	**15207**	-43.99	0.57	**12084**	-41.86	0.55	**12850**	-39.75	0.52
13199	-46.74	0.61	**14218**	-43.96	0.57	**12607**	-41.84	0.55	**14490**	-39.69	0.52
12070	-46.67	0.61	**12703**	-43.94	0.57	**12570**	-41.79	0.54	**12309**	-39.67	0.52
12973	-46.58	0.61	**12477**	-43.79	0.57	**13362**	-41.71	0.54	**12591**	-39.67	0.52
12021	-46.57	0.61	**13958**	-43.76	0.57	**12577**	-41.70	0.54	**12645**	-39.63	0.52
12888	-46.51	0.61	**13196**	-43.57	0.57	**12971**	-41.66	0.54	**13014**	-39.59	0.52
12839	-46.50	0.61	**12909**	-43.51	0.57	**12701**	-41.61	0.54	**14495**	-39.50	0.51
12484	-46.45	0.61	**12097**	-43.49	0.57	**13305**	-41.59	0.54	**12584**	-39.46	0.51
12874	-46.33	0.60	**12262**	-43.25	0.56	**12395**	-41.59	0.54	**12697**	-39.42	0.51
14386	-46.22	0.60	**12458**	-43.18	0.56	**12498**	-41.59	0.54	**12223**	-39.41	0.51
12011	-46.17	0.60	**12991**	-43.13	0.56	**1037**	-41.56	0.54	**13413**	-39.38	0.51
14524	-46.14	0.60	**12871**	-43.10	0.56	**12183**	-41.54	0.54	**13004**	-39.31	0.51
12329	-45.94	0.60	**12890**	-43.09	0.56	**14561**	-41.53	0.54	**12976**	-39.27	0.51
13298	-45.93	0.60	**14522**	-43.09	0.56	**15276**	-41.36	0.54	**12795**	-39.26	0.51
13003	-45.80	0.60	**13296**	-43.09	0.56	**12524**	-41.33	0.54	**14380**	-39.24	0.51
14518	-45.69	0.60	**12643**	-43.07	0.56	**13369**	-41.32	0.54	**12712**	-39.23	0.51
12391	-45.66	0.60	**12007**	-43.06	0.56	**15514**	-41.31	0.54	**14491**	-39.13	0.51
12435	-45.65	0.60	**12623**	-42.97	0.56	**12103**	-41.28	0.54	**12610**	-39.13	0.51
13397	-45.64	0.59	**12503**	-42.92	0.56	**12732**	-41.17	0.54	**14457**	-39.03	0.51
13033	-45.55	0.59	**12620**	-42.85	0.56	**12964**	-41.16	0.54	**12479**	-39.00	0.51
14410	-45.49	0.59	**12415**	-42.84	0.56	**14394**	-40.98	0.53	**12583**	-38.99	0.51
13182	-45.47	0.59	**12595**	-42.78	0.56	**12243**	-40.90	0.53	**12124**	-38.95	0.51
12683	-45.00	0.59	**12131**	-42.75	0.56	**14353**	-40.88	0.53	**13282**	-38.92	0.51
12417	-44.97	0.59	**13013**	-42.74	0.56	**14276**	-40.80	0.53	**13316**	-38.85	0.51
12837	-44.96	0.59	**12891**	-42.69	0.56	**12754**	-40.72	0.53	**12611**	-38.80	0.51
12166	-44.90	0.59	**12869**	-42.68	0.56	**12526**	-40.48	0.53	**15206**	-38.79	0.51
12780	-44.81	0.58	**13053**	-42.67	0.56	**14297**	-40.48	0.53	**13346**	-38.71	0.50
12893	-44.61	0.58	**14270**	-42.62	0.56	**14560**	-40.46	0.53	**12665**	-38.69	0.50
12750	-44.56	0.58	**12069**	-42.61	0.56	**12075**	-40.43	0.53	**14447**	-38.61	0.50
14424	-44.48	0.58	**12248**	-42.50	0.55	**12617**	-40.38	0.53	**12110**	-38.59	0.50
13010	-44.46	0.58	**14319**	-42.46	0.55	**12681**	-40.38	0.53	**12704**	-38.56	0.50
12992	-44.45	0.58	**13135**	-42.42	0.55	**12960**	-40.35	0.53	**13216**	-38.48	0.50
12367	-44.33	0.58	**12916**	-42.42	0.55	**14422**	-40.28	0.53	**14346**	-38.46	0.50
14497	-44.31	0.58	**12000**	-42.38	0.55	**12227**	-40.25	0.52	**14510**	-38.42	0.50
15516	-44.29	0.58	**12109**	-42.38	0.55						

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *L. donovani*

N- myristoyltransferase											
ID	**p**	**p_s**	**p_c**	**ID**	**p**	**p_s**	**p_c**	**ID**	**p**	**p_s**	**p_c**
12416	0.75	0.79	0.76	**12124**	0.55	0.83	0.65	**12433**	0.61	0.63	0.62
12131	0.80	0.68	0.76	**12705**	0.57	0.78	0.65	**14278**	0.61	0.63	0.62
14277	0.77	0.67	0.73	**12052**	0.62	0.70	0.65	**12511**	0.65	0.56	0.62
14473	0.58	0.99	0.73	**12568**	0.61	0.72	0.65	**14489**	0.55	0.74	0.62
12415	0.69	0.79	0.72	**13129**	0.72	0.52	0.65	**12642**	0.68	0.50	0.62
12234	0.72	0.70	0.71	**15517**	0.70	0.56	0.65	**14523**	0.64	0.57	0.62
12656	0.71	0.70	0.71	**12201**	0.56	0.79	0.65	**12791**	0.64	0.57	0.61
12906	0.72	0.67	0.70	**14444**	0.71	0.53	0.65	**12304**	0.54	0.75	0.61
14408	0.76	0.58	0.70	**12778**	0.73	0.50	0.65	**14412**	0.60	0.65	0.61
14485	0.59	0.87	0.69	**14268**	0.63	0.66	0.64	**12018**	0.62	0.60	0.61
13108	0.77	0.56	0.69	**14294**	0.65	0.63	0.64	**12000**	0.61	0.62	0.61
13134	0.77	0.56	0.69	**12466**	0.68	0.58	0.64	**13074**	0.65	0.55	0.61
12150	0.77	0.54	0.69	**12707**	0.62	0.68	0.64	**14378**	0.65	0.55	0.61
12461	0.75	0.59	0.69	**12855**	0.63	0.66	0.64	**12031**	0.64	0.56	0.61
14491	0.77	0.55	0.69	**14439**	0.66	0.60	0.64	**12212**	0.50	0.81	0.61
12515	0.72	0.63	0.69	**14468**	0.65	0.62	0.64	**12866**	0.57	0.68	0.61
13135	0.79	0.51	0.69	**14296**	0.63	0.65	0.64	**12827**	0.60	0.63	0.61
12521	0.78	0.53	0.69	**12162**	0.71	0.50	0.64	**13369**	0.56	0.70	0.61
12155	0.77	0.53	0.68	**13103**	0.65	0.61	0.64	**12861**	0.54	0.73	0.61
13184	0.66	0.72	0.68	**12921**	0.69	0.55	0.64	**12438**	0.55	0.72	0.61
12793	0.62	0.77	0.68	**12015**	0.71	0.51	0.64	**12435**	0.57	0.68	0.61
13137	0.68	0.66	0.67	**12125**	0.52	0.85	0.64	**12657**	0.60	0.63	0.61
13032	0.72	0.59	0.67	**12914**	0.59	0.73	0.64	**12703**	0.55	0.72	0.61
13110	0.72	0.60	0.67	**12634**	0.52	0.84	0.64	**14376**	0.67	0.50	0.61
12567	0.51	0.96	0.67	**13199**	0.64	0.62	0.63	**13116**	0.56	0.68	0.61
12658	0.66	0.69	0.67	**12111**	0.66	0.59	0.63	**12702**	0.59	0.63	0.61
13246	0.72	0.58	0.67	**14459**	0.66	0.58	0.63	**12010**	0.54	0.72	0.61
12395	0.56	0.86	0.67	**14267**	0.58	0.72	0.63	**14319**	0.63	0.56	0.61
12624	0.62	0.75	0.67	**12262**	0.69	0.53	0.63	**12222**	0.55	0.71	0.61
12077	0.64	0.71	0.67	**13294**	0.64	0.60	0.63	**12455**	0.61	0.59	0.60
12899	0.67	0.66	0.67	**12730**	0.59	0.70	0.63	**12611**	0.62	0.57	0.60
12593	0.74	0.54	0.67	**15518**	0.63	0.63	0.63	**15275**	0.66	0.50	0.60
12654	0.70	0.60	0.67	**12020**	0.57	0.72	0.63	**14467**	0.63	0.56	0.60
12901	0.61	0.77	0.67	**12745**	0.64	0.60	0.63	**13297**	0.64	0.54	0.60
13063	0.65	0.70	0.67	**12218**	0.62	0.64	0.63	**12700**	0.59	0.62	0.60
12480	0.74	0.54	0.67	**12370**	0.68	0.54	0.63	**12391**	0.55	0.69	0.60
12460	0.63	0.72	0.66	**14498**	0.69	0.51	0.63	**14430**	0.61	0.58	0.60
13153	0.69	0.62	0.66	**15516**	0.64	0.60	0.63	**12090**	0.56	0.67	0.60
12626	0.74	0.52	0.66	**12655**	0.68	0.53	0.63	**12616**	0.57	0.66	0.60
12323	0.60	0.76	0.66	**13067**	0.69	0.51	0.63	**12050**	0.59	0.62	0.60
12126	0.57	0.82	0.66	**12225**	0.57	0.73	0.63	**12724**	0.66	0.50	0.60
12087	0.74	0.52	0.66	**14281**	0.56	0.75	0.62	**13100**	0.65	0.51	0.60
12139	0.55	0.86	0.66	**12860**	0.55	0.76	0.62	**12801**	0.62	0.57	0.60
12401	0.67	0.63	0.66	**14474**	0.62	0.64	0.62	**12249**	0.62	0.57	0.60
12053	0.61	0.74	0.66	**12731**	0.58	0.70	0.62	**12441**	0.56	0.66	0.60
12898	0.63	0.70	0.65	**14380**	0.61	0.64	0.62	**12520**	0.62	0.55	0.60
14443	0.71	0.56	0.65	**12229**	0.53	0.79	0.62	**12722**	0.56	0.66	0.60
12826	0.66	0.63	0.65	**12210**	0.53	0.79	0.62	**14453**	0.53	0.72	0.60
12900	0.65	0.67	0.65	**12432**	0.60	0.65	0.62	**12070**	0.59	0.61	0.60
12110	0.66	0.64	0.65	**12076**	0.53	0.78	0.62	**12891**	0.50	0.77	0.60
13142	0.67	0.63	0.65	**12577**	0.67	0.52	0.62	**12736**	0.58	0.62	0.60
13399	0.54	0.85	0.65	**14482**	0.60	0.66	0.62	**13318**	0.58	0.63	0.60
13114	0.68	0.60	0.65	**14403**	0.67	0.53	0.62	**13320**	0.57	0.63	0.59
12839	0.60	0.73	0.65								

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *L. donovani*

N- myristoyltransferase											
ID	p	p_s	p_c	ID	p	p_s	p_c	ID	p	p_s	p_c
12868	0.60	0.59	0.59	**12973**	0.52	0.67	0.57	**12049**	0.53	0.58	0.55
12625	0.65	0.50	0.59	**12029**	0.52	0.67	0.57	**14301**	0.50	0.62	0.55
15209	0.55	0.68	0.59	**12639**	0.55	0.60	0.57	**13044**	0.55	0.54	0.55
14270	0.57	0.64	0.59	**12915**	0.55	0.61	0.57	**14341**	0.50	0.62	0.54
12622	0.64	0.52	0.59	**13075**	0.53	0.64	0.57	**12471**	0.57	0.51	0.54
12490	0.54	0.68	0.59	**12468**	0.59	0.54	0.57	**12986**	0.57	0.51	0.54
13104	0.59	0.60	0.59	**14483**	0.60	0.51	0.57	**12002**	0.51	0.60	0.54
13038	0.63	0.53	0.59	**12641**	0.51	0.68	0.57	**13013**	0.52	0.59	0.54
12467	0.56	0.65	0.59	**12246**	0.57	0.57	0.57	**13001**	0.57	0.50	0.54
14273	0.59	0.60	0.59	**14472**	0.52	0.67	0.57	**12047**	0.50	0.62	0.54
12123	0.55	0.67	0.59	**12743**	0.58	0.55	0.57	**12148**	0.56	0.52	0.54
14297	0.52	0.71	0.59	**14293**	0.52	0.65	0.57	**12673**	0.54	0.54	0.54
12988	0.60	0.57	0.59	**12008**	0.56	0.59	0.57	**12165**	0.57	0.50	0.54
15266	0.60	0.57	0.59	**14488**	0.53	0.63	0.57	**14279**	0.52	0.58	0.54
14404	0.57	0.63	0.59	**12618**	0.55	0.60	0.57	**12996**	0.54	0.54	0.54
12408	0.51	0.73	0.59	**13279**	0.60	0.51	0.57	**12159**	0.51	0.59	0.54
12680	0.57	0.63	0.59	**13398**	0.55	0.59	0.57	**12310**	0.51	0.59	0.53
12886	0.58	0.60	0.59	**12878**	0.53	0.64	0.57	**12894**	0.53	0.54	0.53
13216	0.53	0.68	0.59	**14550**	0.60	0.51	0.57	**12687**	0.54	0.52	0.53
14479	0.55	0.66	0.59	**13298**	0.53	0.63	0.56	**12327**	0.52	0.56	0.53
14481	0.60	0.57	0.59	**12694**	0.58	0.53	0.56	**12446**	0.55	0.51	0.53
12067	0.52	0.70	0.59	**12912**	0.56	0.57	0.56	**13011**	0.51	0.58	0.53
12977	0.60	0.56	0.58	**12757**	0.56	0.57	0.56	**12573**	0.52	0.56	0.53
14452	0.61	0.54	0.58	**12504**	0.52	0.63	0.56	**14463**	0.54	0.52	0.53
12559	0.63	0.51	0.58	**14395**	0.54	0.60	0.56	**12867**	0.54	0.52	0.53
14329	0.57	0.61	0.58	**12436**	0.59	0.51	0.56	**12462**	0.53	0.54	0.53
12887	0.53	0.68	0.58	**12016**	0.51	0.65	0.56	**12869**	0.52	0.55	0.53
12976	0.63	0.50	0.58	**12513**	0.52	0.63	0.56	**12442**	0.52	0.53	0.53
12510	0.54	0.66	0.58	**14302**	0.54	0.58	0.56	**14427**	0.53	0.53	0.53
12951	0.54	0.66	0.58	**12305**	0.54	0.58	0.56	**12317**	0.54	0.50	0.53
13088	0.62	0.51	0.58	**12449**	0.52	0.62	0.56	**13406**	0.52	0.53	0.53
13190	0.57	0.60	0.58	**12147**	0.59	0.50	0.56	**13006**	0.54	0.51	0.53
14471	0.61	0.53	0.58	**12069**	0.58	0.51	0.56	**12295**	0.50	0.56	0.53
12045	0.56	0.62	0.58	**12427**	0.58	0.51	0.56	**14343**	0.50	0.56	0.52
12042	0.55	0.63	0.58	**12166**	0.54	0.58	0.55	**12933**	0.51	0.55	0.52
12766	0.54	0.65	0.58	**13185**	0.55	0.56	0.55	**14532**	0.54	0.50	0.52
13183	0.55	0.64	0.58	**12882**	0.55	0.55	0.55	**13378**	0.53	0.51	0.52
12409	0.59	0.56	0.58	**12744**	0.55	0.56	0.55	**14548**	0.53	0.51	0.52
12368	0.58	0.58	0.58	**12420**	0.52	0.61	0.55	**12333**	0.52	0.52	0.52
14348	0.62	0.52	0.58	**12670**	0.51	0.63	0.55	**12662**	0.50	0.55	0.52
12406	0.55	0.63	0.58	**12051**	0.56	0.55	0.55	**13033**	0.50	0.56	0.52
12389	0.58	0.58	0.58	**14457**	0.51	0.62	0.55	**12881**	0.51	0.54	0.52
12909	0.52	0.68	0.58	**12346**	0.55	0.56	0.55	**12316**	0.53	0.50	0.52
12903	0.61	0.52	0.58	**12028**	0.54	0.57	0.55	**14276**	0.52	0.51	0.52
12797	0.59	0.56	0.58	**13055**	0.57	0.52	0.55	**12952**	0.52	0.51	0.52
12257	0.62	0.50	0.58	**12517**	0.51	0.62	0.55	**12857**	0.51	0.52	0.51
12704	0.55	0.64	0.58	**12074**	0.58	0.50	0.55	**14342**	0.52	0.50	0.51
12037	0.60	0.53	0.58	**12403**	0.51	0.62	0.55	**13003**	0.51	0.53	0.51
12871	0.53	0.67	0.58	**13048**	0.55	0.54	0.55	**14462**	0.51	0.50	0.51
13005	0.57	0.59	0.58	**12720**	0.57	0.51	0.55	**12838**	0.51	0.50	0.51
12245	0.56	0.60	0.58	**12852**	0.57	0.51	0.55	**12830**	0.50	0.52	0.51
14507	0.58	0.56	0.58	**12948**	0.57	0.51	0.55	**12349**	0.50	0.50	0.50
12947	0.59	0.56	0.58	**12638**	0.56	0.52	0.55	**12862**	0.50	0.50	0.50

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *L. donovani*

MPK3											
ID	p	p_s	p_c	ID	p	p_s	p_c	ID	p	p_s	p_c
14296	0.634	0.89	0.73	**12594**	0.662	0.51	0.61	**12694**	0.584	0.53	0.56
14468	0.65	0.86	0.73	**12457**	0.64	0.55	0.61	**14384**	0.56	0.56	0.56
12395	0.564	1.00	0.72	**15275**	0.658	0.51	0.60	**12663**	0.568	0.55	0.56
14491	0.77	0.62	0.72	**12537**	0.548	0.70	0.60	**14495**	0.572	0.54	0.56
14440	0.746	0.65	0.71	**12976**	0.63	0.55	0.60	**12751**	0.574	0.53	0.56
14467	0.628	0.86	0.71	**12511**	0.648	0.52	0.60	**12740**	0.566	0.54	0.56
12521	0.776	0.58	0.70	**12075**	0.576	0.64	0.60	**14488**	0.532	0.60	0.56
12778	0.726	0.64	0.69	**12712**	0.634	0.52	0.59	**14462**	0.51	0.63	0.55
13135	0.788	0.50	0.68	**12346**	0.546	0.68	0.59	**12620**	0.502	0.65	0.55
12323	0.602	0.83	0.68	**13045**	0.59	0.59	0.59	**12086**	0.564	0.53	0.55
13212	0.74	0.58	0.68	**14490**	0.614	0.55	0.59	**12933**	0.512	0.62	0.55
12108	0.73	0.57	0.67	**13057**	0.548	0.66	0.59	**13329**	0.514	0.62	0.55
12472	0.742	0.54	0.67	**12520**	0.624	0.53	0.59	**12944**	0.538	0.57	0.55
14474	0.616	0.76	0.67	**13369**	0.558	0.64	0.59	**14553**	0.5	0.63	0.55
12593	0.74	0.51	0.66	**14351**	0.564	0.63	0.59	**12374**	0.548	0.55	0.55
12270	0.7	0.58	0.66	**13040**	0.622	0.53	0.59	**12256**	0.542	0.56	0.55
12953	0.708	0.56	0.66	**12559**	0.628	0.52	0.59	**14472**	0.516	0.60	0.55
14557	0.632	0.68	0.65	**14267**	0.582	0.60	0.59	**13342**	0.558	0.52	0.54
13174	0.722	0.51	0.65	**14552**	0.61	0.55	0.59	**12435**	0.566	0.50	0.54
12667	0.706	0.54	0.64	**12474**	0.574	0.61	0.59	**12076**	0.532	0.56	0.54
14504	0.656	0.62	0.64	**12257**	0.624	0.51	0.58	**14447**	0.544	0.54	0.54
14481	0.598	0.72	0.64	**14524**	0.516	0.70	0.58	**12703**	0.548	0.53	0.54
13026	0.654	0.62	0.64	**14336**	0.6	0.55	0.58	**13327**	0.524	0.57	0.54
12501	0.7	0.54	0.64	**14494**	0.58	0.58	0.58	**14395**	0.54	0.54	0.54
12473	0.678	0.55	0.63	**14471**	0.61	0.53	0.58	**12529**	0.522	0.57	0.54
14500	0.66	0.57	0.63	**12353**	0.604	0.54	0.58	**13183**	0.546	0.52	0.54
14458	0.576	0.72	0.63	**13034**	0.554	0.62	0.58	**12669**	0.546	0.52	0.54
12481	0.608	0.66	0.63	**12903**	0.61	0.52	0.58	**13037**	0.554	0.50	0.54
12554	0.654	0.58	0.63	**12973**	0.52	0.68	0.58	**12742**	0.5	0.60	0.54
15210	0.586	0.69	0.62	**15276**	0.606	0.53	0.58	**12966**	0.526	0.55	0.53
12230	0.654	0.57	0.62	**14348**	0.616	0.51	0.58	**12459**	0.516	0.56	0.53
12965	0.662	0.55	0.62	**14434**	0.616	0.51	0.58	**12787**	0.54	0.52	0.53
13058	0.618	0.63	0.62	**12387**	0.616	0.50	0.57	**12503**	0.504	0.58	0.53
13208	0.628	0.60	0.62	**13190**	0.574	0.58	0.57	**12555**	0.512	0.56	0.53
14268	0.632	0.59	0.62	**12968**	0.51	0.69	0.57	**12964**	0.536	0.51	0.53
15208	0.6	0.64	0.62	**12436**	0.586	0.55	0.57	**13075**	0.532	0.52	0.53
14461	0.61	0.63	0.62	**12633**	0.572	0.57	0.57	**12173**	0.508	0.56	0.53
12139	0.548	0.73	0.62	**14554**	0.58	0.56	0.57	**12676**	0.538	0.50	0.53
14489	0.546	0.74	0.61	**14218**	0.578	0.56	0.57	**12016**	0.506	0.56	0.52
14483	0.604	0.63	0.61	**14374**	0.582	0.54	0.57	**15211**	0.512	0.53	0.52
14460	0.572	0.68	0.61	**12286**	0.568	0.57	0.57	**14293**	0.522	0.51	0.52
12140	0.648	0.54	0.61	**14475**	0.568	0.57	0.57	**12884**	0.504	0.53	0.51
14470	0.664	0.51	0.61	**15204**	0.57	0.56	0.57	**12897**	0.51	0.51	0.51
12977	0.596	0.63	0.61	**13378**	0.53	0.62	0.56				

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *L. donovani*

Pteridine reductase 1											
ID	p	p_s	p_c	ID	p	p_s	p_c	ID	p	p_s	p_c
13348	0.87	0.67	0.80	**12625**	0.65	0.65	0.65	**12888**	0.62	0.61	0.62
13184	0.66	0.96	0.77	**12868**	0.60	0.74	0.65	**13002**	0.54	0.76	0.61
12793	0.62	1.00	0.76	**14352**	0.53	0.86	0.65	**14348**	0.62	0.61	0.61
12779	0.79	0.68	0.75	**12329**	0.67	0.60	0.65	**12409**	0.59	0.65	0.61
13317	0.74	0.70	0.73	**12886**	0.58	0.76	0.65	**12988**	0.60	0.63	0.61
14403	0.67	0.82	0.72	**13183**	0.55	0.82	0.65	**13279**	0.60	0.64	0.61
13134	0.77	0.64	0.72	**12636**	0.63	0.66	0.64	**12887**	0.53	0.76	0.61
12906	0.72	0.73	0.72	**12618**	0.55	0.81	0.64	**13396**	0.66	0.52	0.61
14303	0.70	0.76	0.72	**12262**	0.69	0.56	0.64	**12030**	0.57	0.68	0.61
12131	0.80	0.56	0.72	**15210**	0.59	0.74	0.64	**12468**	0.59	0.65	0.61
12781	0.75	0.64	0.71	**13005**	0.57	0.77	0.64	**12870**	0.58	0.66	0.61
12835	0.75	0.65	0.71	**12568**	0.61	0.70	0.64	**12110**	0.66	0.50	0.60
12908	0.70	0.73	0.71	**12702**	0.59	0.72	0.64	**12008**	0.56	0.69	0.60
13319	0.63	0.84	0.71	**12415**	0.69	0.56	0.64	**12617**	0.65	0.53	0.60
12921	0.69	0.74	0.71	**12634**	0.52	0.85	0.64	**14319**	0.63	0.55	0.60
13135	0.79	0.55	0.70	**13185**	0.55	0.80	0.64	**12744**	0.55	0.70	0.60
14516	0.63	0.82	0.70	**14518**	0.66	0.60	0.64	**12839**	0.60	0.61	0.60
12833	0.67	0.76	0.70	**12838**	0.51	0.85	0.64	**12881**	0.51	0.77	0.60
12521	0.78	0.55	0.70	**12065**	0.64	0.62	0.63	**12067**	0.52	0.74	0.60
14402	0.60	0.84	0.69	**12622**	0.64	0.63	0.63	**12644**	0.62	0.58	0.60
12731	0.58	0.87	0.68	**12613**	0.59	0.71	0.63	**12467**	0.56	0.68	0.60
12115	0.71	0.62	0.68	**12584**	0.70	0.51	0.63	**12112**	0.62	0.57	0.60
12831	0.62	0.78	0.68	**13406**	0.52	0.82	0.63	**12645**	0.65	0.52	0.60
12462	0.53	0.95	0.68	**14458**	0.58	0.73	0.63	**12948**	0.57	0.66	0.60
14491	0.77	0.51	0.68	**13199**	0.64	0.61	0.63	**12458**	0.62	0.56	0.60
12367	0.73	0.58	0.67	**12877**	0.51	0.84	0.63	**12863**	0.52	0.73	0.60
14331	0.61	0.77	0.67	**14404**	0.57	0.74	0.63	**12780**	0.61	0.58	0.60
12791	0.64	0.72	0.67	**12947**	0.59	0.70	0.63	**12119**	0.58	0.63	0.60
12859	0.58	0.84	0.67	**15266**	0.60	0.67	0.63	**13362**	0.63	0.54	0.60
12826	0.66	0.68	0.67	**12743**	0.58	0.71	0.63	**14395**	0.54	0.69	0.60
12724	0.66	0.69	0.67	**13280**	0.57	0.73	0.63	**12912**	0.56	0.66	0.60
12077	0.64	0.71	0.67	**12577**	0.67	0.54	0.63	**13314**	0.56	0.65	0.59
12616	0.57	0.84	0.67	**13011**	0.51	0.84	0.63	**12070**	0.59	0.61	0.59
13137	0.68	0.63	0.66	**12018**	0.62	0.63	0.63	**13196**	0.61	0.57	0.59
12007	0.72	0.56	0.66	**14401**	0.59	0.69	0.62	**12641**	0.51	0.74	0.59
13195	0.67	0.64	0.66	**12801**	0.62	0.64	0.62	**12420**	0.52	0.72	0.59
14329	0.57	0.82	0.66	**12949**	0.62	0.63	0.62	**12482**	0.54	0.69	0.59
12639	0.55	0.84	0.66	**13346**	0.69	0.50	0.62	**12165**	0.57	0.63	0.59
12750	0.70	0.58	0.66	**14517**	0.60	0.66	0.62	**12716**	0.56	0.65	0.59
12099	0.59	0.79	0.66	**12933**	0.51	0.82	0.62	**12712**	0.63	0.51	0.59
12827	0.60	0.76	0.66	**13307**	0.58	0.68	0.62	**12567**	0.51	0.73	0.59
13006	0.54	0.87	0.65	**15516**	0.64	0.58	0.62	**12000**	0.61	0.55	0.59
12754	0.72	0.53	0.65	**12451**	0.57	0.71	0.62	**12830**	0.50	0.74	0.59
13001	0.57	0.81	0.65	**14522**	0.65	0.56	0.62	**12976**	0.63	0.51	0.59
12635	0.64	0.68	0.65	**12878**	0.53	0.78	0.62	**12867**	0.54	0.68	0.59

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **amastigotas** de *L. donovani*

Pteridine reductase 1											
ID	**p**	**p_s**	**p_c**	**ID**	**p**	**p_s**	**p_c**	**ID**	**p**	**p_s**	**p_c**
12893	0.59	0.58	0.59	**12069**	0.58	0.56	0.57	**14524**	0.52	0.60	0.55
14520	0.57	0.62	0.59	**12164**	0.53	0.64	0.57	**14427**	0.53	0.58	0.54
14382	0.56	0.63	0.59	**12074**	0.58	0.55	0.57	**12890**	0.53	0.56	0.54
14553	0.50	0.74	0.59	**12700**	0.59	0.52	0.57	**12484**	0.51	0.61	0.54
13191	0.54	0.66	0.58	**13023**	0.50	0.68	0.57	**12871**	0.53	0.56	0.54
12471	0.57	0.61	0.58	**12391**	0.55	0.60	0.57	**12909**	0.52	0.57	0.54
12638	0.56	0.62	0.58	**12021**	0.54	0.61	0.57	**13003**	0.51	0.60	0.54
12408	0.51	0.71	0.58	**14270**	0.57	0.56	0.56	**12964**	0.54	0.54	0.54
14353	0.61	0.53	0.58	**12513**	0.52	0.65	0.56	**12490**	0.54	0.52	0.54
14450	0.51	0.71	0.58	**12475**	0.53	0.62	0.56	**12996**	0.54	0.52	0.54
15276	0.61	0.54	0.58	**12529**	0.52	0.64	0.56	**12916**	0.52	0.55	0.53
12295	0.50	0.72	0.58	**14386**	0.54	0.60	0.56	**13033**	0.50	0.59	0.53
12734	0.54	0.65	0.58	**12662**	0.50	0.66	0.56	**12124**	0.55	0.51	0.53
12611	0.62	0.51	0.58	**12109**	0.57	0.55	0.56	**12837**	0.50	0.59	0.53
14490	0.61	0.52	0.58	**12086**	0.56	0.55	0.56	**12869**	0.52	0.56	0.53
12045	0.56	0.62	0.58	**12075**	0.58	0.53	0.56	**13013**	0.52	0.56	0.53
12929	0.56	0.61	0.58	**12703**	0.55	0.57	0.56	**12704**	0.55	0.50	0.53
15514	0.60	0.54	0.58	**12643**	0.55	0.56	0.56	**14447**	0.54	0.50	0.53
14380	0.61	0.51	0.58	**12166**	0.54	0.59	0.56	**14276**	0.52	0.53	0.53
12573	0.52	0.68	0.58	**12395**	0.56	0.54	0.56	**14297**	0.52	0.53	0.53
12925	0.54	0.65	0.58	**12091**	0.51	0.63	0.55	**12103**	0.52	0.54	0.52
12435	0.57	0.60	0.58	**12787**	0.54	0.58	0.55	**12503**	0.50	0.56	0.52
14218	0.58	0.57	0.58	**13298**	0.53	0.60	0.55	**13216**	0.53	0.50	0.52
12857	0.51	0.69	0.58	**13398**	0.55	0.55	0.55	**12620**	0.50	0.56	0.52
12476	0.53	0.66	0.57	**14495**	0.57	0.51	0.55	**12891**	0.50	0.56	0.52
13053	0.58	0.56	0.57	**12973**	0.52	0.61	0.55	**14457**	0.51	0.51	0.51
14550	0.60	0.52	0.57	**13369**	0.56	0.54	0.55	**12591**	0.51	0.52	0.51

Ornithine decarboxylase			
ID	**p**	**p_s**	**p_c**
12395	0.564	0.91	0.74
12043	0.646	0.74	0.69
12042	0.552	0.56	0.56

Sesquiterpenos lactonizados classificados como ativos por uma aproximação combinada de uma triagem virtual baseada na estrutura do ligante e baseada na estrutura do receptor, para **promastigotas** de *L. donovani*

N- myristoyltransferase			
ID	**p**	**p_s**	**p_c**
12395	0.51	0.86	0.63
12415	0.51	0.79	0.61
12915	0.55	0.61	0.57
14548	0.57	0.51	0.55

Ornithine decarboxylase			
ID	**p**	**p_s**	**p_c**
12395	0.51	0.91	0.65

MPK3			
ID	**p**	**p_s**	**p_c**
12395	0.51	1.00	0.69
12270	0.51	0.58	0.54

Pteridine reductase 1			
ID	**p**	**p_s**	**p_c**
13191	0.72	0.66	0.70
12415	0.51	0.56	0.52
12395	0.51	0.54	0.52

CONCLUSÕES

Mediante o uso de duas abordagens VS foi possível uma identificação preliminar de SLs potencialmente ativos contra as formas parasitas de *T. cruzi* e *L. donovani*, a partir de um banco de dados de sesquiterpenos lactonizados cadastrado no SISTEMATX.

A combinação das pontuações de probabilidades de VS baseada na estrutura do ligante e do receptor dando relevância aos modelos de Random forest que apresentam uma taxa menor de falso positivos, surge como uma metodologia interessante para esses tipos de estudos sendo que permite a identificação de moléculas promissoras e o possível mecanismo de ação

ANEXOS

ANEXOS I

Resultados

ID	SMILES
1033	[H]C1CC(=C)[C@]2(O)C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)C1
1036	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)C(C)C)[C@@H](OC(C)=O)C(C)=C1CC[C@@](C)(O)[C@H]21
1037	[H][C@@]12C[C@H](O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@]1([H])[C@H](CC2=C)OC(=O)C(=C)CO
11998	[H]\C1=C(C)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12000	[H]\C1=C(C)/C[C@H](O)C=C(C)C[C@@H](OC(=O)C(\CO)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12001	[H]\C1=C(C)/C[C@H](O)\C=C(C)\C[C@@H](OC(=O)C(\COC(C)=O)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12002	[H]\C1=C(C)/C[C@H](O)C=C(C)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12003	[H]\C1=C(C)/CC\C=C(C)\C[C@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12007	[H]\C1=C(C)/CC\C=C(C)\C[C@@H](OC(C)=O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12008	[H]\C1=C(C)/CCC=C(C)C[C@@H](OC(=O)C(\CO)=C/CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12009	[H]\C1=C(C)/CC\C=C(C)\C[C@@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12010	[H]\C1=C(C)/CC\C=C(C)\C[C@@H](OC(=O)[C@](C)(O)[C@@H](C)OC(C)=O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12011	[H]\C1=C(C)/CCC=C(C)C[C@@H](OC(=O)C(\CO)=C\C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12012	[H]\C1=C(C)/CCC=C(C)C[C@@H](OC(=O)C(\COC(C)=O)=C\C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12014	[H]\C1=C(C)/C[C@H](O)\C=C(C)\C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12015	[H]\C1=C(C)/C[C@H](O)\C=C(C)\C[C@@H](OC(=O)[C@@]2(C)O[C@H]2C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12016	[H]\C1=C(C)/C[C@H](O)\C=C(C)\C[C@@H](OC(=O)[C@@]2(CO)O[C@@H]2C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12017	[H]\C1=C(C)/C[C@H](O)\C=C(C)\C[C@@H](OC(=O)[C@@H](CO)[C@H](C)S)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12018	C[C@@H](O)C(=C)C(=O)O[C@@H]1C\C(C)=C\[C@@H](O)C\C(C)=C\[C@@H]2OC(=O)C(=C)[C@H]12
12019	[H]C1C\C(C)=C([H])\[C@@]2([H])OC(=O)C(=C)[C@]2(O)[C@@H](OC(C)=O)[C@H](OC(=O)C2(C)OC2C)C(C)=C1
12020	[H]C1C\C(C)=C([H])\[C@@]2([H])OC(=O)C(=C)[C@]2(O)[C@@H](OC(=O)C2(C)OC2C)[C@H](OC(C)=O)C(C)=C1
12021	[H]\C1=C(C)/C[C@@H](O)C=C(C)[C@@H](O)[C@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12022	[H]\C1=C(C)/[C@@H](O)C\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12024	[H]\C1=C(C)/[C@H](C\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(C)=O
12025	[H]\C1=C(C)/C(C\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)CC(C)C
12026	[H]\C1=C(C)/[C@H](O)CC=C(C)C[C@H](OC(C)=O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12027	[H]\C1=C(C)/[C@@H](O)CC=C(C)C[C@@H](OC(=O)C(\COC(=O)C(\C)=C\C)=C\C)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12028	[H]\C1=C(C)/[C@@H](O)CC=C(C)C[C@@H](OC(=O)C(\C)=C\C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12029	[H]\C1=C(C)/[C@@H](O)CC=C(C)C[C@@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12030	[H]\C1=C(C)/[C@@H](O)C\C=C(C)\[C@@H](O)[C@@H](OC(=O)[C@@H](C)CC)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12031	[H]\C1=C(C)/[C@@H](O)C\C=C(C)\[C@H](O)[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12032	[H]\C1=C(C)/CC\C=C(C)\[C@@H](O)C[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12033	[H]\C1=C(C)/CC\C=C(C)\[C@H](C[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)CC
12034	[H]\C1=C(C)/CC\C=C(C)\[C@H](C[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)C(C)C
12035	[H]\C1=C(C)/CC\C=C(C)\[C@H](C[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)CC(C)C
12036	[H]\C1=C(C)/CC\C=C(C)\[C@H](C[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)[C@@H](C)CC
12037	C[C@@H]1O[C@]1(C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)\C=C(C)\CC\C=C(C)\[C@H]1OC(C)=O
12042	[H]\C1=C(C)/[C@H](CC=C(C)C[C@@H](OC(=O)C(\C)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(C)=O
12043	[H]\C1=C(C)/[C@H](CC=C(C)C[C@@H](OC(=O)C(\CO)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(C)=O
12045	[H]\C1=C(C)/[C@@H](O)[C@H](O)\C=C(C)\C[C@@H](OC(=O)C(\COC(C)=O)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12046	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12047	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)[C@@H](C)CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12048	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)C(=C)CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12049	[H]\C1=C(CO)/CC\C=C(C)\C[C@@H](OC(=O)C(\C)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12050	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)C(=C)C[C@H](O)CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12051	[H]\C1=C(CO)/CC\C=C(CO)\C[C@H](OC(=O)C(C)C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12052	[H]\C1=C(CO)/CCC=C(CO)C[C@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12053	[H]\C1=C(CO)/CCC=C(CO)C[C@H](OC(=O)C(\C)=C/CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12054	[H]\C1=C(CO)/CC\C=C(CO)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12055	[H]\C1=C(CO)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12056	[H]\C1=C(/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(=O)CC(C)C
12057	[H]\C1=C(COC(=O)C=C(C)C)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12058	[H]\C1=C(COC(=O)C(C)C)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12059	[H]\C1=C(COC(=O)[C@@H](C)CC)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12060	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)C(C)C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12061	[H]\C1=C(COC(=O)C(=C)CO)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12062	[H]\C1=C(COC(=O)[C@]2(C)CO2)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12063	[H]\C1=C(COC(=O)[C@H](C)CO)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12064	[H]\C1=C(COC(=O)[C@](C)(O)CO)/CC\C=C(C)\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12065	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)[C@@H](C)CC)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12066	[H]\C1=C(CO)/CC\C=C(C)\C[C@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12067	[H]\C1=C(COC(C)=O)/CCC=C(COC(=O)C(C)C)C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12068	[H]\C1=C(COC(C)=O)/CCC=C(COC(=O)C(C)CC)C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12069	[H]\C1=C(COC(C)=O)/CC\C=C(CO)\C[C@@H](OC(=O)C(\C)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12070	[H]\C1=C(COC(C)=O)/CC\C=C(CO)\C[C@@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12071	[H]\C1=C(COC(C)=O)/CCC=C(COC(=O)C(C)=C)C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12074	[H]\C1=C(C)/C[C@H](O)\C=C(CO)\C[C@@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12075	[H]\C1=C(C)/C[C@H](O)C=C(CO)C[C@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12076	[H]\C1=C(C)/C[C@H](O)C=C(COC(C)=O)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12077	[H]\C1=C(C)/C[C@H](OC(C)=O)C=C(CO)C[C@@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12079	[H]\C1=C(C)/CC\C=C(CO)\C[C@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12080	[H]\C1=C(C)/CC\C=C(COC(C)=O)\C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12083	[H]\C1=C(C)/CC\C=C(\C[C@H](OC(=O)[C@@H](C)CC)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C(O)=O
12084	[H]\C1=C(CO)/CC\C=C(\C[C@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C=O
12085	[H]\C1=C(CO)/CC\C=C(\C[C@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C=O
12086	[H]\C1=C(C)/C[C@@H]2OC(=O)C(C[C@H](OC(=O)C(C)=C)[C@@]3([H])C(=C)C(=O)O[C@]13[H])=C2

ID	SMILES
12087	[H]\C1=C(C)/C[C@@H]2OC(=O)C(C[C@@H](OC(=O)C=C(C)C)[C@@]3([H])C(=C)C(=O)O[C@]13[H])=C2
12088	[H]\C1=C(C)/C[C@@H]2OC(=O)C(C[C@@H](OC(=O)[C@@H](C)CC)[C@@]3([H])C(=C)C(=O)O[C@]13[H])=C2
12089	[H]\C1=C(C)/C[C@@H]2OC(=O)C(C[C@@H](O)[C@@]3([H])C(=C)C(=O)O[C@]13[H])=C2
12090	[H]\C1=C(C)/C[C@@H](O[C@]2([H])O[C@H](COC(C)=O)[C@@H](O)[C@H](O)[C@H]2O)\C=C(\CC[C@@]2([H])C(=C)C(=O)O[C@]12[H])C(O)=O
12091	[H]\C1=C(/CC\C=C(C)\C[C@@H](OC(=O)\C=C(/C)CC)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C(O)=O
12092	[H]\C1=C(/CC\C=C(COC(=O)[C@H](C)CC)\C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C(O)=O
12093	[H]\C1=C(CO)/CC[C@@]2([H])O[C@]2(COC(C)=O)C[C@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12094	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(C)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12095	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(C)C[C@@H](OC(=O)[C@@]2(C)O[C@H]2C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12096	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(CO)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12097	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(COC(C)=O)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12098	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(COC(C)=O)C[C@@H](OC(=O)[C@@]2(C)O[C@H]2C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12099	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(COC(C)=O)C[C@@H](OC(=O)C(C)=C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12100	[H]\C1=C2/CC[C@@]3([H])O[C@@]3(C[C@H](OC(=O)C(C)=C)[C@@]3([H])C(=C)C(=O)O[C@]13[H])[C@H](O)OC2
12103	[H]\C1=C2/CC[C@@]3([H])O[C@@]3(C[C@H](OC(=O)C(=C)CO)[C@@]3([H])C(=C)C(=O)O[C@]13[H])[C@H](O)OC2
12104	[H][C@]12O[C@]1(C)CC\C=C(C)\CC[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12105	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@H](O)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12108	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@@H](OC(=O)C(\C)=C\CO)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12109	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@@H](OC(=O)C(\CO)=C\CO)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12110	[H][C@]12O[C@]1(CO)CC\C=C(C)\C[C@@H](OC(=O)C(\C)=C\CO)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12111	[H][C@]12O[C@]1(CO)CC\C=C(C)\C[C@@H](OC(=O)C(\C)=C\COC(=O)C[C@H](O)CCCCCCCCCCCCCCCC)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12112	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@H](OC(C)=O)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12114	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12115	[H][C@]12O[C@]1(C)C[C@H](O)\C=C(C)\C[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]21[H]
12117	[H][C@]12O[C@]1(COC(C)=O)\C=C/C=C(/[C@@H](O)[C@@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@]1([H])C(=C)C(=O)O[C@]21[H])C(=O)OC
12118	[H][C@]12O[C@]1(C)C[C@H]1OC(=O)C(C[C@H](OC(=O)C(C)=C)[C@@]3([H])C(=C)C(=O)O[C@]23[H])=C1
12119	[H][C@]12O[C@]1(C)C[C@H]1OC(=O)C(C[C@H](OC(=O)C=C(C)C)[C@@]3([H])C(=C)C(=O)O[C@]23[H])=C1
12123	[H][C@]12O[C@]1(C)C[C@H](OC(C)=O)\C=C(C)\C[C@H](OC(=O)C(C)=C)C1=C(COC(C)=O)C(=O)O[C@]21[H]
12124	[H][C@]12O[C@]1(C)C[C@H](OC(C)=O)\C=C(C)\C[C@H](OC(C)=O)C1=C(COC(C)=O)C(=O)O[C@]21[H]
12125	[H][C@]12O[C@]1(C)C[C@H](OC(C)=O)\C=C(C)\C[C@H](OC(=O)[C@]1(C)CO1)C1=C(OC(C)=O)C(=O)O[C@]21[H]
12126	[H][C@]12O[C@]1(C)CC\C=C(C)\C[C@H](OC(=O)C(\C)=C\C)C1=C(COC(C)=O)C(=O)O[C@]21[H]
12128	[H][C@]12OC(=O)[C@H](C)[C@]1([H])[C@H](O)CC1=C[C@@H](C\C(C)=C\2)OC1=O
12129	[H]\C1=C(C)/CC\C=C(C)\CC[C@@]2([H])[C@H](C)C(=O)O[C@]12[H]
12130	[H]\C1=C(C)/[C@@H](O)C\C=C(C)\CC[C@@]2([H])[C@@H](C)C(=O)O[C@]12[H]
12131	[H]\C1=C(C)/[C@@H](C\C=C(C)\C[C@H](OC(C)=O)[C@@]2([H])[C@@H](C)C(=O)O[C@]12[H])OC(C)=O
12133	[H]\C1=C(C)/CC\C=C(C)\C[C@H](O)[C@@]2([H])[C@H](C)C(=O)O[C@]12[H]
12134	[H]\C1=C(C)/CC\C=C(C)\C[C@H](OC(C)=O)[C@@]2([H])[C@H](C)C(=O)O[C@]12[H]
12135	[H]\C1=C(C)/CC\C=C(C)\C[C@H](O)[C@]2([H])[C@]1([H])OC(=O)[C@]2(C)O
12137	[H]\C1=C(C)/CC\C=C(C)\[C@H](C[C@@]2([H])[C@H](C)C(=O)O[C@]12[H])OC(C)=O
12138	[H]\C1=C(CO)/CC\C=C(C)\CC[C@@]2([H])[C@H](C)C(=O)O[C@]12[H]
12139	[H]\C1=C(C)/CC\C=C(\CC[C@@]2([H])[C@H](C)C(=O)O[C@]12[H])OC(=O)C[C@]1([H])O[C@H](COC(C)=O)[C@@H](O)[C@H](O)[C@H]1O
12140	[H]\C1=C(C)\CC[C@@]2([H])O[C@]2(C)[C@](Cl)([H])C[C@@]2([H])[C@H](C)C(=O)O[C@]12[H])OC(C)=O
12141	[H][C@]12O[C@]1(C)CC\C=C(C)\CC[C@@]1([H])[C@H](C)C(=O)O[C@]21[H]
12142	[H][C@]12O[C@]1(C)[C@@H](O)[C@H](O)\C=C(C)\C[C@@H](OC(=O)C(\C)=C\C)[C@@]1([H])[C@H](C)C(=O)O[C@]21[H]
12143	[H][C@]12O[C@]1(C)[C@@H](O)C\C=C(C)\C[C@@H](OC(=O)C(\C)=C\C)[C@@]1([H])[C@H](C)C(=O)O[C@]21[H]
12145	[H][C@]12O[C@]1(C)CC\C=C(C)\[C@H](C[C@@]1([H])[C@H](C)C(=O)O[C@]21[H])OC(C)=O
12146	[H][C@]12O[C@]1(C)C[C@H]1OC(=O)C(C[C@H](OC(=O)C(C)=C)[C@@]3([H])[C@H](C)C(=O)O[C@]23[H])=C1
12147	C[C@@H]1O[C@]1(C)C(=O)O[C@@H]1C\C(C)=C\C[C@H](OC(C)=O)\C(C)=C\[C@@H]2OC(=O)C(=C)[C@H]12
12148	C[C@H](OC(C)=O)[C@](C)(O)C(=O)O[C@@H]1C\C(C)=C\C[C@H](OC(C)=O)\C(C)=C\[C@@H]2OC(=O)C(=C)[C@H]12
12150	C[C@@H]1O[C@]1(C)C(=O)O[C@@H]1C\C(C)=C\C[C@H](O)\C(C)=C\[C@@H]2OC(=O)C(=C)[C@H]12
12151	[H][C@@]12OC(=O)[C](=C)=C1C(=O)C\C(C)=C\CC\C(C)=C\2
12154	[H]\C1=C(C)/[C@H](O)C\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12155	[H]\C1=C(C)/[C@@H](C\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(C)=O)OC(C)=O
12156	[H]\C1=C(C)/[C@H](O)C\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(C)=O
12157	[H]C1C\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@H](O)\C([H])=C1/C
12159	[H]\C1=C(C)/[C@H](C\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O)OC(C)=O
12160	[H]\C1=C(CO)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12161	[H]\C1=C(COC(C)=O)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12162	[H]\C1=C(COC(C)=O)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(C)=O
12163	[H]\C1=C(COC(=O)C(\COC(C)=O)=C/C)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12164	[H]\C1=C(COC(=O)C(\CO)=C/C)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12165	[H]\C1=C(COC(=O)C(\C)=C\CO)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12166	[H]\C1=C(CO)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(=O)C(\C)=C\CO
12167	[H]\C1=C(COC(C)=O)/CC\C=C(C)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(=O)[C@](C)(O)CO
12168	[H]\C1=C(COC(C)=O)/CC\C=C(COC(C)=O)\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O
12169	[H]\C1=C(COC(C)=O)/CC\C=C(\C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1O)C=O
12170	[H]C1=C2CC\C=C(C)\C[C@]3([H])OC(=O)C(=C)[C@@]3([H])[C@@H]1OC2=O
12171	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(C)C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(C)=O
12173	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(C)C[C@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(=O)C(\C)=C/C
12174	[H]C1=C2CC[C@]3([H])O[C@@]3(C)C[C@]3([H])OC(=O)C(=C)[C@@]3([H])[C@@H]1OC2=O
12175	[H]C1=C2[C@@H](C[C@]3([H])O[C@@]3(C)C[C@]3([H])OC(=O)C(=C)[C@@]3([H])[C@@H]1OC2=O)OC(C)=O
12176	[H]C1=C2C(=O)O[C@H]1[C@]1([H])C(=C)C(=O)O[C@@]1([H])C[C@]1(C)O[C@@]1([H])[C@]1([H])O[C@]21[H]
12177	[H][C@@]12O[C@@]1(C)CC\C=C(C)\C[C@]1([H])OC(=O)C(=C)[C@@]1([H])[C@@H]2O
12180	[H]\C1=C(C)/[C@H](O)C\C=C(C)\C[C@]2([H])OC(=O)[C@H](C)[C@@]2([H])[C@@H]1O
12182	[H]\C1=C(C)/CC\C=C(C)\C[C@]2([H])OC(=O)[C@H](C)[C@@]2([H])[C@@H]1O
12183	[H]\C1=C(CO)/CC\C=C(C)\C[C@]2([H])OC(=O)[C@@H](C)[C@@]2([H])[C@@H]1O

ID	SMILES
12185	[H]C1=C2[C@H]3O[C@H]3[C@]3([H])O[C@@]3(C)C[C@]3([H])OC(=O)[C@H](C)[C@@]3([H])[C@@H]1OC2=O
12186	[H]C1=C2[C@@H](C[C@]3([H])O[C@@]3(C)C[C@]3([H])OC(=O)[C@H](C)[C@@]3([H])[C@@H]1OC2=O)OC(C)=O
12187	[H]\C1=C(C)/CC\C=C(C)\C[C@@]2([H])OC(=O)C(=C)[C@@]2([H])C1
12188	[H]\C1=C(C)/CC\C=C(C)\C[C@@]2([H])OC(=O)C(=C)[C@@]2([H])[C@@H]1OC(=O)C=C(C)C
12189	[H]\C1=C(C)/CC[C@@]2([H])O[C@]2(C)C[C@@]2([H])OC(=O)C(=C)[C@@]2([H])C1
12190	[H][C@]12C[C@]3([H])C(=C)C(=O)O[C@]3([H])C\C(C)=C\CC[C@]1(C)O2
12193	[H]\C1=C(C)/CC\C=C(C)\C[C@@]2([H])OC(=O)[C@@H](C)[C@@]2([H])C1
12194	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@H](O)\C(=C/CC\C(C)=C\2)C(=O)OC
12195	[H][C@]12OC(=O)C(=C)[C@]1([H])C(OC(=O)C(\C)=C/C)[C@H](OC(C)=O)\C(=C/CC\C(C)=C\2)C(=O)OC
12196	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(OC(C)=O)[C@H](C)OC(C)=O)[C@H](OC(C)=O)\C(=C/CC\C(C)=C\2)C(=O)OC
12197	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(O)[C@H](C)OC(C)=O)[C@H](OC(=O)C(C)C)\C(=C/CC\C(C)=C\2)C(=O)OC
12198	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@H](C)OC(C)=O)[C@H](OC(C)=O)\C(=C/CC\C(C)=C\2)C(=O)OC
12199	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@H](C)OC(C)=O)[C@H](OC(=O)CC(C)(C)O)\C(=C/CC\C(C)=C\2)C(=O)OC
12200	[H][C@@]12C[C@@H](C)CC\C=C(/[C@@H](OC(C)=O)[C@@H](OC(=O)[C@@](C)(O)[C@H](C)O)[C@@]1([H])C(=C)C(=O)O2)C(=O)OC
12201	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@@H](C)O)[C@H](OC(=O)[C@@H](C)[C@@H](C)O)\C(=C/CC\C(C)=C\2)C(=O)OC
12203	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)C(C)=O)[C@H](OC(=O)CCC)\C(=C/CC\C(C)=C\2)C(=O)OC
12204	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(O)C(C)=O)[C@H](OC(=O)C(C)C)\C(=C/CC\C(C)=C\2)C(=O)OC
12205	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@H](C)O)[C@H](OC(C)=O)\C(=C/CC\C(C)=C\2)C(=O)OC
12206	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@H](OC(=O)[C@H](C)[C@H](C)O)\C(=C/CC\C(C)=C\2)C(=O)OC
12207	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(O)[C@H](C)O)[C@H](OC(=O)[C@@H](C)CC)\C(=C/CC\C(C)=C\2)C(=O)OC
12208	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@H](C)O)[C@H](OC(=O)C(C)C)\C(=C/CC\C(C)=C\2)C(=O)OC
12209	[H][C@]12OC(=O)C(=C)[C@]1([H])C(OC(=O)[C@](C)(O)C(C)=O)[C@H](OC(=O)[C@H](C)[C@H](C)O)\C(=C/CC\C(C)=C\2)C(=O)OC
12210	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(C)=O)\C(=C/2)C(=O)OC
12211	[H]C1OC(=O)C(=C)C([H])[C@@H](C\C(=C/CC\C(C)=C\1)C(=O)OC)OC(=O)[C@](C)(O)[C@H](C)OC(C)=O
12212	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@@H]1C)[C@H](OC(C)=O)\C(=C/CC\C=C\2)C(=O)OC
12213	[H][C@@]12CC[C@@]3(C)CCC=C(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12214	[H][C@@]12CC[C@@]3(C)[C@@H](O)CC=C(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12216	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@@]1(C)[C@@H](O)CC=C(C)[C@]21[H])OC(C)=O
12217	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)CCC=C(C)[C@]21[H])OC(=O)C(\C)=C/C
12218	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)CCC=C(C)[C@]21[H])OC(=O)[C@@]1(C)O[C@H]1C
12220	[H][C@@]12CC[C@@]3(C)[C@@H](O)C[C@H]4O[C@@]4(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12222	[H][C@@]12C[C@@H](O)[C@@]3(C)[C@H](C[C@@H]4O[C@]4(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C/C
12223	[H][C@@]12[C@@H](O)C[C@@]3(C)[C@H](CC[C@]4(CO4)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(\CO)=C/C
12225	[H][C@@]12[C@@H](O)C[C@@]3(C)[C@H](CC[C@@]4(CO4)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C/C
12226	[H][C@@]12[C@H](C[C@@]3(C)[C@H](CC[C@@]4(CO4)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(\CO)=C/C)OC(=O)C(\C)=C/C
12227	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@@]1(C)[C@H](CC[C@@]3(CO3)[C@]21[H])OC(=O)C(\CO)=C/C)OC(=O)[C@@H](C)CC
12229	[H][C@@]12[C@@H](O)C[C@@]3(C)[C@H](CC[C@@]4(CO4)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)[C@@]1(C)O[C@H]1C
12230	[H]C1CCC(C)=C2[C@@]3([H])OC(=O)C(=C)[C@]3([H])CC[C@@]12C
12232	[H]C1CCC(C)=C2[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](O)C[C@@]12C
12233	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)[C@H](O)CCC(C)=C21)OC(=O)C(\C)=C/C
12234	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)[C@H](O)CCC(C)=C21)OC(=O)[C@@]1(C)O[C@H]1C
12236	[H][C@@]12CC[C@@]3(C)[C@@H](O)CC(=O)C(C)=C3[C@@]1([H])OC(=O)C2=C
12237	[H]C1C[C@@]2(C)[C@H](O)CCC(=C)[C@]2(O)[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12238	[H]C1C[C@@]2(C)C([H])CCC(=C)[C@]2(O)[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12239	[H]C1CCC(=C)[C@]2(O)[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](O)C[C@@]12C
12240	[H][C@@]12CC[C@@]3(C)CCCC(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12241	[H][C@@]12CC[C@@]3(C)[C@H](O)CCC(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12242	[H][C@@]12CC[C@@]3(C)[C@@H](O)C[C@@H](O)C(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12243	[H][C@@]12CC[C@@]3(C)[C@H](O)C[C@H](O)[C@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12244	[H][C@@]12[C@@H](O)C[C@@]3(C)[C@@H](O)CCC(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12245	[H][C@@]12[C@@H](O)C[C@@]3(C)[C@H](CCC(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(=C)CO
12246	[H][C@@]12[C@H](C[C@@]3(C)[C@@H](O)CCC(=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(=C)CO
12248	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)[C@H](O)CCC(=C)[C@]21[H])OC(=O)C(\C)=C/C
12249	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)C[C@@]1(C)[C@H](CCC(=C)[C@]21[H])OC(=O)C(C)C
12253	[H]C1C[C@@]2(C)C=CC=C(C)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12254	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@@]1(C)C=CC=C(C)[C@]21[H])OC(=O)CC(C)C
12255	[H][C@@]12CC[C@@]3(C)C=CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12256	[H]C1C[C@@]2(C)C(=O)C=C[C@@](C)(O)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12257	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@@]1(C)C(=O)C=C[C@@](C)(OC(C)=O)[C@]21[H])OC(C)=O
12258	[H][C@@]12CC[C@@]3(C)[C@H](O)CC[C@@](C)(O)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12259	[H][C@@]12CC[C@@]3(C)[C@H](O)CC[C@](O)(COC(=O)C(C)=C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12260	[H]C1CC[C@@](C)(O)[C@]2([H])[C@@]3([H])OC(=O)C(=C)[C@]3([H])CC[C@@]12C
12262	[H][C@@]12CC[C@@]3(C)[C@H](C[C@@H](OC(C)=O)C(C)=C3[C@]1([H])OC(=O)C2=C)OC(C)=O
12263	[H][C@@]12CC[C@@]3(C)[C@H](O)CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12264	[H][C@@]12C[C@H](OC(C)=O)[C@@]3(C)[C@H](CC=C(C)[C@@]3([H])[C@@]1([H])OC(=O)[C@@H]2C)OC(=O)C(\C)=C/C
12265	[H]C1C=C(C)[C@]2([H])[C@]3([H])OC(=O)[C@@]([H])(C)[C@]3([H])CC[C@@]2(C)[C@@H]1O
12268	[H][C@]1(C)C(=O)O[C@]2([H])C3=C(C)CC[C@H](O)[C@]3(C)CC[C@@]12[H]
12270	[H][C@@]12CC[C@@]3(C)C(=O)CCC(C)=C3[C@@]1([H])OC(=O)[C@@H]2C
12272	[H][C@@]12CC[C@@]3(C)CCC(=O)C(C)=C3[C@@]1([H])OC(=O)[C@H]2C
12273	[H]C1C[C@@]2(C)C=CC(=O)C(C)=C2[C@@]2([H])OC(=O)[C@@]([H])(C)[C@]12[H]
12276	[H][C@@]1(C)C(=O)O[C@]2([H])C3=C(C)C(=O)C=C[C@]3(C)C[C@H](O)[C@@]12[H]
12277	[H]C1C[C@@]2(C)C=CC(=O)C(C)=C2[C@@]2([H])OC(=O)[C@](C)(O)[C@]12[H]
12279	[H][C@@]12CCC3=C(C)C=C(O)C(C)=C3[C@]1([H])OC(=O)[C@H]2C
12280	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])CC[C@@]1(C)CCCC(=C)[C@]21[H]
12281	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])CC[C@@]1(C)[C@H](O)CCC(=C)[C@]21[H]
12282	[H][C@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])CC[C@@]1(C)[C@H](O)CCC(=C)[C@]21O
12284	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])CC[C@@]1(C)[C@@H](O)C[C@@H](O)C(=C)[C@]21[H]

ID	SMILES
12286	[H]C1C[C@@]2(C)C(=O)C=C[C@@](C)(O)[C@]2([H])[C@@]2([H])OC(=O)[C@]([H])(C)[C@]12[H]
12287	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@@H](O)C[C@@]1(C)C(=O)C=C[C@@](C)(O)[C@]21[H]
12288	[H][C@@]12CC[C@@]3(C)CCC[C@@](C)(O)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12289	[H][C@@]12CC[C@@]3(C)[C@@H](O)CC(=O)[C@@]([H])(C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12291	[H][C@]12[C@H](C)C(=O)O[C@]1([H])[C@@]1([H])[C@@]([H])(C)C(=O)CC[C@]1(C)C[C@@H]2O
12295	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](O)C=C(C)[C@]21O)OC(=O)C(\C)=C/C
12296	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](O)C=C(C)[C@]21[H])OC(=O)C(\C)=C\C
12297	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](O)C=C(C)[C@]21[H])OC(=O)C(\C)=C/COC(C)=O
12298	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](OC(C)=O)C=C(C)[C@]21O)OC(=O)C(\C)=C/C
12299	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](OC(C)=O)C=C(C)[C@]21[H])OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C
12300	[H][C@]12CC[C@@]3(C)O[C@@]13[C@H]1OC(=O)C(=C)[C@@H]1CC[C@H]2C
12301	[H]C1[C@@H](OC(=O)C(\C)=C/CO)[C@@H]2[C@H](OC(=O)C2=C)[C@]2([H])[C@@]([H])([C@H](O)[C@H]3O[C@@]23C)[C@@]11CO1
12302	[H][C@@]12CC=C(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](O)C[C@@]21CO1
12303	[H][C@@]12CC=C(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]21CO1)OC(=O)C(\CO)=C/CO
12304	[H][C@@]12CC=C(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]21CO1)OC(=O)C1=COC=C1
12305	C\C=C(\C)C(=O)O[C@@H]1[C@@H]2OC(=O)C2=C2CC=C(C)[C@]2(C)[C@H]2OC(=O)C(=C)[C@H]12
12306	[H]C1[C@@H](OC(=O)C(\C)=C/C)[C@@H]2[C@H](OC(=O)C2=C)[C@]2(O)[C@@]([H])([C@H](O)[C@H]3O[C@@]23C)[C@@]11CO1
12307	[H]C1[C@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C/C)[C@@H]2[C@@H](OC(=O)C2=C)[C@@]2([H])[C@@]([H])([C@@H](OC(C)=O)[C@H]3O[C@@]23C)[C@@]11CO1
12308	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](OC(=O)[C@@H](C)CC)[C@@H](O)[C@]3(CO3)[C@]1([H])[C@H](O)[C@H]1O[C@@]21C
12309	[H]C1[C@@H](OC(=O)CC(C)C)[C@@H]2[C@H](OC(=O)C2=C)[C@]2([H])[C@@]([H])([C@H](O)[C@H]3O[C@@]23C)[C@@]11CO1
12310	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](OC(=O)C(\C)=C\C)[C@@H](O)[C@]3(CO3)[C@]1([H])[C@H](O)[C@H]1O[C@@]21C
12314	[H][C@@]12[C@H]3OC(=O)C(CO)=C3CC[C@@]1(C)CCC[C@@]2(C)O
12316	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@@H](O)C[C@@]1(C)[C@H](CCC(=C)[C@]21[H])OC(=O)C(C)=C
12317	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@@H](O)C[C@@]1(C)[C@H](CCC(=C)[C@]21[H])OC(=O)C(=C)CO
12318	[H][C@@]12CC[C@@]3(C)[C@H](O)CC(=O)C(C)=C3[C@@]1([H])OC(=O)[C@H]2C
12320	[H][C@@]12[C@@H]3OC(=O)[C@H](C)C3=CC[C@@]1(C)CCC(=O)[C@@H]2C
12321	[H][C@]12OC(=O)[C@@H](C)C1=CC[C@@]1(C)CC[C@H](O)[C@@](C)(O)[C@]21[H]
12322	[H][C@]12C[C@@]3(C)[C@H](O)CC=C(C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12323	[H][C@@]12CC[C@@]3(C)[C@@H](CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C)O[C@@H]1O[C@H](COC(C)=O)[C@@H](O)[C@H](O)[C@H]1O
12325	[H][C@]12C[C@@]3(C)[C@H](O)CCC(C)=C3C[C@]1([H])C(=C)C(=O)O2
12327	[H][C@]12C[C@@]3(C)[C@H](O)CCC(=C)[C@]3([H])[C@@H](OC(=O)C(\C)=C/C)[C@]1([H])C(=C)C(=O)O2
12328	[H][C@]12C[C@@]3(C)[C@H](O)CCC(=C)[C@]3([H])[C@@H](OC(C)=O)[C@]1([H])[C@@H](C)C(=O)O2
12329	[H][C@]12C[C@@]3(C)[C@H](O)CC[C@@](C)(O)[C@]3([H])[C@@H](OC(=O)C(\C)=C/C)[C@]1([H])C(=C)C(=O)O2
12330	[H]C1C(=O)C=C(C)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]12C
12331	[H][C@@]12C[C@@]3(C)[C@@H](O)C(=O)C=C(C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12333	[H][C@@]12C[C@@]3(C)[C@@H](OC(C)=O)[C@@H](O)C=C(C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12334	[H]C1[C@@H](O)C=C(C)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]12C
12336	[H][C@@]12C[C@@]3(C)[C@H](O)CCC(C)=C3[C@H](OC(=O)C(\C)=C\C)[C@]1([H])C(=C)C(=O)O2
12337	[H][C@@]12C[C@@]3(C)[C@@H](CCC(C)=C3[C@H](OC(=O)C(\C)=C\C)[C@]1([H])C(=C)C(=O)O2)OC(C)=O
12338	[H][C@@]12C[C@@]3(C)C=CC(=O)C(C)=C3C[C@]1([H])C(=C)C(=O)O2
12339	[H]C1CC(=C)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)C1
12340	[H]C1CC(=C)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)[C@@H]1O
12341	[H][C@@]12C[C@@]3(C)[C@@H](O)[C@@H](O)CC(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12342	[H]C1[C@@H](O)C(=C)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)[C@H]1O
12343	[H][C@@]12C[C@@]3(C)C[C@@H](O)CC(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12344	[H][C@@]12C[C@@]3(C)C[C@H](CC(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2)OC(C)=O
12345	[H][C@@]12C[C@@]3(C)C[C@@H](O)[C@H](O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12346	[H][C@@]12C[C@@]3(C)C[C@@H](O)[C@H](OC(C)=O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12347	[H][C@@]12C[C@@]3(C)C[C@@H](OC(C)=O)[C@H](O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12348	[H][C@@]12C[C@@]3(C)C[C@@H](OC(=O)C(\C)=C/C)[C@H](O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12349	[H][C@@]12C[C@@]3(C)C[C@@H](OC(=O)C=C(C)C)[C@H](O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12350	[H]C1C[C@]2(C)C[C@@]3([H])OC(=O)C(=C)[C@@]3([H])C[C@@]2([H])C(=C)[C@@H]1O
12352	[H][C@@]12C[C@@]3(C)C=CC(=O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12353	[H][C@@]12C[C@@]3(C)CCC(=O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12354	[H][C@@]12C[C@@]3(C)C=C[C@H](O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12355	[H][C@@]12C[C@@]3(C)C=C[C@H](OC(C)=O)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12357	[H]C1C[C@H](C)C2=C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)C1[H]
12358	[H]C1C[C@H](C)C2=C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]2(C)[C@@H]1O
12359	[H]C1[C@@H](O)C[C@H](C)C2=C[C@]3([H])C(=C)C(=O)O[C@]3([H])C[C@@]12C
12360	[H][C@]12C[C@@]3(C)CCCC(C)=C3C[C@]1([H])C(=C)C(=O)O2
12361	[H][C@@]12C[C@@]3(C)CC(=O)C[C@H](C)C3=C[C@]1([H])C(=C)C(=O)O2
12362	[H][C@@]12C[C@@]3(C)[C@H](O)CC[C@@]4(C)O[C@@]34C[C@]1([H])C(=C)C(=O)O2
12363	[H][C@@]12C[C@@]3(C)C=C[C@H]4O[C@@]4(C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12364	[H][C@@]12C[C@@]3(C)[C@@H](O)CC[C@@](C)(O)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12365	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@@H](OC(=O)C(C)C)[C@@]1([H])[C@](C)(O)CC[C@@H](OC(C)=O)[C@]1(C)[C@H]2OC(C)=O
12366	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])[C@](C)(O)CC[C@@H](OC(C)=O)[C@]1(C)[C@H]2OC(C)=O
12367	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@@H](OC(=O)C(C)=C)[C@@]1([H])[C@](C)(O)CC[C@@H](OC(C)=O)[C@]1(C)[C@H]2OC(C)=O
12368	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@H](OC(=O)C(C)C)[C@@]1([H])[C@@]3(C)CC[C@@H](O3)[C@]1(C)[C@@H]2OC(C)=O
12369	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@@]1([H])[C@@]3(C)CC[C@@H](O3)[C@]1(C)[C@@H]2OC(C)=O
12370	[H][C@]12OC(=O)C(=C)[C@@]1([H])[C@H](OC(=O)C(C)=C)[C@@]1([H])[C@@]3(C)CC[C@@H](O3)[C@]1(C)[C@@H]2OC(C)=O
12372	[H][C@]12C[C@@]3([H])[C@H](C)C(=O)CC[C@]3(C)C[C@@H]1OC(=O)C2=C
12373	[H][C@@]1(C)C(=O)O[C@]2([H])C[C@@]3(C)C=CC(=O)C(=C)[C@]3([H])C[C@]12[H]
12374	[H][C@@]12C[C@@]3(C)C=CC(=O)C(=C)[C@]3([H])C[C@]1([H])[C@@](C)(O)C(=O)O2
12375	[H][C@@]12C[C@@]3(C)CCC[C@H](C)C3=C[C@]1([H])[C@H](C)C(=O)O2
12377	[H][C@@]1(C)C(=O)O[C@]2([H])C[C@@]3(C)CCCC(=C)[C@]3([H])C[C@]12[H]
12378	[H][C@]1(C)C(=O)O[C@]2([H])C[C@@]3(C)[C@@H](O)C[C@@H](O)C(=C)[C@]3([H])C[C@]12[H]
12379	[H][C@@]12C[C@@]3(C)CC[C@H](O)C(=C)[C@]3([H])C[C@]1([H])[C@](C)(O)C(=O)O2

ID	SMILES
12381	[H][C@]12C[C@@]3(C)CCCC(=C)[C@]3([H])CC1=C(C)C(=O)O2
12382	[H][C@@]12CC3=C(C)C(=O)O[C@@]3(O)C[C@@]1(C)CCCC2=C
12384	[H][C@@]12C[C@@]3(C)CCC[C@H](C)C3=CC1=C(C)C(=O)O2
12385	[H][C@@]12C[C@@]3(C)CCC[C@H](C)C3=CC1=C(CO)C(=O)O2
12386	[H][C@@]12CC3=C(C)C(=O)OC3=C[C@@]1(C)CCCC2=C
12387	[H][C@@]12CC3=C(C[C@@]1(C)CCCC2=C)OC(=O)[C@@H]3C
12389	[H][C@@]12CC[C@]3(C)[C@@]4(C)C(=O)C=C(C)[C@@]34[C@@]1([H])OC(=O)[C@H]2C
12390	[H][C@@]12CC(CO)=C([C@@H](C)CCCO)[C@@H](OC(=O)C(C)=C)[C@]1([H])C(=C)C(=O)O2
12391	[H][C@@]12CC(CO)=C([C@@H](C)CCCO)[C@@H](OC(=O)C(\C)=C/C)[C@]1([H])C(=C)C(=O)O2
12392	[H]C1C([C@@H](C)CCC(=O)OC)=C(C)C[C@@]2([H])OC(=O)C(=C)[C@@]12[H]
12395	[H][C@]1(CC(C)=O)OC(=O)C(=C)[C@H]1CC(=O)[C@H](C)CCCO[C@]1([H])O[C@H](COC(C)=O)[C@@]([H])(O)[C@]([H])(O)[C@@]1([H])O
12396	[H][C@]1(CC2=C(C)C=CC=C2C)CCOC(=O)C1=C
12397	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)C(C)=O)[C@H](OC(C)=O)\C(=C/CC\C(C)=C\2)C(=O)OC
12398	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=C)[C@]1(C)O[C@@H]1C)[C@@H](O)\C(=C/CC\C(COC(C)=O)=C\2)C(=O)OC
12399	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(C)=O)[C@H](OC(=O)[C@@]1(C)O[C@H]1C)\C(=C/CC\C(C)=C\2)C(=O)OC
12400	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@H](OC(C)=O)\C(=C/CC\C(CO)=C\2)C(=O)COC
12401	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@H](C)\C(=C/CC\C(CO)=C\2)C(=O)OC
12402	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(OC(C)=O)[C@H](C)OC(C)=O)[C@H](OC(C)=O)\C(=C/CC\C(CO)=C\2)C(=O)OC
12403	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)C)[C@@H](OC(=O)C(C)(C)O)\C(C=O)=C/CC\C(C)=C\2
12405	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](O)[C@@H](NC(=O)C(C)C)\C(C)=C/CC\C(C)=C\2
12406	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(CC)OC(C)=O)[C@H](OC(C)=O)\C(C=O)=C/CC\C(C)=C\2
12407	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@@H](O)\C(C=O)=C/CC\C(C)=C\2
12408	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@H](OC(C)=O)\C(C=O)=C/CC\C(CO)=C\2
12409	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C\C(C=O)=C/CC\C(CO)=C\2)OC(=O)[C@@H](C)CC
12411	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@H](O)\C(C=O)=C/CC\C(CO)=C\2
12412	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@H](OC(C)=O)\C(CO)=C/CC\C(CO)=C\2
12413	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@H](C)CC)[C@H](OC(=O)CCCCCCCCCCCCCCCC)\C(C=O)=C/CC\C(CO)=C\2
12414	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@H](OC(=O)CCCCCCCCCCCCCCCCCC)\C(C=O)=C/CC\C(CO)=C\2
12415	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@H](C)CC)[C@H](OC(=O)CCCCCCC\C=C\C\C=C\CCCCC)\C(C=O)=C/CC\C(CO)=C\2
12416	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@H](C)CC)[C@H](OC(=O)CCCCCCC\C=C\C\C=C\C\C=C\CC)\C(C=O)=C/CC\C(CO)=C\2
12417	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@H](OC(C)=O)\C(C=O)=C/CC\C(CO)=C\2
12418	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@H](OC(C)=O)\C(=C/CC\C(CO)=C\2)C(=O)OC
12419	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@H](OC(C)=O)\C(=C/CC\C(CO)=C\2)C(=O)OC
12420	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@H](OC(C)=O)\C(C=O)=C/CC\C(CO)=C\2
12421	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](O)\C(=C/2)C(=O)OC
12422	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@](C)(O)[C@H](C)OC(C)=O)[C@@H](O)\C(=C/2)C(=O)OC
12424	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(=O)[C@H](C)CC)\C(=C/2)C(=O)OC
12425	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@](C)(O)[C@H](C)OC(C)=O)[C@@H](OC(C)=O)\C(=C/2)C(=O)OC
12426	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)[C@@]([H])(C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(=O)[C@H](C)CC)\C(=C/2)C(=O)OC
12427	[H][C@]12O[C@@]1([H])\C(C)=C\[C@@]1([H])OC(=O)[C@]([H])(C)[C@]1([H])[C@H](OC(=O)[C@@](C)(O)[C@H](C)OC(C)=O)[C@@H](OC(C)=O)\C(=C/2)C(=O)OC
12428	[H][C@]12O[C@]1(C)CC\C=C(/[C@H](O)[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]21[H])C(=O)OC
12431	[H][C@@]12[C@H](OC(C)=O)[C@H]3O[C@@]3(C)[C@]1(O)[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]21CO1)OC(=O)C(\C)=C/C
12432	[H][C@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@]3(O)CO[C@](O)([C@H]4O[C@@]14C)[C@]23[H])OC(=O)[C@@H](C)CC
12433	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@]3(O)CO[C@@H]([C@]13[H])[C@]1(O)O[C@@]21CO)OC(=O)CC(C)C
12435	[H]C1[C@H](OC(=O)C(=C)[C@H](C)O)[C@@H]2[C@H](OC(=O)C2=C)[C@@]2([H])C(=C)C(=O)C[C@@]2([H])C1=C
12436	[H]C1C([H])C(=C)[C@]2([H])CC(=O)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
12437	[H][C@@]12C[C@H](O)[C@@]3(CO3)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@]1(C)CO1
12438	[H][C@@]12C[C@H](O)[C@@]3(CO3)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@](C)(O)CCl
12439	[H][C@@]12C[C@H](O)[C@@]3(CO3)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)CC2=C
12440	[H][C@@]12C[C@H](O)[C@]3(CO3)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@H]1O[C@@H]1C
12441	[H][C@@]12C[C@H](O)[C@]3(CO3)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)C(C)=C
12442	[H][C@@]12C[C@H](O)[C@@]3(CO3)[C@]1([H])[C@@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)C(=C)CO
12446	[H][C@]12O[C@]1(C)CC\C=C(/[C@H](OC(C)=O)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H]2OC(=O)C1=C)C(=O)OC
12447	[H][C@]12O[C@]1(C)CC\C=C(/[C@H](OC(C)=O)[C@@H](OC(=O)C(\C)=C/C)[C@@H]1[C@@H]2OC(=O)C1=C)C(=O)OC
12448	[H][C@@](OC(=O)C(\C)=C/C)(C1=C(C)C(=O)O[C@H]1O)[C@]1(C)[C@@H](C)CCC=C1C=O
12449	[H][C@]12O[C@]1(C)CC\C=C(/[C@H](OC(C)=O)[C@@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H]1[C@@H]2OC(=O)C1=C)C(=O)OC
12451	[H][C@]12O[C@]1(C)CC\C=C(/[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(C)=O)[C@@H]1[C@@H]2OC(=O)C1=C)C(=O)OC
12453	[H][C@]12O[C@]1(C)\C=C/C=C(\[C@H](O)[C@@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H]1[C@@H]2OC(=O)C1=C)C(=O)OC
12454	[H][C@]12O[C@@]1([H])[C@@]1(C)O[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(C)=O)\C(=C/2)C(=O)OC
12455	[H][C@@](OC(=O)CC(C)C)(C1=C(C)C(=O)O[C@H]1O)[C@]1(C)[C@@H](C)CCC=C1C=O
12456	[H][C@@]12CC\C(C)=C\[C@H]3OC(=O)C(=C)[C@@H]3[C@H](OC(=O)[C@@H](C)CC)[C@@H](O)[C@@]1(O2)C(=O)OC
12457	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](CC(C)=C1CC=C(C)[C@H]21)OC(C)=O
12458	COC(=O)C1=C/CC[C@H](CO)C[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\1)OC(=O)C(C)C
12459	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)CC(C)=C1CC=C(C)[C@H]21
12460	CC[C@H](C)C(=O)O[C@@H]1C\C(=C/CC[C@H](CO)C[C@H]2OC(=O)C(=C)[C@H]12)C(=O)OC
12461	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(C)=O)[C@@H](OC(C)=O)C(C)=C1C=CC(=C)[C@]21[H]
12462	COC(=O)C1=C/CC[C@H](CO)C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(C)C)[C@H]\1OC(C)=O
12463	CC[C@H](C)C(=O)O[C@H]1[C@@H]2[C@@H](C[C@@H](CO)CC\C=C(/[C@@H]1OC(C)=O)C(=O)OC)OC(=O)C2=C
12464	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)C(C)C)[C@@H](OC(C)=O)C(C)=C1C=CC(=C)[C@]21[H]
12465	COC(=O)C1=C/CC[C@H](C)C[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\1)OC(=O)C(\C)=C/C
12466	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(C)=O)[C@@H](OC(C)=O)C(C)=C1CC[C@@](C)(O)[C@]21[H]
12467	COC(=O)C1=C/CC[C@H](C)C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C/C)[C@H]\1O
12468	COC(=O)C1=C/CC[C@H](C)C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)CC(C)C)[C@H]\1O
12471	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)[C@@H](C)CC)[C@@H](OC(C)=O)C(C)=C1CC[C@](C)(O)[C@]21[H]
12472	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](CC(C)=C1C(=O)C=C(C)[C@]21[H])OC(C)=O
12473	[H]C1CC(C)=C2C(=O)C=C(C)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12474	[H]C1CC(C)=C2C(=O)C=C(CO)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]

ID	SMILES
12475	[H][C@@]12[C@@H](CC(CO)=C3C(=O)C=C(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C\C
12476	[H][C@@]12[C@H](CC(CO)=C3C(=O)C=C(C)[C@]3(O)[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C\C
12477	[H][C@@]12CC[C@@](C)(O)[C@H]3CC(=O)C(C)=C3[C@@]1([H])OC(=O)C2=C
12478	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)CC(C)=C1C[C@H](O)[C@@](C)(O)[C@H]21
12479	[H]C1[C@H](C)C(C)=C2C[C@H]3O[C@@]3(C)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12480	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(C)=O)[C@H](C)C(C)=C1C[C@H]3O[C@@]3(C)[C@]21[H]
12481	[H][C@@]12[C@@H](O)[C@H](OC(=O)C(\C)=C/C)C(C)=C3C[C@H]4O[C@@]4(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12482	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)C(\C)=C/C)[C@H](OC(C)=O)C(C)=C1C[C@H]3O[C@@]3(C)[C@]21[H]
12483	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](CC(C)=C1[C@@H]2[C@]2(C)O[C@@H]2C1=O)OC(=O)C=C(C)C
12484	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@@](C)(O)[C@@]13O[C@@H]1[C@@H]1O[C@]1(C)[C@]23[H])OC(C)=O
12485	[H][C@@]12CC[C@]3(C)O[C@@]33C[C@@H]4O[C@]4(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12486	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)C[C@]1(C)O[C@@]11[C@@H]3O[C@@H]3[C@@](C)(O)[C@]21[H]
12488	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)C=C(C)[C@]1(O)CC=C(C)[C@]21[H]
12489	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)C(=C)[C@@H](C)OC(C)=O)C=C(C)[C@]1(O)CC=C(COC(C)=O)[C@]21[H]
12490	[H][C@]12OC(=O)C(=C)[C@]1([H])C(=O)[C@H]1O[C@@]1(C)[C@@]1(CC=C(CO)[C@]21[H])OC(=O)C(\C)=C\C
12491	C\C1=C/C[C@H]2[C@@H](C\C(C=O)=C/CC1)OC(=O)C2=C
12494	[H]C1C[C@@]2([H])C(C)=C[C@H](OC(=O)C=C(C)C)[C@@]3([H])C(=C)C(=O)O[C@]3([H])[C@@]2([H])C1=C
12496	[H]C1C=C(C)[C@]2([H])C[C@H](OC(=O)C(\C)=C/C)C(=C)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12497	[H]C1C=C(C)[C@]2([H])C[C@H](OC(=O)C=C(C)C)C(=C)[C@]2([H])[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12498	[H][C@@]12C[C@@H](OC(C)=O)[C@@](C)(OC(C)=O)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@@H](OC(=O)C(=C)[C@H](C)OC(C)=O)C=C2C
12499	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)C=C(C)[C@@]1(CC=C(C)[C@]21[H])OO
12501	[H][C@@]12CC=C(C)[C@]3([H])CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)C2=C
12502	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)CC(=C)[C@]1(O)CC=C(C)[C@]21[H]
12503	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](O)CC(=C)[C@@H]1CC=C(C)[C@]21[H]
12504	[H][C@]12CC=C(C)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@@H](CC2=C)OC(=O)C(\CO)=C\CO
12505	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)CC(=C)[C@@]1(CC=C(C)[C@]21[H])OO
12506	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](CC(=C)[C@@]1(CC=C(C)[C@]21[H])OO)OC(C)=O
12507	[H][C@@]12[C@H](CC(=C)[C@]3(O)C[C@@H]4O[C@]4(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(C)=O
12508	[H]C1[C@H]2O[C@@]2(C)[C@]2([H])[C@@]3([H])OC(=O)C(=C)[C@]3([H])C([H])CC(=C)[C@]12[H]
12509	[H][C@@]12[C@@H](CC(=C)[C@]3([H])[C@H](O)[C@H]4O[C@@]4(C)[C@]3(O)[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C/C
12510	[H][C@@]12[C@@H](CC(=C)[C@]3([H])[C@H](OC(C)=O)[C@H]4O[C@@]4(C)[C@]3(O)[C@@]1([H])OC(=O)C2=C)OC(=O)C(\C)=C/C
12511	[H][C@@]12C[C@H]3O[C@@]3(C)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](CC2=C)OC(=O)C(C)=C
12512	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@](C)(O)[C@]1([H])[C@H](O)[C@H]3O[C@@]3(C)[C@]21[H])OC(=O)[C@@H](C)CC
12513	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@](C)(O)[C@]1([H])[C@H](O)[C@H]3O[C@@]3(C)[C@]21[H])OC(=O)CC(C)C
12514	[H][C@@]12C[C@@H](OC(C)=O)[C@@](C)(O)[C@]34OO[C@](C)(C=C3)[C@]4([H])[C@@]1([H])OC(=O)C2=C
12515	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])C(=O)C=C(C)[C@]21[H])OC(=O)C(\C)=C\C
12517	[H][C@]12OC(=O)[C@@](C)(O)[C@]1([H])[C@@H](CC(C)=C1CC=C(C)[C@H]21)OC(=O)C(\CO)=C/CO
12518	[H]C1CC(C)=C2C(=O)C=C(C)[C@]2([H])[C@@]2([H])OC(=O)[C@@]([H])(C)[C@]12[H]
12519	[H]C1CC(C)=C2C(=O)C=C(C)[C@]2(O)[C@@]2([H])OC(=O)[C@@]([H])(C)[C@]12[H]
12520	[H][C@@]1(C)C(=O)O[C@]2([H])[C@@]3([H])C(C)=CC(=O)C3=C(C)C[C@H](O)[C@@]12[H]
12521	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@H](CC(C)=C1C(=O)C=C(C)[C@]21[H])OC(C)=O
12524	[H][C@@]12OC(=O)C(=C1)[C@@H](COCC)CC\C(CO)=C/[C@@]1([H])OC(=O)C(=C)[C@]21[H]
12526	[H][C@@]12OC(=O)C(=C1)[C@@H](COC(C)=O)CC\C(CO)=C/[C@@]1([H])OC(=O)C(=C)[C@]21[H]
12528	[H][C@@]12OC(=O)C(=C1)[C@@H](COC(=O)C(C)C)CC\C(CO)=C/[C@@]1([H])OC(=O)C(=C)[C@]21[H]
12529	[H][C@@]12OC(=O)C(=C1)[C@@H](COC(=O)[C@@H](C)CC)CC\C(CO)=C/[C@@]1([H])OC(=O)C(=C)[C@]21[H]
12530	[H][C@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@H](O)C=C(C)[C@]1([H])CCC(=C)[C@]21[H]
12531	[H][C@@]12C[C@H](CO)CC[C@H](COC(C)=O)C3=C[C@@]([H])(OC3=O)[C@@]1([H])C(=C)C(=O)O2
12532	[H][C@@]12CC=C(C)[C@]3([H])CCC(=C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12537	[H][C@]12CC=C(C)[C@@]1([H])[C@]1([H])OC(=O)[C@@](C)(OC(C)=O)[C@]1([H])[C@H](C[C@]2(C)O)OC(=O)C=C(C)C
12538	[H][C@]12CC=C(C)[C@@]1([H])[C@]1([H])OC(=O)[C@@](C)(OC(C)=O)[C@]1([H])[C@H](C[C@]2(C)OC(C)=O)OC(=O)C(\C)=C/C
12548	[H]C1[C@H]2O[C@@]2(C)[C@@]2([H])C1=C(C)CC([H])[C@@]1([H])[C@@H](C)C(=O)O[C@]21[H]
12549	[H][C@]12OC(=O)[C@H](C)[C@]1([H])[C@@H](CC(C)=C1[C@H](OC(C)=O)[C@H]3O[C@@]3(C)[C@]21[H])OC(C)=O
12552	[H][C@@]12CC[C@@]3(C)O[C@]33CC=C(C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12553	[H][C@]12OC(=O)[C@@H](C)[C@]1([H])[C@@H](C[C@]1(C)O[C@@]11CC=C(C)[C@]21[H])OC(C)=O
12554	[H][C@@]12CC[C@]3(C)O[C@@]33C(=O)C=C(C)[C@]3([H])[C@@]1([H])OC(=O)[C@@H]2C
12555	[H]C1C([H])\C(C)=C\C[C@H](O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12
12556	[H][C@]12OC(=O)[C@H](C)[C@]1([H])[C@@H](O)C[C@]1(C)O[C@@]11C(=O)C=C(C)[C@]21[H]
12558	[H][C@]12OC(=O)[C@H](C)[C@]1([H])[C@@H](CC1=C3[C@@H](C[C@@](C)(O)[C@]23[H])OC1=O)OC(=O)C(\C)=C\C
12559	[H][C@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])C(=O)[C@H]1O[C@@]1(C)[C@@]1(CC=C(CO)[C@]21[H])OC(=O)C(\C)=C\C
12560	[H][C@@]12CC(=O)[C@@H](C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@H](CC2=C)OC(C)=O
12561	[H][C@@]12CC=C(C)[C@]3([H])CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12562	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])[C@H](O)CC(=C)[C@@]1([H])CC=C(C)[C@]21[H]
12563	[H]C1[C@H](OC(=O)C(\C)=C/C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@H](O)C\C=C1/C
12564	[H]C1[C@H](OC(=C)C(\C)=C/C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@@H](O)C\C=C1/C
12565	[H]\C1=C(C)/C[C@@H](O)[C@@]2([H])C(=C)C(=O)O[C@]2([H])\C([H])=C(C)\[C@@H](C1)OC(C)=O
12567	[H]\C1=C(C)/C[C@@H](OC(=O)C(\COC(C)=O)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]2([H])\C([H])=C(C)/[C@@H](C1)OC(C)=O
12568	[H]C1[C@@H](OC(=O)C(C)C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@H](C\C=C1/C)OC(C)=O
12570	[H]\C1=C(C)/C[C@@H](OC(=O)C(\COC(=O)C(\C)=C\C)=C\CO)[C@@]2([H])C(=C)C(=O)O[C@]2([H])\C([H])=C(C)/[C@H](C1)OC(C)=O
12572	[H]C1[C@@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@H](C\C=C1/C)OC(C)=O
12573	[H]\C(C)=C(\C)C(=O)O[C@@H]1[C@H](O)\C(C)=C([H])\C[C@H](OC(C)=O)\C(C)=C([H])\[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12574	[H][C@@]12C[C@H]3O[C@@]3(C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@@H](O)CC2=C
12575	[H][C@@]12C[C@H]3O[C@@]3(C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@H](CC2=C)OC(C)=O
12576	[H][C@@]12C[C@H](O)[C@]3(CO3)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@H](CC2=C)OC(=O)C(C)=C
12577	[H][C@@]12CC=C(C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@H](C[C@@]21CO1)OC(C)=O
12578	[H][C@@]1(C)C(=O)O[C@@]2([H])[C@@]1([H])CCC(=C)[C@]1([H])CCC(=C)[C@]21[H]
12579	[H][C@@]12C[C@H](O)C(=C)[C@]1([H])[C@@]1([H])OC(=O)[C@](O)(CO)[C@]1([H])CCC2=C
12580	[H][C@@]12C[C@H](O)C(=C)[C@]1([H])[C@@]1([H])OC(=O)[C@](O)(COC(C)=O)[C@]1([H])CCC2=C

ID	SMILES
12581	[H]C1C[C@@]2([H])C(=C)CC[C@@]3(O)[C@@]([H])(C)C(=O)O[C@]3([H])[C@@]2([H])C1=C
12582	[H][C@@]12CC(=O)C(=C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@@H](O)CC2=C
12583	[H][C@@]12C[C@H](O)C(=C)[C@]3([H])CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12584	[H][C@@]12CCC(=C)[C@]3([H])CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12585	[H][C@@]12C[C@H](O)[C@@H](C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@@H](O)CC2=C
12588	[H][C@@]12CC(=O)[C@@H](C)[C@]1([H])[C@@]1([H])OC(=O)[C@@H](C)[C@]1([H])[C@@H](O)CC2=C
12590	[H][C@]12OC(=O)[C@@H](C)[C@]1([H])[C@H](C[C@@](C)(O)C1=CCC(C)=C21)OC(C)=O
12591	[H][C@@]12CC[C@@](C)(O)[C@H]3CC(=O)C(C)=C3[C@@]1([H])OC(=O)[C@H]2C
12592	[H][C@@]12CC[C@](C)(O)C3=CCC(C)=C3[C@@]1([H])OC(=O)[C@H]2C
12593	[H][C@@]12CC[C@@H](C)C3=CC(=O)[C@@H](C)[C@]3([H])[C@@]1([H])OC(=O)[C@H]2C
12594	[H][C@]12CCC(C)=C1[C@@]1([H])OC(=O)C(=C)[C@]1([H])CC[C@H]2C
12595	C[C@@H](O)[C@@H]1C[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@]11O[C@@H]1[C@@H](Cl)[C@]2(C)O
12596	[H][C@]12C[C@@H](C)[C@]3(CCC(C)=C3[C@@]3([H])OC(=O)C(=C)[C@]13[H])O2
12597	[H][C@@]12CC(=O)[C@H](C)[C@]1(O)[C@]1([H])OC(=O)C(=C)[C@]1([H])[C@H]1C[C@@]21C
12598	[H][C@@]12CC(=O)C(C)=C1[C@]1([H])OC(=O)C(=C)[C@]1([H])[C@H]1C[C@@]21C
12599	[H][C@@]12[C@H]3C[C@@]3(C)[C@]3([H])C[C@H](OC(C)=O)[C@H](C)[C@]3(O)[C@]1([H])OC(=O)C2=C
12600	[H]\C1=C(C)/[C@H](OC(=O)[C@]2(C)O[C@@H]2C)[C@H](O)[C@H]2[C@H](OC(=O)C2=C)\C([H])=C(C)/[C@H](C1)OC(C)=O
12603	[H]C1C\C=C(C)\C([H])[C@@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C1\C
12604	[H]C1[C@@H](OC(=O)C(\CO)=C\CO)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@H](C\C=C1/C)OC(=O)[C@@H](O)C(C)C
12605	[H]\C(CO)=C(\CO)C(=O)O[C@@H]1C\C(C)=C([H])\C[C@H](OC(=O)CC(C)C)\C(C)=C([H])\[C@]2([H])OC(=O)C(=C)[C@@]12[H]
12606	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](O)CC(C)=C1C(=O)C=C(CO)[C@]21[H]
12607	[H]\C1=C(C)/C[C@@H](OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C)=C\CO)[C@@H]2[C@H](OC(=O)C2=C)\C([H])=C(C)\[C@H](C1)OC(C)=O
12608	CC[C@H](C)C(=O)O[C@H]1[C@@H](O)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@H](C\C=C1/C)OC(C)=O
12609	C\C=C(\C)C(=C)O[C@H]1C\C(C)=C\CC(=O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12
12610	[H][C@@]12[C@@H](CC(C)=C3C(=O)C=C(C)[C@]3([H])[C@@]1([H])OC(=O)C2=C)OC(=O)CC1=CC=C(O)C=C1
12611	CC(=C)C(=O)O[C@H]1C\C(C)=C\CC\C(CO)=C/[C@@H]2OC(=O)C(=C)[C@H]12
12612	[H][C@@]12[C@@H](OC(=O)C1=C)[C@@]1([H])C(C)=C[C@@H](O)[C@@]1([H])C(=C)C[C@H]2OC(=O)C(\C)=C/C
12613	C\C=C(/C)C(=O)O[C@@H]1C\C(C)=C\CC\C(CO)=C/[C@@H]2OC(=O)C(=C)[C@H]12
12614	[H][C@@]12[C@@H](OC(=O)C1=C)[C@@]1(O)C(C)=C[C@@H](OC(C)=O)[C@@]1([H])C(=C)C[C@H]2OC(=O)C(\C)=C/C
12615	CC(=O)OC\C1=C\[C@@H]2OC(=O)C(=C)C2[C@@H](C\C(C)=C\CC1)OC(=O)C(\C)=C\CO
12616	CC(=O)O[C@@H]1C\C(COC(=O)C(C)=C)=C\CC\C(CO)=C\[C@H]2OC(=O)C(=C)[C@H]12
12617	[H][C@@]12[C@@H](OC(=O)C1=C)[C@@]1(O)C(C)=C[C@@H](O)[C@@]1([H])C(=C)C[C@H]2OC(=O)C(\C)=C\C
12618	C\C=C(/C)C(=O)OC\C1=C/CC\C(CO)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C1)OC(C)=O
12620	[H][C@@]12[C@@H](OC(=O)C1=C)[C@]1([H])C(=C)C[C@@H](O)[C@@]1([H])C(=C)C[C@H]2O
12622	[H][C@@]12C[C@H](O)\C(CO)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12623	[H][C@]12CCC(=C)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@H](CC2=C)OC(=O)C(\C)=C\C
12624	[H][C@@]12C[C@H](OC(C)=O)\C(CO)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12625	[H][C@@]12C[C@H](O)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12626	[H][C@@]12C[C@H](O)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(\C)=C\C
12629	[H][C@@]12CC=C(C)[C@@]1([H])[C@@H]1OC(=O)C(=C)[C@]1([H])[C@@H](O)C[C@@]2(C)O
12633	[H][C@@]12CCC(C)=C3[C@@H]([C@H]1OC(=O)[C@@H]2C)C(CO)=CC3=O
12634	[H][C@@]12C[C@H](OC(C)=O)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@H](C[C@@]1(C)O2)OC(=O)CC(C)C
12635	[H][C@@]12C[C@H](O)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)C
12636	[H][C@@]12C[C@H](OC(C)=O)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12637	[H][C@@]12C[C@H](OC(=O)C(C)=C)\C(C)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12638	[H][C@@]12C[C@H](O)\C(C=O)=C/[C@@]3([H])OC(=O)C(=C)[C@]3([H])[C@@H](C[C@@]1(C)O2)OC(=O)C(C)=C
12639	[H][C@@]12O[C@]1([H])[C@@]1([H])O[C@]1(C)C[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]1([H])\C=C2\CO
12640	[H][C@@]12O[C@]1([H])[C@@]1([H])O[C@]1(C)C[C@@H](OC(=O)C(\C)=C/COC(C)=O)[C@@]1([H])C(=C)C(=O)O[C@]1([H])\C=C2\CO
12641	[H][C@@]12O[C@]1([H])[C@@]1([H])O[C@]1(C)C[C@@H](OC(=O)C(\COC(C)=O)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]1([H])\C=C2\CO
12642	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C[C@@H]1O)\C(C)=C/2)OC(=O)C(C)C
12643	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@@H](C[C@@H]1O)\C(C)=C/2)OC(=O)C(\C)=C/C
12644	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C[C@@H]1O)\C(C)=C/2)OC(=O)C(\C)=C/C
12645	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C[C@@H]1O)\C(C)=C/2)OC(=O)C(C)=C
12646	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C[C@@H]1OC(C)=O)\C(C)=C/2)OC(=O)C(C)=C
12647	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)C[C@H](O)[C@@](O)(O1)\C(C)=C/2)OC(=O)C(C)C
12648	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)C[C@H](OC(C)=O)[C@@](O)(O1)\C(C)=C/2)OC(=O)C(C)C
12650	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)CC[C@@](O)(O1)\C(C)=C/2)OC(=O)C(\C)=C/C
12651	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)C[C@H](OC(C)=O)[C@@](O)(O1)\C(C)=C/2)OC(=O)C(\C)=C\C
12652	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)C[C@H](OC(C)=O)[C@@](O)(O1)\C(C)=C/2)OC(=O)[C@@H](C)CC
12653	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(CC1=O)\C(C)=C/2)OC(=O)C(C)=C
12654	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(C)C
12655	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(C)=C
12656	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)CC(C)C
12657	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(\C)=C\C
12658	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(CO)=C/2)OC(=O)C(\C)=C/C
12659	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@H](O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12660	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@H](OC(=O)CC(C)C)[C@]1(C)OC(=CC1=O)\C(CO)=C/2
12661	[H][C@@]12CCC(=C)[C@]1([H])[C@@H](OC(C)=O)[C@H]1[C@H](CC2=C)OC(=O)C1=C
12662	[H][C@@]12CCC(=C)[C@]1([H])[C@@H](OC(=O)C(C)C)[C@H]1[C@H](CC2=C)OC(=O)C1=C
12663	[H][C@@]12CC[C@@](C)(O)[C@]1([H])C[C@H]1[C@H](CC2=C)OC(=O)C1=C
12664	[H][C@@]12CC[C@@]3(C)O[C@@]13C[C@H]1[C@H](C[C@@H]2C)OC(=O)C1=C
12665	[H][C@@]12C[C@H](OC(C)=O)C(=C)[C@]1(O)C[C@H]1[C@H](C[C@@]22CO2)OC(=O)C1=C
12666	[H][C@@]12C[C@H]3[C@H](C[C@H](C)C1=CC(=O)[C@@H]2C)OC(=O)C3=C
12667	[H]C1[C@H]2[C@H](C[C@H](C)[C@]3([H])CC(=O)C(C)=C13)OC(=O)C2=C
12668	[H][C@@]12CC(=O)C(C)=C1[C@H](O)[C@H]1[C@H](C[C@@H]2C)OC(=O)C1=C
12669	C[C@@H]1[C@H]2CC3=C(C)CCC3=C(C)C[C@@H]2OC1=O
12670	[H][C@]12C[C@@H](OC(=O)C(\C)=C\C)[C@@]3(C)O[C@@]13[C@@H](O)[C@@H]1[C@H](CO)C(=O)O[C@H]1[C@@H]1O[C@]21C

ID	SMILES
12671	[H][C@@]12[C@@H](C[C@](C)(O)C1=C[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C)OC(C)=O
12672	[H]C1C[C@@]2([H])C(=C)C[C@@]3([H])OC(=O)C(=C)[C@@]3([H])C[C@@]2([H])C1=C
12673	[H][C@@]12CC(=C)[C@]3([H])C[C@H](OC(=O)CC(C)C)C(=C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12674	[H][C@@]12CCC(=C)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])CCC2=C
12675	[H]C1C[C@@]2([H])C(=C)C[C@H](OC(=O)C=C(C)C)[C@]3([H])[C@H](OC(=O)C3=C)[C@@]2([H])C1=C
12676	[H]C1CC(=C)[C@]2([H])C[C@@H](O)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12678	[H]C1CC(=C)[C@]2([H])C[C@H](OC(C)=O)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12679	[H]C1CC(=C)[C@]2([H])C[C@H](OC(=O)C=C(C)C)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12680	[H][C@@]12C[C@H](O[C@]3([H])O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)C(=C)[C@]1([H])[C@@]1([H])OC(=O)C(=C)[C@]1([H])[C@@H](C)CC2=C
12681	[H]C1C[C@@]2([H])C(=C)CC([H])[C@]3(O)[C@H](OC(=O)C3=C)[C@@]2([H])C1=C
12682	[H]C1CC(=C)[C@]2([H])C[C@H](OC(=O)C(\C)=C/C)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12683	[H][C@@]12[C@H](OC(=O)C1=C)[C@@]1([H])C(=C)[C@@H](O)C[C@@]1([H])C(=C)C[C@@H]2O
12687	[H][C@@]12C[C@@H](O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@]1([H])[C@@H](CC2=C)OC(=O)C(C)=C
12688	[H][C@@]12[C@H](OC(=O)C1=C)[C@@]1([H])C(=C)[C@H](O)C[C@@]1([H])C(=C)C[C@H]2OC(=O)[C@](C)(O)CCl
12689	[H]C1CC(=C)[C@]2([H])C[C@@H](OC(=O)CC(C)C)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12692	[H]C1CC(=C)[C@]2([H])C[C@H](OC(=O)[C@@H](C)CC)C(=C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@]12[H]
12694	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(=C)[C@]1([H])[C@@H](O)[C@@H](O)C2=C)OC(=O)C(C)=C
12696	[H][C@@]12CCC(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1C[C@@H](OC(=O)C=C(C)C)C2=C
12697	[H][C@@]12CCC(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1C[C@@H](OC(=O)C(\C)=C/C)C2=C
12698	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)CC2=C
12699	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@](C)(CO)OCC
12700	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@](C)(O)CO
12701	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@](C)(O)CCl
12702	[H][C@@]1(CCl)[C@@H](O)C[C@@]2([H])C(=C)C[C@H](OC(=O)C(=C)CO)[C@@H]3[C@H](OC(=O)C3=C)[C@@]12[H]
12703	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)C(=C)CO
12704	[H][C@@]12C[C@H](O)[C@](O)(CCl)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)C(C)=C
12705	[H][C@@]12C[C@H](OC(C)=O)[C@](O)(CO)[C@]1([H])[C@@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)[C@](C)(O)CCl
12707	[H][C@@]12C[C@H](O)[C@@](C)(O)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC2=C)OC(=O)C(=C)CO
12709	[H]C1C[C@@]2([H])C(C)=C[C@H](OC(=O)CC(C)C)[C@@]3([H])C(=C)C(=O)O[C@]3([H])[C@@]2([H])C1=C
12710	[H]C1C[C@@]2([H])C(=C)C[C@@H](OC(=O)CC(C)C)[C@]3([H])[C@H](OC(=O)C3=C)[C@@]2([H])C1=C
12712	[H]C1C[C@@]2([H])C(C)=CC([H])[C@@]3([H])C(=C)C(=O)O[C@]3([H])[C@@]2([H])C1=C
12715	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](OC(=O)C(\C)=C\C)[C@@H](O)[C@]3(CO3)[C@]1([H])[C@H](O)C=C2C
12716	[H]\C1=C(C)/[C@H](C[C@@]2([H])O[C@]2(C)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H])OC(C)=O
12718	[H][C@]12[C@@H](C)C(=O)O[C@@]1([H])C1=C(C)CCC(=O)[C@]1(C)C[C@@H]2O
12719	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](OC(C)=O)C=C(C)[C@@]1(CC=C(C)[C@]21[H])OO
12720	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@H](OC(=O)C=C(C)C)[C@]1(C)OC(=CC1=O)\C(CO)=C/2
12721	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@H](OC(=O)C(\C)=C/C)[C@]1(C)OC(=CC1=O)\C(CO)=C/2
12722	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(CO)=C/2)OC(=O)C(C)=C
12724	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(CO)=C/2)OC(=O)CC(C)C
12726	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@@H](O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12729	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\CO)=C\C)[C@H](O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12730	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(CO)=C/2)OC(=O)C(C)C
12731	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)C(=C)[C@H]2O)OC(=O)CC(C)C
12732	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C=C1)\C(C)=C/2)OC(=O)C(\COC(C)=O)=C/C
12733	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(C)O[C@](O)(C=C1)\C(C)=C/2)OC(=O)C(C)C
12734	[H][C@@]12C[C@]([H])(C)[C@@]3(O)C[C@H](O)[C@](C)(C[C@@H](OC(=O)C(C)C)[C@@]1([H])C(=C)C(=O)O2)O3
12735	[H][C@@]12C[C@]([H])(C)[C@]3(CC[C@](C)(C[C@@H](OC(=O)C(C)C)[C@@]1([H])C(=C)C(=O)O2)O3)OCC
12736	[H][C@@]12C[C@]([H])(C)[C@@]3(O)CC[C@](C)(C[C@@H](OC(=O)C(C)C)[C@@]1([H])C(=C)C(=O)O2)O3
12737	[H][C@@]12C[C@]([H])(C)C3=CC[C@](C)(C[C@@H](OC(=O)C(C)=C)[C@@]1([H])C(=C)C(=O)O2)O3
12738	[H][C@@]12CC(=C)[C@]3([H])CC=C(C)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12739	[H][C@@]12C[C@]3([H])C(=C)C(=O)O[C@]3([H])CC(C)=C1C(=O)C=C2C
12740	[H][C@@]12CC(C)=C3CC[C@@](C)(O)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12741	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C\C(C)=C\CC[C@@]1(C)O[C@@H]21)OC(=O)C(\CO)=C/C
12742	[H][C@@]12CC(C)=C3CC[C@@](C)(OC(C)=O)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12743	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)O[C@@H]1\C=C/C(/CO)=C\2)OC(=O)C(\CO)=C/C
12744	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)O[C@@H]1\C=C/C(/C)=C\2)OC(=O)C(\C)=C/C
12745	[H][C@@]1(C[C@@](C)(O)[C@]2([H])C[C@]3([H])C(=C)C(=O)O[C@]3([H])CC(C)=C12)OC(C)=O
12746	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)O[C@@H]1\C=C/C(/C)=C\2)OC(=O)C(\CO)=C/C
12747	[H][C@]12C[C@]3([H])[C@]([H])([C@H](C[C@@]3(C)O)OC(C)=O)C(C)=C[C@@]1([H])OC(=O)C2=C
12748	[H][C@@]12CC(=C)[C@@]3([H])[C@@H](O)C[C@@](C)(O)[C@]3([H])C[C@]1([H])C(=C)C(=O)O2
12750	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@@]1(C)O[C@@H]1C[C@H](OC(C)=O)\C(C)=C/2)OC(=O)C(C)C
12751	[H]C1CC2=C(C)C(=C)[C@@]3([H])OC(=O)[C@@H](C)[C@@]3(C)[C@@]2([H])[C@]1(C)O
12752	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@@]3([H])OC(=O)[C@@](C)([C@]13[H])[C@](O)(O2)C(C)C
12753	[H][C@]12C[C@]3([H])C(C[C@@H](O)[C@@]3(C)O)=C(C)C(=O)[C@@]1([H])OC(=O)[C@H]2C
12754	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@@]3([H])OC(=O)[C@@](C)([C@]13[H])[C@](O)(O2)[C@@H](C)CC
12755	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@@]3([H])OC(=O)[C@@](C)([C@]13[H])[C@](O)(O2)C(C)=C
12756	[H][C@]12C[C@]3([H])[C@@]([H])(C[C@@H](O)[C@@]3(C)O)C(C)=C[C@@]1([H])OC(=O)[C@H]2C
12757	[H][C@]12C[C@]3([H])[C@@]([H])(C[C@@H](OC(=O)C(\C)=C/C)[C@@]3(C)O)C(C)=C[C@@]1([H])OC(=O)[C@H]2C
12758	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@@]3([H])OC(=O)[C@@](C)([C@]13[H])[C@](O)(O2)C(\C)=C\C
12759	[H][C@]12OC(=O)C(=C)[C@@]1([H])C[C@]1([H])[C@]([H])([C@@H](O)C[C@@]1(C)O)C(C)=C2
12761	[H][C@@]12C[C@H](C)[C@@]3([H])CC(=O)C(C)=C3[C@@H](O)[C@]1([H])[C@@H](C)C(=O)O2
12762	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@@]3([H])OC(=O)[C@@](C)([C@]13[H])[C@](O)(O2)[C@]1(C)O[C@H]1C
12765	[H][C@]12C[C@H](C)C3=CC(=O)[C@](C)(C[C@]4([H])O[C@](O)(C(C)=C)[C@@](C)(C(=O)O1)[C@@]24[H])O3
12766	[H][C@@]12C[C@@H](C)[C@]3([H])[C@H](C[C@@]4(C)O[C@@]34C[C@]1([H])C(=C)C(=O)O2)OC(=O)[C@@H](C)CC
12767	[H][C@@]12CC[C@@](C)(O)[C@]1([H])C[C@]1([H])C(=C)C(=O)O[C@]1([H])C=C2C
12768	[H][C@]12OC(=O)[C@@](O)(CCl)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(\C)=C/C
12771	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(CC1=CC[C@H]3C[C@@H]1C3(C)C)=C/2)OC(=O)C(\C)=C/C

ID	SMILES
12772	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(\C)=C/C
12774	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@@H](C[C@]1(C)OC(=CC1=O)\C(C)=C/2)OC(=O)C(C)=C
12775	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@@H](O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12777	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@H](O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12778	[H]C(=O)C1=C/C[C@]2([H])C(=C)C(=O)O[C@@]2([H])C\C(C)=C\CC\1
12779	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@H](OC(=O)C(C)=C)[C@]1([H])C(=C)C(=O)O2
12780	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(CO)=C/[C@H](OC(=O)C(C)=C)[C@]1([H])C(=C)C(=O)O2
12781	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@H](OC(=O)C(\C)=C\C)[C@]1([H])C(=C)C(=O)O2
12784	[H][C@]12C[C@]3(C)OC(=CC3=O)\C(C)=C/[C@H](OC(=O)[C@@]3(C)O[C@H]3C)[C@]1([H])C(=C)C(=O)O2
12785	[H][C@]12C[C@]3(C)OC(=CC3=O)C(=C)[C@H](O)[C@H](OC(=O)C(C)=C)[C@]1([H])C(=C)C(=O)O2
12786	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@](C)(CC(O)=O)OC(=O)\C(C)=C/2)OC(=O)C(\C)=C\C
12787	[H]C1C\C(C)=C/[C@@H](O)C\C(C)=C/[C@@]2([H])OC(=O)C(=C)[C@]12[H]
12791	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C\C(C=O)=C/CC\C(CO)=C\2)OC(=O)CC(C)C
12793	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C\C(C=O)=C/[C@@H](O)C\C(COC(C)=O)=C/2)OC(=O)CC(C)C
12794	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@@]([H])(OC(C)=O)\C(=C/CC\C(CO)=C/2)C(=O)OC
12795	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)C)[C@@]([H])(OC(C)=O)\C(=C/CC\C(CO)=C/2)C(=O)OC
12796	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@H](C)CC)[C@@]([H])(OC(C)=O)\C(=C/CC\C(CO)=C/2)C(=O)OC
12797	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(C)=O)[C@@]([H])(OC(=O)[C@@H](C)CC)\C(=C/CC\C(CO)=C/2)C(=O)OC
12798	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)C)[C@@]([H])(OC(C)=O)\C(=C/[C@@H](O)C\C(CO)=C/2)C(=O)OC
12799	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\C)=C/C)[C@@]([H])(OC(C)=O)\C(=C/[C@@H](O)C\C(CO)=C/2)C(=O)OC
12801	[H][C@@]1(OC(C)=O)\C=C(CO)/CC\C=C(CO)/C[C@@H]2OC(=O)C(=C)[C@@H]12
12803	[H]\C1=C(C)/CC[C@@H](O)C(=C)CC[C@@H]2[C@@H]1OC(=O)C2=C
12805	[H][C@]12OC(=O)C(=C)[C@]1([H])CCC(=C)[C@H](O)C[C@H](O)\C(C)=C\2
12806	C\C1=C/[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](O)CC(=C)[C@H](O)CC1
12807	[H]\C1=C(C)/CCC(=O)C(=C)CC[C@@H]2[C@@H]1OC(=O)C2=C
12808	[H][C@]12O[C@]1(C)CC[C@@H](OO)C(=C)CC[C@@H]1[C@@H]2OC(=O)C1=C
12810	C[C@H]1[C@@H]2CCC(=C)[C@H](O)CC\C(C)=C\[C@H]2OC1=O
12811	C\C1=C/[C@@H](O)[C@H]2[C@H](CC(=C)C(=O)CC1)OC(=O)C2=C
12812	CC(=O)O[C@@H]1\C=C(C)\CCC(=O)C(=C)C[C@@H]2OC(=O)C(=C)[C@@H]12
12814	[H][C@]12O[C@]1(C)C[C@H](OC(C)=O)[C@@H]1[C@H](OC(=O)C1=C)[C@H](O)C(=C)[C@H]1O[C@@H]21
12815	[H][C@]12O[C@]1(C)C[C@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O[C@]1([H])[C@H](O)[C@@]1(CO1)[C@H]1O[C@@H]21
12816	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(=O)C(C)C)\C(=C/CCC(=C)[C@@H]2O)C(=O)OC
12817	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)[C@@]1(C)O[C@H]1C)[C@@H](OC(=O)[C@@H](C)CC)\C(=C/CCC(=C)[C@@H]2O)C(=O)OC
12818	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\CO)=C/C)[C@@H](OC(=O)C(C)C)\C(=C/CCC(=C)[C@@H]2O)C(=O)OC
12819	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(\CO)=C/C)[C@@H](OC(=O)[C@@H](C)CC)\C(=C/CCC(=C)[C@@H]2O)C(=O)OC
12820	[H]\C1=C(C)/[C@H](O)CC\C(C)=C\[C@H](O)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12821	[H][C@@]12C\C(C)=C\CCC(=C)[C@]([H])(O)C[C@]1([H])C(=C)C(=O)O2
12822	[H][C@]12C\C(C)=C\CC(=O)[C@@H](C)C[C@@H](O)[C@]1([H])[C@@H](C)C(=O)O2
12823	[H][C@@]12C[C@H](C)C(=O)C\C=C(C)\C[C@H](OC(C)=O)[C@@]1([H])[C@@H](C)C(=C)O2
12825	C\C1=C/[C@@H](O)[C@H]2[C@@H](CC(=C)[C@@H](O)CC1)OC(=O)C2=C
12826	[H]C1C[C@H](O)\C(C)=C/[C@@H](OC(=O)[C@@H](C)CC)[C@@]2([H])C(=C)C(=O)O[C@]2([H])\C=C1\C
12827	[H]C1C[C@H](O)\C(C)=C/[C@@H](OC(=O)[C@@H](C)CC)[C@@]2([H])C(=C)C(=O)O[C@]2([H])\C=C1\CO
12830	[H][C@@]1(O)C[C@H](O)\C(C)=C/[C@H]2OC(=O)C(=C)[C@@H]2C[C@H](OC(=O)C(\C)=C/C)[C@@]1(C)O
12831	C\C=C(\C)C(=O)O[C@H]1C[C@@H]2[C@H](OC(=O)C2=C)\C=C(C)/[C@@H](O)CC[C@]1(C)O
12832	[H][C@]12CC(=C)[C@@H]3CC[C@](C)(O)[C@H](O3)[C@@H](O)[C@]1([H])C(=C)C(=O)O2
12833	[H][C@@]12C[C@H](C)C3=CC(=O)[C@](C)(C[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O2)O3
12835	[H][C@@]12C[C@H](C)C3=CC(=O)[C@](C)(C[C@H](OC(=O)C(C)=C)[C@@]1([H])C(=C)C(=O)O2)O3
12837	C\C=C(\C)C(=O)O[C@@H]1[C@@H](OC(C)=O)[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@@]1(C)O)OC(=O)C2=C
12838	C\C=C(\C)C(=O)O[C@H]1[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@@]1(C)O)[C@@H]1OC(C)=O)OC(=O)C2=C
12839	CC(C)CC(=O)O[C@@H]1C[C@](C)(O)C(=O)\C=C/[C@H](C)C[C@H]2OC(=O)C(=C)[C@H]12
12840	CC(C)CC(=O)O[C@H]1[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@](C)(O)[C@@H]1OC(C)=O)OC(=O)C2=C
12841	C[C@@H]1C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(C)=C)[C@@H](OC(C)=O)[C@@](C)(O)C(=O)\C=C/1
12844	C[C@@H]1C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(C)=C)[C@@H](OC(=O)C(C)=C)[C@@](C)(O)C(=O)\C=C/1
12845	C\C=C(\C)C(=O)O[C@@H]1[C@@H](OC(=O)C(C)=C)[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@]1(C)O)OC(=O)C2=C
12846	CC(C)C(=O)O[C@@H]1[C@@H](OC(=O)C(C)=C)[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@]1(C)O)OC(=O)C2=C
12847	[H][C@@]12O[C@]1([H])C(=O)[C@](C)(O)[C@H](OC(=O)C(\C)=C/C)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
12848	[H][C@@]12O[C@]1([H])C(=O)[C@](C)(O)[C@H](OC(=O)CC(C)C)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
12849	[H][C@@]12O[C@]1([H])C(=O)[C@](C)(O)[C@H](OC(=O)C(C)=C)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
12850	[H][C@@]12O[C@]1([H])C(=O)[C@](C)(O)[C@H](OC(=O)C(C)C)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
12851	[H][C@@]12O[C@]1([H])C(=O)[C@](C)(O)[C@H](OC(C)=O)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
12852	C[C@@H]1C[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(C)=O)[C@@H](OC(=O)C(C)=C)[C@@](C)(O)C(=O)\C=C/1
12853	CC[C@H](C)C(=O)O[C@@H]1C[C@@]2(C)O[C@](O)(C[C@@H]2O)[C@@H](C)C[C@H]2OC(=O)C(=C)[C@H]12
12854	[H][C@@]12C[C@H](C)[C@@](O)(O1)[C@@H](OC(=O)C(\C)=C/C)[C@H]1[C@@H](OC(=O)C1=C)[C@H](OC(=O)[C@@H](C)CC)[C@](C)(O)C2
12855	CC[C@@H](C)C(=O)O[C@H]1[C@@H]2OC(=O)C(=C)[C@H]2CC(=O)[C@@H](C)CCC[C@@]1(C)O
12856	C\C=C(\C)C(=O)O[C@H]1[C@@H]2OC(=O)C(=C)[C@H]2CC(=O)[C@@H](C)CCC[C@@]1(C)O
12857	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O
12858	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)[C@@]3(CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O)OC(C)=O
12859	CCC(=O)O[C@@H]1C[C@](C)(O)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(COC(C)=O)=C12
12860	CCC(=O)O[C@@H]1C[C@](C)(O)[C@@]2(CC[C@@](C)(O2)\C=C2\OC(=O)C(COC(C)=O)=C12)OC(C)=O
12861	CCC(=O)O[C@@H]1C[C@](C)(O)[C@@]2(CC[C@@](C)(O2)\C=C2\OC(=O)C(COC(C)=O)=C12)OC
12862	CO[C@@]12CC[C@@](C)(O1)\C=C1\OC(=O)C(COC(C)=O)=C1[C@@H](C[C@]2(C)OC(C)=O)OC(C)=O
12863	CC(=O)OCC1=C2[C@@H](C[C@](C)(OC(C)=O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O
12865	CC(=C)C(=O)O[C@@H]1C[C@](C)(O)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(C)=C12
12866	CC1=C2[C@@H](C[C@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)[C@]1(C)CO1
12867	CC(=O)OCC1=C2[C@@H](C[C@@](O)(CO)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
12868	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
12869	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)[C@]1(C)CO1

ID	SMILES
12870	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(=C)CO
12871	CC(=O)OCC1=C2[C@@H](C[C@](C)(O)C3=CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
12874	CO[C@@]12CC[C@@](C)(O1)\C=C1\OC(=O)C(COC(C)=O)=C1[C@@H](C[C@@]2(C)O)OC(=O)C(\C)=C\C
12877	C\C=C(/C)C(=O)O[C@@H]1C[C@@](C)(O)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(COC(C)=O)=C12
12878	CC(=C)C(=O)O[C@@H]1C[C@@](C)(O)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(CO)=C12
12879	CC(=O)OCC1=C2[C@@H](O)C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O
12880	CCO[C@@]12CC[C@@](C)(O1)\C=C1\OC(=O)C(COC(C)=O)=C1[C@@H](C[C@@]2(C)O)OC(=O)C(C)=C
12881	CCOCC1=C2[C@@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
12882	CC(=O)O[C@@H]1C[C@](C)(O)\C=C2\OC(=O)C(CO)=C2[C@@H](C\C(C)=C/1)OC(=O)C(C)=C
12883	CC(=O)OCC1=C2[C@@H](C\C(C)=C/[C@@H](C[C@](C)(O)\C=C2\OC1=O)OC(C)=O)OC(=O)C(C)=C
12884	CC(=O)O[C@H]1C[C@](C)(O)C[C@@H](OC(=O)C(C)=C)C2=C(CO)C(=O)O\C2=C\C(\C)=C/1
12885	CCOCC1=C2[C@@H](CC(C)(C)C[C@H](OC(C)=O)\C=C(\C)/C=C2/OC1=O)OC(=O)C(C)=C
12886	CCO[C@]12O[C@H]([C@H]3OC(=O)C(=C)[C@@H]3[C@H](C\C(C)=C/1)OC(=O)C(C)=C)C(C)=C2
12887	[H][C@@]12OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]1(C)CC[C@@H](O)[C@@H](COC(C)=O)[C@H]2O1)OC(=O)C(\COC(C)=O)=C/C
12888	[H][C@@]12OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]1(C)CC[C@@H](OC(C)=O)[C@@H](CO)[C@H]2O1)OC(=O)C(C)=C
12889	[H][C@@]12OC(=O)C(=C)[C@@H]1[C@@H](C[C@@]1(C)CC[C@@H](OC(C)=O)[C@@H](COC(C)=O)[C@H]2O1)OC(=O)C(\COC(C)=O)=C/C
12890	[H]C1[C@@H](OC(=O)C(\C)=C/C)[C@@H]2[C@@H](OC(=O)C2=C)\C=C(C)/[C@@H](O)[C@H]2O[C@H]2[C@]1(C)O
12891	C\C=C(\C)C(=O)O[C@@H]1\C=C(C)/[C@@H](O)C[C@H](OC(C)=O)\C(C)=C/[C@H]2OC(=O)C(=C)[C@@H]12
12892	C[C@@H]1O[C@]1(C)C(=O)O[C@@H]1\C=C(C)/[C@@H](O)C[C@H](OC(C)=O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@@H]12
12893	C\C=C(\C)C(=O)O[C@@H]1\C=C(C)/[C@@H](O)C[C@H](OC(C)=O)\C(CO)=C/[C@@H]2OC(=O)C(=C)[C@@H]12
12894	C\C=C(\C)C(=O)O[C@@H]1\C=C(C)/[C@@H](O)C[C@H](O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@@H]12
12896	[H][C@]12O[C@]1(C)CC[C@H](OC(C)=O)[C@](H)(C)C[C@H](OC(C)=O)C1=C(COC(C)=O)C(=O)O[C@H]21
12897	C\C=C(/C)C(=O)O[C@@H]1C[C@@]2(OC(C)=O)O[C@]1(C)CCC1=C(COC(C)=O)C(=O)O[C@@H]1\C=C2/C
12898	[H][C@]12O[C@]1(C)CCC(=O)[C@@](C)(C[C@H](OC(C)=O)C1=C(COC(C)=O)C(=O)O[C@H]21)OC(C)=O
12899	[H][C@]12O[C@]1(C)CCC(=O)[C@@](C)(C[C@H](OC(=O)C(C)=C)C1=C(COC(C)=O)C(=O)O[C@H]21)OC(C)=O
12900	[H][C@]12O[C@]1(C)CCC(=O)[C@@](C)(C[C@H](OC(=O)C(=C)CO)C1=C(COC(C)=O)C(=O)O[C@H]21)OC(C)=O
12901	[H][C@]12O[C@]1(C)CCC(=O)[C@@](C)(C[C@H](OC(=O)CC)C1=C(COC(C)=O)C(=O)O[C@H]21)OC(C)=O
12902	[H][C@@]12C[C@H](C)[C@H](O)C\C=C(C)\CC[C@@]1([H])[C@@H](C)C(=O)O2
12903	[H][C@@]12C[C@H](C)C(=O)C\C=C(C)\CC[C@@]1([H])[C@@H](C)C(=O)O2
12905	[H][C@@]12C[C@H](C)\C=C\C[C@@](C)(O)CCC[C@H](OC(=O)C(C)=C)[C@@]1([H])CC(=O)O2
12906	[H][C@]12OC(=O)C(OC(C)=O)=C1[C@H](CC(C)=[C@]=CC\C(C)=C/2)OC(C)=O
12907	[H][C@]12O[C@]1(C)CC=[C@]=C(C)C[C@H](OC(C)=O)C1=C(OC(C)=O)C(=O)O[C@]21[H]
12908	CC(=O)O[C@H]1CC(C)=[C@]=CC[C@@](C)(O)\C=C2\OC(=O)C(OC(C)=O)=C12
12909	[H]\C1=C(C)/C[C@H](O)C=C(C)C[C@@H](OC(=O)C(=C)CO)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12910	[H]\C1=C(C)/CCC=C(CO)C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12911	[H]\C1=C(C)/CCC=C(C[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H])C(O)=O
12912	[H]\C1=C(C)/CCC=C(C)[C@@H](O)[C@@H](OC(=O)C(\C)=C/C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12913	[H]\C1=C(C)/CC\C=C(C)\[C@@H](O)[C@@H](OC(=O)CC(C)C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12914	[H]\C1=C(C)/CCC=C(C)[C@@H](O)[C@@H](OC(=O)C2(C)OC2C)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12915	[H]\C1=C(C)/CCC=C(C)C[C@@H](OC(=O)CC(C)COC(=O)CCCCCCCCCCCCCCCC)[C@@]2([H])C(=C)C(=O)O[C@]12[H]
12916	[H][C@@]12C[C@H](C)C3=CC(=O)[C@@](C)(O3)[C@@H](O)[C@@H](OC(=O)C(\C)=C/C)[C@@]1([H])C(=C)C(=O)O2
12918	[H][C@]12OC(=O)[C@@]3(CO3)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@H](OC(C)=O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12921	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](OC(=O)C(C)=C)[C@H](OC(C)=O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12922	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](O)[C@H](OC(=O)C(C)=C)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12923	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](O)[C@H](OC(=O)C(\C)=C/COC(C)=O)[C@]1(C)OC(=CC1=O)\C(C)=C/2
12924	C\C=C(/C)C(=O)O[C@H]1[C@@H]2OC(=O)[C@]22CCC\C(=C\[C@H]3OC(=O)C(=C)[C@H]13)[C@H]2O
12925	[H][C@]1(O)O[C@@]23[C@H](OC(=O)C2=C)[C@@]1(C)CC\C=C(C)\[C@@H](OC(C)=O)[C@H]3OC(=O)C(\C)=C/C
12927	[H][C@]1(O)O[C@@]23[C@H](OC(=O)C2=C)[C@@]1(C)CC\C=C(C)\[C@@H](OC(C)=O)[C@H]3OC(=O)[C@@]1(C)O[C@H]1C
12929	[H][C@]1(O)O[C@@]23[C@H](OC(=O)C2=C)[C@@]1(C)CC\C=C(C)\[C@@H](O)[C@H]3OC(=O)C(\C)=C/C
12931	[H][C@]1(O)O[C@@]23[C@H](OC(=O)C2=C)[C@@]1(C)CC\C=C(C)\[C@@H](O)[C@H]3OC(=O)[C@@]1(C)O[C@H]1C
12933	C\C=C(\C)C(=O)O[C@H](C\C(C)=C\CO)[C@@H]1[C@H](OC(=O)C1=C)\C=C(/C)CO
12937	[H]C1CC(=C)[C@]2([H])CC(=O)[C@@H](C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
12938	[H][C@@]12CC(=O)[C@@H](C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)CC2=C
12939	[H][C@@]12[C@H](O)C=C(C)[C@]1(O)[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](C[C@]2(O)CCl)OC(=O)C(\C)=C/C
12940	[H][C@@]12[C@H](OC(C)=O)C=C(C)[C@]1(O)[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](C[C@]2(O)CCl)OC(=O)C(\C)=C/C
12941	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](O)(CCl)[C@]1([H])[C@H](OC(C)=O)C=C2C)OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C
12943	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@@](C)(O)[C@]12[H])OC(C)=O
12944	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3C([H])C[C@@](C)(O)[C@]12[H]
12945	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](O)(CCl)[C@]1([H])[C@H](O)C=C2C)OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C
12946	H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](O)(COC(=O)C(\CO)=C\C)[C@]1([H])[C@H](O)C=C2C)OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C
12947	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@@](C)(O)[C@]12[H])OC(=O)C(C)C
12948	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@@](C)(O)[C@]12[H])OC(=O)C(C)=C
12949	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@@](C)(O)[C@]12[H])OC(=O)C(\C)=C\C
12950	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@H](C)[C@]1([H])C(=O)C=C2CO)OC(=O)CCC1=CC=CC=C1
12951	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@H](C)[C@]1([H])C(=O)C=C2CO)OC(=O)C(C)=C
12952	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](C)(O)[C@]1([H])C(=O)C(O)=C2C)OC(=O)C(\C)=C\C
12953	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@H](C)[C@]1([H])C(=O)C=C2C
12954	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3C[C@@H](OC(=O)C(\C)=C/C)[C@@](C)(O)[C@@]11OO[C@]2(C)C=C1
12955	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(=O)C(\C)=C/C)[C@@](C)(OC(C)=O)[C@@]11OO[C@]2(C)C=C1
12957	[H][C@]12O[C@@]1(C)[C@]1(O)[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](O)(CCl)[C@]1([H])[C@@H]2O)OC(=O)C(\C)=C/C
12958	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@](O)(CCl)[C@]1([H])[C@H](O)C=C2C)OC(=O)C(\C)=C/COC(C)=O
12959	[H]O[C@H]1C[C@@](C)(O)[C@]2([H])C([H])[C@@]3([H])O[C@@]3(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
12960	[H][C@@]12O[C@]1(C)[C@]1([H])[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@](C)(O)[C@]11O[C@]21[H]
12962	[H][C@]12O[C@@]1(C)[C@]1([H])[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)C[C@@](C)(O)[C@@]11O[C@@]21[H]
12964	[H][C@]12CCC(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1CC[C@@]2(C)O
12965	[H][C@@]12CCC(=C)[C@]1([H])[C@H]1OC(=O)C(C)=C1CCC2=C
12966	[H][C@@]12CCC(=C)[C@]1([H])[C@H]1OC(=O)C(CO)=C1CCC2=C

ID	SMILES
12967	C[C@@H](O)C[C@H](O)C1=C[C@@H]2OC(=O)C(=C)[C@@H]2CC[C@@H]1C
12968	C[C@@H](O)C[C@H](OC(C)=O)C1=C[C@@H]2OC(=O)C(=C)[C@@H]2CC[C@@H]1C
12969	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)C=C1[C@@H](O)CC(C)=O
12970	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)C=C1[C@H](CC(C)=O)OC(C)=O
12971	C[C@@H](O)CCC1=C[C@@H]2OC(=O)C(=C)[C@@H]2C[C@@H](O)[C@@H]1C
12972	C[C@H](O)C[C@H](OC(C)=O)C1=CC[C@H]2[C@H](C[C@@H]1C)OC(=O)C2=C
12973	C[C@@H](C[C@H](O)C1=CC[C@H]2[C@H](C[C@@H]1C)OC(=O)C2=C)OC(C)=O
12974	C[C@H](O)CCC1=CC[C@H]2[C@H](C[C@@H]1C)OC(=O)C2=C
12976	C[C@H]1C[C@@H]2OC(=O)C(=C)[C@H]2CC=C1CCC(C)=O
12977	C[C@H]1C[C@@H]2OC(=O)C(=C)[C@H]2CC=C1[C@H](CC(C)=O)OC(C)=O
12979	C[C@H]1C[C@H]2OC(=O)C(=C)[C@H]2CC=C1[C@H](O)CC(C)=O
12980	C[C@H]1C[C@@H]2OC(=O)C(=C)[C@H]2CC=C1\C=C/C(C)=O
12982	C[C@H](O)C[C@H](O)C1=CC[C@H]2[C@H](C[C@@H]1C)OC(=O)C2=C
12983	[H][C@@]12C[C@H]3[C@@H](C[C@]1(C)[C@H]2CCC(C)=O)OC(=O)C3=C
12985	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@@H](O)C[C@@]1(C)C=C)C(=C)CO
12986	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@H](C[C@@]1(C)C=C)OC(=O)C(C)=C)C(=C)CO
12988	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@H](C[C@@]1(C)C=C)OC(=O)[C@H](C)CO)C(=C)CO
12989	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@H](C[C@@]1(C)C=C)OC(=O)C(=C)CO)C(=C)C=O
12990	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)C[C@@]1(COC(=O)C2=C)C=C
12991	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@@]1(COC(=O)C2=C)C=C)OC(=O)C(=C)CO
12992	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C\C)[C@@H]2OC(=O)[C@]12C=C)C(=C)C=O
12993	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C[C@]11CO[C@@H](O)C1)OC(=O)C(\C)=C/C)C(C)=C
12994	[H][C@]1([C@H]2OC(=O)[C@@H](C)[C@@H]2CC[C@@]1(C)C=C)C(C)=C
12995	[H][C@]1([C@H]2OC(=O)[C@@H](C)[C@@H]2[C@@H](O)C[C@@]1(C)C=C)C(=C)CO
12996	[H][C@]1([C@H]2OC(=O)[C@@H](C)[C@@H]2[C@H](C[C@@]1(C)C=C)OC(=O)[C@H](C)CO)C(=C)CO
12997	[H][C@]1([C@H]2OC(=O)[C@@H](C)[C@@H]2[C@@H](O)C[C@@]1(C)C=C)C(C)=C
12998	[H][C@@]12[C@@H](O)[C@H]3[C@H](C[C@@]1(COC(=O)C2=C)C=C)OC(=O)C3=C
12999	[H][C@]12[C@H](OC(=O)C1=C)[C@H]1[C@@H](OC(=O)C1=C)[C@@H](O)[C@@]2(C)C=C
13000	[H][C@@]12[C@H]3OC(=O)C1=COC=C[C@]2(C)C[C@@H]1OC(=O)C(=C)[C@@H]31
13001	[H][C@]1([C@H](OC(=O)C(\C)=C/C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)C=C)C(=C)C=O
13002	[H][C@]1([C@H](OC(=O)C(C)=C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)C=C)C(=C)C=O
13003	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(\C)=C/C)[C@@]1(C)C=C)C(=C)C=O
13004	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(C)=C)[C@@]1(C)C=C)C(=C)C=O
13005	[H][C@]1([C@H](OC(=O)C(\C)=C/C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(C)=O)[C@@]1(C)C=C)C(=C)C=O
13006	[H][C@]1([C@H](OC(C)=O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(\C)=C/C)[C@@]1(C)C=C)C(=C)C=O
13007	[H][C@]1([C@H](OC(=O)C(\C)=C/C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13008	[H][C@]1([C@H](OC(=O)C(C)=C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13009	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(\C)=C/C)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13010	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(C)=C)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13011	[H][C@]1([C@H](OC(=O)[C@@H](C)CC)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13012	[H][C@]1([C@H](OC(=O)C(C)C)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13013	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)[C@@H](C)CC)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13014	[H][C@]1([C@H](O)[C@H]2[C@@H](OC(=O)C2=C)[C@@H](OC(=O)C(C)C)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13016	[H][C@]1([C@H](OC(=O)C(C)=C)[C@H]2[C@H](OC(=O)C2=C)[C@@H](OC(C)=O)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13017	[H][C@]1([C@H](OC(=O)C(\C)=C/C)[C@H]2[C@H](OC(=O)C2=C)[C@@H](OC(=O)C(\C)=C/C)[C@@]1(C)[C@@H]1CO1)C(=C)C=O
13018	[H][C@@]12O[C@@H](O[C@H]1O)C(=C)[C@]1([H])C[C@H]3[C@@H](C[C@@]21COC(C)=O)OC(=O)C3=C
13019	[H][C@]1(C[C@H]2[C@@H](C[C@@]1(C)C=C)OC(=O)C2=C)C(C)=C
13023	[H][C@]1([C@H](OC(=O)C(C)C)[C@H]2[C@H](OC(=O)C2=C)[C@@H](O)[C@@]1(C)C=C)C(=C)C=O
13024	[H][C@]1([C@H](OC(=O)C(C)C)[C@H]2[C@H](OC(=O)C2=C)[C@@H](OC(C)=O)[C@@]1(C)C=C)C(=C)COC(C)=O
13026	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13027	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)CC[C@@]12O
13028	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(COC(C)=O)C(=O)CC[C@@]12O
13029	CC(=O)OC[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)CC[C@@]12O
13030	CC(C)C(=O)OC[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)CC[C@@]12O
13031	[H][C@@]12CCC(=O)[C@@]1(COC(C)=O)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13032	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2COC(=O)C(C)C
13033	[H][C@@]12CCC(=O)[C@@]1(COC(C)=O)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2COC(C)=O
13034	[H][C@@]12[C@@H](O)CC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13036	[H][C@@]12C[C@@H](O)C(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13037	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)C[C@@H]2C
13038	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1[C@H](C[C@@H]2C)OC(C)=O
13039	C[C@@]12[C@@H]3OC(=O)C(=C)[C@@H]3CC[C@@H]3CO[C@H](CC1=O)[C@]23O
13040	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13041	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)C=C[C@@]12O
13043	CC(=O)OC[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)C=C[C@@]12O
13044	[H][C@@]12C=CC(=O)[C@@]1(COC(C)=O)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13045	[H][C@@]12[C@H]3O[C@H]3C(=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13046	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)C(=O)CC=C12
13047	[H][C@@]12[C@@H](O)C[C@@H](O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13048	[H][C@@]12[C@H](C[C@@H](O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C)OC(C)=O
13049	[H][C@@]12C[C@@H](O)[C@@H](O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13050	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(COC(C)=O)[C@@H](O)CC[C@@]12O
13051	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(COC(C)=O)[C@H](CC[C@@]12O)OC(C)=O
13052	CC(=O)O[C@H]1CC[C@@]2(O)[C@@H](CO)CC[C@@H]3[C@@H](OC(=O)C3=C)[C@]12C
13053	CC(=O)OC[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]2(C)[C@H](CC[C@@]12O)OC(C)=O
13054	[H][C@@]12CC[C@@H](OC(C)=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2O
13055	[H][C@@]12CC[C@@H](OC(C)=O)[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2CO

ID	SMILES
13056	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H]1OC(=O)[C@]3(CO3)[C@@H]1CC[C@@H]2C
13057	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)[C@@H](C)[C@@H]1CC[C@@]2(C)O
13058	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H]1OC(=O)[C@H](C)[C@@H]1CC[C@@H]2C
13059	C[C@@H]1[C@@H]2CC[C@H](C)[C@]3(O)CCC(=O)[C@@]3(C)[C@@H]2OC1=O
13060	[H][C@@]12CCC(=O)[C@@]1(CO)[C@@H]1OC(=O)[C@@H](C)[C@@H]1CC[C@@H]2C
13061	[H][C@@]12CCC(=O)[C@@]1(C)[C@H](O)[C@H]1[C@H](C[C@@H]2C)OC(=O)C1=C
13062	[H][C@@]12CCC(=O)[C@@]1(C)C[C@H]1[C@H](C[C@]22CO2)OC(=O)C1=C
13063	[H][C@@]12[C@H](C[C@H](OC(C)=O)[C@@]1(C)[C@@H](O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C)OC(C)=O
13064	[H][C@@]12CCC(=O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13065	C[C@H]1C[C@H]2OC(=O)C(=C)[C@H]2C[C@]2(C)C(=O)CC[C@@]12O
13066	[H][C@@]12[C@@H](O)CC(=O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13067	[H][C@@]12[C@@H](C[C@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C)OC(C)=O
13068	[H][C@@]12C[C@H](O)[C@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13069	[H][C@@]12C[C@H](OC(C)=O)[C@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13070	[H][C@@]12C[C@H](OC(C)=O)[C@H](OC(C)=O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13071	[H][C@@]12CC[C@@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13072	[H][C@@]12CC(=O)[C@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13073	[H][C@@]12CC[C@H](O)[C@@]1(CO)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13074	[H][C@@]12CC[C@H](OC(C)=O)[C@@]1(CO)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13075	[H][C@@]12CC[C@H](O)[C@@]1(COC(C)=O)C[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
13076	C[C@H]1CC[C@@H]2[C@@H](OC(=O)C2=C)[C@](C)(O)[C@@]11CCC(=O)O1
13078	[H][C@@]12CCC(=O)O[C@@]1(C)[C@@H]1OC(=O)C(=C)[C@@H]1CC[C@@H]2C
13079	CC1=C2CCC(=O)O[C@@]2(C)[C@@H]2OC(=O)C(=C)[C@@H]2CC1
13080	CCOC(=O)CCC1=C(C)CC[C@@H]2[C@@H](OC(=O)C2=C)[C@]1(C)O
13082	[H][C@]12CC(=O)O[C@]11[C@@H](C)CC[C@@H]3[C@@H](OC(=O)C3=C)[C@]21C
13085	[H][C@@]12C=CC(=O)[C@@]1(C)C[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13086	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](O)[C@H]1[C@@H](OC(=O)C1=C)[C@H](OC(C)=O)[C@H]2C
13087	[H][C@@]12C=CC(=O)[C@@]1(C)[C@H](O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13088	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(C)=O)[C@H]1[C@@H](OC(=O)C1=C)[C@H](OC(C)=O)[C@H]2C
13090	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(\C)=C/C)[C@H]1[C@@H](OC(=O)C1=C)[C@H](O)[C@H]2C
13091	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C=C(C)C)[C@H]1[C@@H](OC(=O)C1=C)[C@H](O)[C@H]2C
13092	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(C)=O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13095	C[C@@H]1C[C@@H]2OC(=O)C(=C)[C@H]2[C@@H](OC(C)=O)[C@]2(C)C(=O)CC=C12
13096	C[C@@H]1C[C@@H]2OC(=O)C(=C)[C@H]2[C@H](O)[C@]2(C)C(=O)CC=C12
13097	[H][C@@]12O[C@]1([H])[C@@]1([H])[C@H](C)C[C@@H]3OC(=O)C(=C)[C@H]3C[C@]1(C)[C@@H]2O
13098	[H][C@@]12[C@@H](O)C[C@@H](O)[C@@]1(C)C[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13099	[H][C@@]12[C@H](O)C[C@@H](O)[C@@]1(C)[C@H](OC(C)=O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13100	[H][C@@]12[C@@H](O)C[C@@H](O)[C@@]1(C)[C@H](OC(C)=O)[C@H]1[C@@H](OC(=O)C1=C)[C@H](OC(C)=O)[C@H]2C
13101	[H][C@@]12[C@@H](C[C@@H](O)[C@@]1(C)[C@@H](O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C)OC(C)=O
13102	[H][C@@]12[C@@H](C[C@@H](O)[C@@]1(C)[C@@H](OC(C)=O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C)OC(C)=O
13103	[H][C@@]12[C@H](C[C@@H](O)[C@@]1(C)[C@H](OC(C)=O)[C@H]1[C@@H](OC(=O)C1=C)[C@H](O)[C@H]2C)OC(C)=O
13104	[H][C@@]12[C@H](C[C@@H](O)[C@@]1(C)[C@H](OC(C)=O)[C@H]1[C@@H](OC(=O)C1=C)[C@H](O)[C@H]2C)OC(=O)CC(C)C
13105	[H][C@@]12O[C@]1([H])[C@@]1(C)[C@H](OC(C)=O)[C@H]3[C@H](C[C@@H](C)[C@]1([H])[C@@H]2O)OC(=O)C3=C
13106	[H][C@@]12C=CC(=O)[C@@]1(C)[C@H](O)[C@H]1[C@H](C[C@H]2C)OC(=O)[C@@H]1C
13107	[H][C@@]12C=CC(=O)[C@@]1(C)[C@H](OC(C)=O)[C@H]1[C@H](C[C@H]2C)OC(=O)[C@@H]1C
13108	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(\C)=C/C)[C@H]1[C@H](C[C@H]2C)OC(=O)[C@@H]1C
13109	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(\C)=C/C)[C@@H]1[C@H](C)C(=O)O[C@H]1[C@H](O)[C@H]2C
13110	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C=C(C)C)[C@@H]1[C@H](C)C(=O)O[C@H]1[C@H](O)[C@H]2C
13111	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(C)=C)[C@@H]1[C@H](C)C(=O)O[C@H]1[C@H](O)[C@H]2C
13112	[H][C@@]12C=CC(=O)[C@@]1(C)C[C@H]1[C@@H](OC(=O)[C@]1(O)CO)[C@H](OC(C)=O)[C@H]2C
13113	[H][C@@]12[C@@H](O)C[C@@H](O)[C@@]1(C)C[C@H]1[C@H](C[C@H]2C)OC(=O)[C@@H]1CN1CCCCC1
13114	[H][C@@]12[C@@H](O)CC(=O)[C@@]1(C)[C@@H](OC(=O)C(\C)=C/C)[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13115	[H][C@@]12[C@@H](CC(=O)[C@@]1(C)[C@@H](O)[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C)OC
13116	[H][C@@]12[C@H](CC(=O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C)O[C@@]1([H])O[C@H](COCC)[C@@H](O)[C@H](O)[C@H]1O
13117	[H][C@@]12[C@H]3CC(=O)[C@@]1(CO3)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13129	[H][C@@]12[C@@H](OC(C)=O)[C@@H](OC(C)=O)[C@H](OC(C)=O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13132	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(C)=C)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13133	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(C)C)C1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13134	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)[C@@H](C)CC)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13135	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)CC(C)C)C1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13137	[H][C@@]12[C@@H](O)CC(=O)[C@@]1(C)[C@@H](OC(=O)C=C(C)C)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13138	[H][C@@]12[C@H](CC(=O)[C@@]1(C)[C@@H](O)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@]1(O)CO)OC(=O)[C@@H](C)CC
13139	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H](O)[C@@H]1[C@H](C)C(=O)O[C@@H]1C[C@H]2C
13142	[H][C@@]12[C@@H](OC(C)=O)[C@@H](OC(C)=O)[C@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@H]1C
13143	[H][C@@]12[C@@H](O)C[C@@H](O)[C@@]1(C)[C@@H](O)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@H]1CN1CCCCC1
13144	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](O)[C@H]1[C@@H](C[C@H]2C)OC(=O)[C@]1(O)CO
13145	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C=C(C)C)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13146	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(\C)=C/C)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13148	[H][C@@]12O[C@]1([H])[C@@]1([H])[C@H](C)C[C@H]3OC(=O)C(=C)[C@H]3[C@H](O)[C@]1(C)[C@H]2O
13149	[H][C@@]12CC(=O)OC[C@@]1(C)C[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13151	[H][C@@]12C[C@H](O)O[C@@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13152	[H][C@@]12[C@@H](OC(C)=O)[C@H](O)O[C@@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13153	[H][C@@]12[C@@H](OC(=O)C(\C)=C\C)[C@H](O)O[C@@H](O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13154	[H][C@@]12[C@@H](OC(C)=O)[C@H](O)O[C@@H](OCC)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13155	[H][C@@]12[C@@H](OC(=O)C(\C)=C\C)[C@H](O)O[C@@H](OCC)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13156	[H][C@@]12C[C@H](O)O[C@@H](OCC)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13157	[H][C@@]12CCOC(=O)[C@@]1(C)C[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C

ID	SMILES
13160	[H][C@]12[C@H](C)C[C@@H]3OC(=O)C(=C)[C@H]3C[C@]1(C)[C@H](O)OC2=O
13161	C[C@@H]1C[C@H]2OC(=O)C(=C)[C@H]2C[C@H]2CC(=O)C(C)=C12
13162	C[C@H]1[C@@H](O)[C@H]2[C@@H](CC(C)=C1CCCO)OC(=O)C2=C
13163	C[C@H]1[C@@H](OC(C)=O)[C@H]2[C@@H](CC(C)=C1CCCO)OC(=O)C2=C
13165	[H][C@@]12C[C@H]3[C@@H](C[C@@H](C)[C@]1([H])C=CC2=O)OC(=O)C3=C
13166	C[C@H]1CCC=C2C[C@H]3OC(=O)C(=C)[C@H]3C[C@]12C
13167	C[C@@H]1[C@H]2C[C@]3(C)[C@@H](C)CCC=C3C[C@H]2OC1=O
13168	[H][C@]12C[C@H]3OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C(=O)C=C2
13169	[H][C@]12C[C@@]3(O)OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C(=O)C=C2
13170	[H][C@]12C[C@@]3(OC)OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C(=O)C=C2
13171	[H][C@]12C[C@@]3(OCC)OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C(=O)C=C2
13172	C[C@H]1CCC=C2C=C3OC(=O)C(C)=C3C[C@]12C
13174	C[C@H]1CCC=C2C=C3OC(=O)C(C)=C3[C@@H](OC(C)=O)[C@]12C
13175	CO[C@@H]1C2=C(C)C(=O)OC2=C[C@@]2(O)CCC[C@H](C)[C@@]12C
13176	[H][C@]12CCC[C@H](C)[C@@]1(C)CC1=C(C)C(=O)O[C@H]1C2
13177	[H][C@@]12C[C@@H]3OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C[C@H](C2)OC(=O)C(\C)=C\C
13179	[H][C@@]12C[C@@H]3OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C[C@H](C2)OC(=O)\C=C/SC
13181	[H][C@]12CCC[C@H](C)[C@@]1(C)[C@H](O)C1=C(C)C(=O)O[C@H]1C2
13182	[H][C@]12CCC[C@H](C)[C@@]1(C)[C@H](O)C1=C(C)C(=O)O[C@@]1(O)C2
13183	[H][C@@]12C[C@@H]3OC(=O)C(C)=C3C[C@]1(C)[C@@H](C)C[C@H](C2)OC(=O)C(\CO)=C/C
13184	[H][C@]12C[C@@H](C[C@H](C)[C@@]1(C)CC1=C(C)C(=O)O[C@@]1(O)C2)OC(=O)C(\CO)=C/C
13185	[H][C@]12C[C@@H](C[C@H](C)[C@@]1(C)CC1=C(C)C(=O)O[C@@]1(O)C2)OC(=O)C(\C)=C/C
13186	[H][C@@]12CCC[C@@H]3C(=O)O[C@@H](C4=C(C)C(=O)O[C@H]4C1)[C@@]23C
13187	[H][C@@]12CCC[C@@H]3C(=O)O[C@@H](C4=C(C)C(=O)O[C@@]4(O)C1)[C@@]23C
13189	[H][C@@]12CCC[C@@H]3C(=O)O[C@@H](C4=C(C1)OC=C4C)[C@@]23C
13190	[H][C@@]12CCC[C@@H]3C(=O)O[C@@H](C4=C(OC=C4C)[C@@H]1OC(=O)C(\C)=C\C)[C@@]23C
13191	C[C@H]1CCCC2=C[C@]3(OC(=O)C(C)=C3C[C@]12C)[C@]12OC(=O)C(C)=C1C[C@]1(C)[C@@H](C)CCCC1=C2
13193	[H]C1[C@@]2([H])CC([H])C([H])[C@H](C)[C@@]2(C)C[C@@]11C(=C)COC1=O
13194	[H]C1C([H])[C@@H](OC(=O)C(\C)=C/C)[C@]2([H])[C@@H](O)[C@@]3(C[C@]2(C)[C@H]1C)C(=C)COC3=O
13195	[H]C1C([H])[C@@H](OC(=O)C(\C)=C/C)[C@]2([H])[C@@H](OC(C)=O)[C@@]3(C[C@]2(C)[C@H]1C)C(=C)COC3=O
13196	[H]C1C([H])[C@@H](OC(=O)\C=C/SC)[C@]2([H])[C@@H](OC(C)=O)[C@@]3(C[C@]2(C)[C@H]1C)C(=C)COC3=O
13198	[H]C1C([H])[C@H](OC(=O)C(\C)=C/C)[C@]2([H])[C@@H](OC(=O)C(\C)=C/C)[C@@]3(C[C@]2(C)[C@H]1C)C(=C)COC3=O
13199	[H]C1C([H])[C@H](OC(=O)[C@@H](C)CC)[C@]2([H])[C@@H](OC(C)=O)[C@@]3(C[C@]2(C)[C@H]1C)C(=C)COC3=O
13200	[H]C1[C@@]2([H])C[C@@H](OC(=O)C(\C)=C/C)C([H])[C@H](C)[C@@]2(C)C[C@@]11C(=C)COC1=O
13201	[H]C1[C@@]2([H])CC([H])[C@@H](OC(=O)C(\C)=C\C)[C@H](C)[C@@]2(C)C[C@@]11C(=C)COC1=O
13202	[H]C1C[C@]2([H])[C@@H](OC(C)=O)[C@@]3(C[C@]2(C)[C@@H](C)C1[H])C(=C)COC3=O
13203	[H][C@@]12CC[C@@]3(C)O[C@@]45[C@H](C[C@]1(C)[C@@]4(OC(=O)C5=C)[C@@]23C)OC(=O)C(\C)=C\C
13204	[H][C@@]12CC[C@@]3(C)O[C@@]45[C@H](C[C@]1(C)[C@@]4(OC(=O)C5=C)[C@@]23C)OC(=O)C(C)=C
13205	[H]C1[C@@H]2C[C@@]34[C@@H]5C[C@@H](O)[C@H]3[C@]2(COC(=O)[C@@H](C)CC)C(=O)O[C@@H]1C4=CO[C@H]5OC(=O)CC(C)(C)OC(C)=O
13206	CC(C)CC(=O)O[C@H]1[C@@H]2C[C@@]34[C@@H]5C[C@@H](O)[C@H]3[C@]2(C)C(=O)O[C@@H]1C4=CO[C@H]5OC(=O)CC(C)(C)OC(C)=O
13207	[H]C1[C@@H]2C[C@@]34[C@@H]5C[C@@H](O)[C@H]3[C@]2(C)C(=O)O[C@@H]1C4=CO[C@H]5OC(=O)CC(C)(C)OC(C)=O
13208	[H]C1[C@@H]2C[C@@]34[C@@H]5CC[C@H]3[C@]2(C)C(=O)O[C@@H]1C4=CO[C@H]5OC(=O)CC(C)C
13209	[H]C1[C@@H]2C[C@@]34[C@@H]5CC[C@H]3[C@]2(C)C(=O)O[C@@H]1C4=CO[C@H]5OC(=O)[C@@H](C)CC
13211	[H][C@@]12[C@H]3OC(=O)[C@@H](C)[C@@H]3[C@H](CC(C)=C1C=C[C@@]2(C)O)OC(C)=O
13212	[H][C@@]12[C@H]3OC(=O)[C@@H](C)[C@@H]3CC[C@@H](C)[C@]1([H])C(=O)C=C2C
13216	[H][C@]12CC=C(C)[C@@]1([H])[C@H]1OC(=O)[C@](C)(OC(C)=O)[C@@H]1[C@H](C[C@]2(C)O)OC(=O)C(\C)=C/C
13220	[H][C@@]12CC[C@@]3(C)O[C@H]3[C@@]11OC(=O)C(=C)[C@@H]1CC[C@H]2C
13234	C[C@@H]1C(=O)OC2=C1C(C)=C1[C@@H](C)CCCC1=C2O
13244	[H][C@@]12CC3=C(C[C@@]1(C)CC[C@H](OC(C)=O)C2=C)OC=C3C
13245	[H][C@@]12CC3=C(C[C@@]1(C)CC[C@H](O)C2=C)OC=C3C
13246	[H][C@]12C[C@H](C)C3=CC(=O)[C@](C)(C[C@]4([H])O[C@](O)(O)(\C=C/C)[C@@](C)(C(=O)O1)[C@@]24[H])O3
13247	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@H](C[C@]13[C@H](C)CC[C@@](CO[C@@H]1O)(C1=CC=CC=C1)[C@]23[H])OC(=O)C(C)=C
13248	[H]C1[C@H]2[C@H](C)C(=O)O[C@@H]2[C@@]2([H])C(=C)CC[C@@]2([H])C(=C)[C@H]1O
13249	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](C[C@]1(CO1)[C@]1([H])[C@H](O)C=C(C)[C@]21[H])OC(=O)C(\COC(=O)C(\CO)=C\C)=C\C
13250	[H][C@@]12[C@H](O)C[C@@H](O)[C@@]1(C)[C@H](O)[C@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13279	[H]C1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)O)[C@](C)(O)[C@@H]1OC(C)=O
13280	CC(C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)O)[C@](C)(O)[C@@H]1OC(C)=O
13281	[H]C1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@](C)(O)[C@@H]1OC(C)=O
13282	C[C@@]1(OO)C=C[C@@H]2[C@H]1[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)[C@@H](O)[C@@]2(C)O
13283	CC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@](C)(O)[C@@H]1O
13284	CC(=O)O[C@@H]1[C@H](O)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@]1(C)O
13285	CC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@](C)(O)[C@@H]1OC(C)=O
13286	CC(C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@](C)(O)[C@@H]1OC(C)=O
13287	C\C=C(/C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@@H](C=C[C@@]2(C)OO)[C@](C)(O)[C@@H]1OC(C)=O
13291	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13292	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(C)=O)[C@@H](O)[C@@](C)(O)[C@@H]12
13293	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(C)=O)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13294	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(C)C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13295	[H]C1[C@H](O)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@H]([C@](C)([H])O)C=C2C)[C@]1(C)O
13296	[H][C@]1(OO)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(C)C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13297	[H]C1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@H]([C@](C)([H])OC(C)=O)C=C2C)[C@](C)(O)[C@@H]1OC(C)=O
13298	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(C)=O)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13299	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13300	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(=O)C(C)C)[C@@](C)(O)[C@@H]12
13301	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(=O)CC(C)C)[C@@](C)(O)[C@@H]12
13302	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(=O)[C@H](C)CC)[C@@](C)(O)[C@@H]12
13303	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(C)C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12

ID	SMILES
13304	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](O)[C@@](C)(O)[C@@H]12
13305	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(\C)=C/C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13306	[H][C@@]1(O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](O)[C@@](C)(O)[C@@H]12
13307	[H][C@@]1(O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(\C)=C/C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13308	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)[C@@H](C)CC)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13309	[H][C@@]1(OC(C)=O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)CC(C)C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13310	[H][C@@]1(O)C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13311	[H]C1([H])C=C(C)[C@@H]2[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(\C)=C\C)[C@@H](OC(C)=O)[C@@](C)(O)[C@@H]12
13314	[H]C1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O
13316	CC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1O
13317	CC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O
13318	CC(C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O
13319	CCC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O
13320	C\C=C(/C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@H](O)C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O
13321	CC(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@@H](C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O)OO
13322	CC(C)C(=O)O[C@@H]1[C@@H]2[C@H](OC(=O)C2=C)C2=C(C)[C@@H](C[C@@H]2[C@](C)(O)[C@@H]1OC(C)=O)OO
13323	C\C=C(\C)C(=O)O[C@H]1CC(=C)[C@H]2C[C@H]3O[C@@]3(C)[C@@H]2[C@H]2OC(=O)C(=C)[C@H]12
13324	CC(C)CC(=O)O[C@H]1CC(=C)[C@H]2C[C@H]3O[C@@]3(C)[C@@H]2[C@H]2OC(=O)C(=C)[C@H]12
13327	[H]C1[C@H]2[C@H](C[C@](I)([H])(C)[C@]3(O)CC(=O)C(C)=C13)OC(=O)C2=C
13329	[H]C1[C@H]2[C@H](C[C@@](C)(O)[C@]3([H])CC(=O)C(C)=C13)OC(=O)[C@]2([H])C
13330	[H]C1C[C@@](C)(O)[C@@]23C=C[C@@](C)(C[C@@]22[C@H]4CC5=C(C)C(=O)C[C@@H]5[C@@H](C)C[C@@H]4OC2=O)[C@]3([H])[C@H]2OC(=O)C(=C)[C@H]12
13331	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](O)C[C@@](C)(O)[C@]11C=C[C@@]2(C)C[C@]11[C@H]2CC3=C(C)C(=O)C[C@@H]3[C@@H](C)C[C@@H]2OC1=O
13333	CC(=O)O[C@@H]1[C@@H]2O[C@]2(C)CC\C=C(C)\C[C@@H]2OC(=O)C(=C)[C@@H]12
13334	CC(=O)O[C@H]1C[C@@H]2O[C@]2(C)C[C@@H]2OC(=O)C(=C)[C@H]2[C@H](O)\C=C1/C
13335	CC(=O)O[C@H]1C[C@@H](O)C(=C)C[C@@H](O)[C@@H]2[C@H](OC(=O)C2=C)\C=C1/C
13336	CC(=O)O[C@H]1C[C@@H](OO)C(=C)C[C@@H](O)[C@@H]2[C@H](OC(=O)C2=C)\C=C1/C
13340	[H]C1[C@H](OC(=O)C=C(C)C)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C1/C
13342	[H]C1[C@H](OC(=O)C(C)C)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C1/C
13343	[H]C1[C@H](OC(=O)[C@@H](C)CC)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C1/C
13344	[H]C1[C@H](OC(=O)CC(C)C)[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C1/C
13345	[H]C1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C(C)\[C@@H]1O
13346	[H]C1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C(C)\[C@@H]1OC(C)=O
13348	CC(=O)O[C@@H]1C\C(C)=C/C[C@@H](OC(C)=O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12
13349	[H][C@@]1(C)[C@H]2[C@@H](OC1=O)\C=C(C)/[C@@H](C\C=C(C)/C[C@H]2O)OC(C)=O
13350	C[C@@]12CC[C@@H]3[C@H](OC(=O)C3=C)[C@@H]3O[C@]3(C)CC[C@H]1O2
13354	C\C1=C/[C@@H](O)[C@H]2[C@@H](C[C@]3(C)O[C@H]3CC1)OC(=O)C2=C
13360	C\C=C(\C)C(=O)O[C@H]1C[C@@]2(C)[C@@H](O)CC=C(C)[C@@H]2[C@H]2OC(=O)C(=C)[C@H]12
13362	CC(C)C(=O)O[C@H]1C[C@@]2(C)[C@@H](O)CC=C(C)[C@@H]2[C@H]2OC(=O)C(=C)[C@H]12
13363	CC1=CC[C@H](O)[C@]2(C)C[C@H](O)[C@@H]3[C@H](OC(=O)C3=C)[C@@H]12
13366	[H][C@@]12[C@@H](O)[C@H]3[C@H](C[C@@]1(C)[C@H](O)CC=C2C)OC(=O)C3=C
13368	[H][C@@]12[C@@H](O)[C@H]3[C@H](C[C@@]1(C)[C@H](O)CCC2=C)OC(=O)C3=C
13369	CC(C)=CC(=O)C\C(C)=C\C[C@H]1COC(=O)C1=C
13370	CC(C)=CC(=O)\C=C(\C)CC[C@H]1COC(=O)C1=C
13374	[H][C@]1(C)[C@@H]2CCC(C)=C3C(=O)C=C(CO[C@]4(C)O[C@H](CO)[C@@](C)(O)[C@@](C)(O)[C@@]4(C)O)[C@]3([H])[C@H]2OC1=O
13376	[H][C@]1(C)[C@H]2[C@H](OC1=O)[C@@]1([H])C(CO)=CC(=O)C1=C(C)C[C@@H]2O
13378	[H][C@@]12C[C@H](O[C@]3([H])O[C@@](C)(CO)[C@](C)(O)[C@](C)(O)[C@]3(C)O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1CCC2=C
13396	[H][C@@]12CCC(=O)[C@@]1(C)[C@@H](OC(C)=O)[C@@H]1[C@H](C[C@H]2C)OC(=O)C1=C
13397	C[C@@H](CCCOC(C)=O)C1=C(C)C[C@@]2(C)OC(=O)C(=C)[C@@]2(C)[C@H]1O
13398	C[C@@H](CCCOC(C)=O)C1=C(C)C[C@@]2(C)OC(=O)C(=C)[C@@]2(C)[C@H]1OC(C)=O
13399	C[C@H]1C[C@@H]2OC(=C)C(=C)[C@H]2CC2=C(C)[C@@H]3[C@H](OC(C)=O)[C@]12C[C@@]31[C@H]2[C@H](CC(C)=C(CCCCOC(C)=O)[C@@H]2O)OC1=O
13400	C[C@H]1C[C@@H]2OC(=O)C(=C)[C@H]2CC2=C(C)[C@@H]3[C@H](OC(C)=O)[C@]12C[C@@]31[C@H]2[C@H](OC1=O)\C=C(C)/CC\C=C(C)/C[C@H]2O
13401	C[C@H]1C[C@@H]2OC(=O)C(=C)[C@H]2CC2=C(C)[C@@H]3[C@H](OC(C)=O)[C@]12C[C@@]31[C@H]2[C@H](CC(C)=C(CCCCOC(C)=O)[C@@H]2O)OC1=O
13405	C[C@@H](CCCOC(C)=O)C1=C[C@]2(C)[C@H](C)C(=O)O[C@]2(C)C[C@@]1(C)O
13406	CC[C@@H](C)C(=O)O[C@H]1C([C@@H](C)CCCO)=C(C)C[C@@]2(C)OC(=O)C(=C)[C@@]12C
13407	CC(=O)CC[C@@H]1[C@H](OC(=O)C1=C)C1=C(C)CCC1=O
13408	[H]C1[C@@H](OC)C(C)=C([C@H]2OC(=O)C(=C)[C@@H]2CCC(C)=O)C1=O
13410	CC(=O)CC[C@@H]1[C@H](OC(=O)C1=C)C1=C(C)[C@H](O)[C@@H](Cl)C1=O
13411	CC(=O)CC[C@@H]1[C@H](OC(=O)C1=C)C1=C(C)C(=O)C[C@@H]1O
13412	[H][C@]1([C@H]2OC(=O)C(=C)[C@@H]2CCC(C)=O)C(=O)C=C[C@@]1(C)O
13413	C[C@H]1[C@H](CCC(C)=O)[C@H](OC1=O)C1=C(C)CCC1=O
13946	C[C@H]1C[C@H]2OC(=O)C(=C)[C@H]2C[C@H]2O[C@@]12\C=C\C(C)=O
13947	[H][C@@]12C[C@H](C)[C@@]3(O)C=C(C(C)=O)[C@@]4(C[C@]3([H])C[C@]1([H])C(=C)C(=O)O2)C(=O)O[C@]1([H])C[C@H](C)C(\C=C\C(C)=O)=CC[C@]41[H]
13948	[H][C@@]12C[C@H](C)[C@@]3(O)C=C(C(C)=O)[C@@]4(C[C@]3([H])C[C@]1([H])C(=C)C(=O)O2)C(=O)O[C@]1([H])C[C@H](C)[C@@]2(O[C@@H]2C[C@]41[H])\C=C\C(C)=O
13949	[H][C@@]12C[C@H]3[C@@H](C[C@H](C)C1=CC(=O)O2)OC(=O)C3=C
13950	[H][C@]12C[C@H](C)[C@]34C=C[C@]5(C)C[C@@]3(C(=O)O1)[C@@]2([H])C[C@@]4([H])O5
13952	[H][C@@]1(COC(=O)C1=C)[C@@H](O)\C=C(/C)CC(=O)C=C(C)C
13953	[H][C@@]1(COC(=O)C1=C)[C@@H](OC(C)=O)\C=C(/C)CC(=O)C=C(C)C
13954	CC(C)CC(=O)O[C@@H]1[C@@H](O)[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@]1(C)O)OC(=O)C2=C
13955	CC(C)CC(=O)O[C@H]1[C@@H]2[C@@H](C[C@@H](C)\C=C/C(=O)[C@](C)(O)[C@@H]1O)OC(=O)C2=C
13958	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)C(C)C)[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13963	[H][C@@]12C=CC(=O)[C@@]1(C)[C@@H](OC(=O)CC(C)C)[C@H]1[C@@H](C[C@H]2C)OC(=O)C1=C
13968	[H]C1[C@H]2[C@@H](C[C@@H](C)[C@]3([H])[C@H](C[C@@H](OC(C)=O)[C@@]13C)OC(C)=O)OC(=O)C2=C
14040	[H][C@]12[C@H](C)C(=O)O[C@@H]1[C@@]1([H])C(=C)[C@@H](O)C[C@@]1([H])C(=C)C[C@@H]2O
14218	[H][C@@]12C[C@@H](OC(=O)C(\C)=C\C)[C@@H]3[C@@H](OC(=O)C3=C)\C=C(C)/[C@@H](O)C[C@H]1O2
14267	[H][C@@]12C[C@@H](OC(=O)C(C)C)[C@@H]3[C@@H](OC(=O)C3=C)\C=C(C)/[C@@H](O)C[C@H]1O2
14268	[H][C@@]12C[C@@H](OC(=O)[C@@H](C)CC)[C@@H]3[C@@H](OC(=O)C3=C)\C=C(C)/[C@@H](O)C[C@H]1O2
14270	C\C=C(\C)C(=O)O[C@@H]1C[C@@]2(O)O[C@@H]2C[C@H](O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12

ID	SMILES
14273	CC[C@H](C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)\C(CO)=C/[C@@H]2OC(=O)C(=C)[C@H]12
14276	CC(C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)C(=C)[C@H](O)[C@@H]2OC(=O)C(=C)[C@H]12
14277	[H]C1[C@@H]2OC(=O)C(=C)[C@@H]2[C@@H](C[C@@]2(C)OC(=CC2=O)C1=C)OC(=O)C(C)C
14278	[H]C1[C@@H]2OC(=O)C(=C)C2[C@@H](C[C@@]2(C)OC(=CC2=O)C1=C)OC(=O)C(C)=C
14279	[H]C1[C@@H]2OC(=O)C(=C)C2[C@@H](C[C@@]2(C)OC(=CC2=O)C1=C)OC(=O)C(\C)=C/C
14281	[H][C@]1(C)[C@H]2CC\C(C)=C\[C@H](C\C(C)=C\[C@H]2OC1=O)O[C@]1([H])O[C@@H](COC(O)=O)[C@H](OC(O)=O)[C@H]1OC(O)=O
14282	[H]OC[C@H]1O[C@@]([H])(O[C@H]2C\C(C)=C\[C@H]3OC(=O)[C@@]([H])(C)[C@H]3CC\C(C)=C\2)[C@@H](O[H])[C@@H](O[H])[C@@H]1O[H]
14283	[H]O[C@H]1C\C(C)=C\[C@H]2OC(=O)[C@@]([H])(C)[C@H]2CC\C(C)=C\1
14284	[H][C@]1(C)[C@H]2CC\C(C)=C\[C@H](C\C(C)=C\[C@H]2OC1=O)OC(O)=O
14285	[H][C@]1(C)[C@H]2CC[C@@]3(C)C[C@@H](O)CC(=C)[C@@H]3[C@H]2OC1=O
14288	C[C@@]12C[C@H]3OC(=O)C(=C)[C@H]3C[C@@H]1C(=C)[C@@H](OC(O)=O)C=C2
14290	C[C@@]12CC[C@H](C[C@H]1C(=C)C(=O)C=C2)C(=C)C(O)=O
14293	[H]O[C@@H]1C[C@@H]2[C@H](OC(=O)C2=C)[C@@]2([H])C(=C)[C@H](C[C@@]2([H])C1=C)O[C@]1([H])O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O
14294	[H][C@]1(C[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)O[C@H]1C\C=C(CO)\CC[C@H]2[C@H](C)C(=O)O[C@@H]2\C=C1/C
14296	[H]C1CC(C)=C2C(=O)C=C(CO[C@]3([H])O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)[C@]2([H])[C@H]2OC(=O)[C@@]([H])(C)[C@H]12
14297	[H]O[C@@H]1C[C@@H]2[C@H](OC(=O)[C@@]2([H])C)[C@@]2([H])C(=C)[C@H](O[C@]3([H])O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)C(=O)[C@@]2([H])C1=C
14299	C\C=C(\C)C(=O)O[C@H]1C[C@]2(C)CC[C@@](O)(O2)\C(C)=C/[C@H]2CC(=O)C(=C)[C@H]12
14301	[H][C@]12C[C@@]3(O)[C@H](CC[C@H](C)[C@@]3(C)C[C@@H]1[C@H](C)C(=O)O2)OC(=O)C(\C)=C/C
14302	[H][C@@]12C[C@]3(O)OC(=O)[C@@H](C)[C@H]3C[C@]1(C)[C@@H](C)CC[C@@H]2OC(=O)C(\C)=C/C
14303	C\C=C(\C)C(=O)O[C@H]1CC[C@H](C)[C@@]2(C)C[C@H]3[C@H](C)C(=O)O[C@@]3(O)C[C@@]12O
14319	C\C=C(\C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)\C(C)=C/[C@H]2OC(=O)C(=C)[C@H]12
14320	C\C=C(/C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)\C(C)=C/[C@H]2OC(=O)C(=C)[C@H]12
14329	C\C=C(/C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)\C(CO)=C/[C@H]2OC(=O)C(=C)[C@H]12
14331	C\C=C(\C)C(=O)O[C@@H]1C[C@@]2(C)OC(=CC2=O)\C(CO)=C/[C@H]2OC(=O)C(=C)[C@H]12
14336	C[C@H]1[C@H]2C=C3C(C)=CC(=O)C3=C(C)C[C@H]2OC1=O
14341	C[C@]12CC[C@@H](O)\C(C1)=C\C[C@@H]1[C@H](OC(=O)C1=C)[C@H]2O
14342	C\C=C(\C)C(=O)O[C@@H]1CC[C@]2(C)O[C@@H]2[C@H](O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12
14343	C\C=C(\C)C(=O)O[C@@H]1CC[C@]2(C)O[C@@H]2[C@H](OC(C)=O)\C(C)=C/[C@@H]2OC(=O)C(=C)[C@H]12
14346	C\C1=C/[C@H]2OC(=O)C(=C)[C@@H]2[C@H](O)C\C(CO)=C\CC1
14348	[H][C@@]12CC[C@@](O)(O1)\C(CO)=C/[C@H]1OC(=O)C(=C)[C@@H]1[C@H](CC(=O)C(C)C)C2
14349	CC[C@@H](C)C(=O)O[C@@H]1C[C@@]2(C)CC[C@@](O)(O2)\C(CO)=C/[C@H]2OC(=O)C(=C)[C@H]12
14350	CC[C@@H](C)C(=O)O[C@@H]1C[C@@]2(C)CC[C@@](O)(O2)\C(COC)=C/[C@H]2OC(=O)C(=C)[C@H]12
14351	C\C=C(\C)C(=O)O[C@@H]1C[C@@]2(C)O[C@@H]2C[C@H](O)\C(CO)=C/[C@H]2OC(=O)C(=C)[C@H]12
14352	CC[C@@H](C)C(=O)O[C@@H]1C[C@@]2(C)O[C@@H]2C[C@H](O)\C(CO)=C/[C@H]2OC(=O)C(=C)[C@H]12
14353	C[C@@]12C[C@H](OC(=O)C(=C)[C@@H](O)CO)[C@@H]3[C@H](OC(=O)C3=C)[C@H]1[C@H](CC[C@H]2O)C=O
14354	C[C@]12CCC[C@H](C=O)[C@@H]1[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)C2
14355	C[C@@H]1CC[C@@H](N)N[C@H]1C1=C(C\C=C/C=C\NO)C=CC=C1O
14366	C[C@@H](CCCOC(C)=O)C1=C(C)[C@H](O)[C@H]2OC(=O)C(=C)[C@H]2[C@@H]1O
14374	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]1(C)[C@H](O)CC=C2C
14376	CC1C2CC(OC(C)=O)\C(C)=C\CC(OC(C)=O)\C(C)=C\[C@H]2OC1=O
14377	CC1CC2OC(=O)C(=C)C2(C)CC2C1=CC[C@]2(C)O
14378	CC(=O)O[C@H]1C[C@@H]2[C@H](OC(=O)C2=C)\C=C(C)\CC\C=C1/C
14380	[H][C@@](C)(OC(C)=O)[C@@](C)(OC(C)=O)C(=O)O[C@H]1C[C@@]2(C)[C@H](O)CC=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
14381	[H][C@@](C)(OC(C)=O)[C@@](C)(O)C(=O)O[C@H]1C[C@@]2(C)[C@H](O)CC=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
14382	[H][C@@](C)(OC(C)=O)[C@@](C)(O)C(=O)O[C@H]1C[C@@]2(C)[C@@H](O)CC(=O)C(C)=C2[C@H]2OC(=O)C(=C)[C@H]12
14383	[H]C([H])=C1[C@@H]2CC[C@H](C)CC(=O)C[C@H](C)C[C@H]2OC1=O
14384	CC1C2CCC(C)=C3C(C2OC1=O)C(C)=C(O)C3=O
14386	[H]O[C@@](C)(C(=O)O[C@H]1C[C@@]2(C)[C@@H](CC=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12)OC(C)=O)[C@]([H])(C)OC(C)=O
14390	CC1=C(O)C(=O)C2=C(C)CCC3C(CN4CCCC4=O)C(=O)OC3C12
14391	COC(=O)C1=C/CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)[C@]2(C)O[C@@H]2C)[C@H]\1OC(C)=O
14392	COC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)[C@]2(C)O[C@@H]2C)[C@H]\1OC(C)=O
14393	COC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C/C)[C@H]\1OC(C)=O
14394	COC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\1)OC(=O)C(\C)=C/C
14395	[H]C(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C/C)[C@H]\1O
14396	COC(=O)C1=C/CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)C(\C)=C/C)[C@H]\1OC(C)=O
14397	C\C1=C/CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2C[C@@H]1O
14398	OCC(=C)C(=O)O[C@H]1CC(=C)C2C[C@H](O)[C@](O)(CO)C2[C@H]2OC(=O)C(=C)C12
14401	CCOCC1=C2[C@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O
14402	CC(=O)OCC1=C2[C@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O
14403	CC(=O)OCC1=C2[C@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
14404	CCOCC1=C2[C@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OC(=O)C(C)=C
14405	CC(=O)OCC1=C2[C@@H](O)C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O
14406	CCOCC1=C2[C@@H](O)C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O
14407	C[C@]12CC[C@](O)(O1)[C@](C)(O)C[C@H](O)C1=C(CO)C(=O)O\C1=C\2
14408	[H][C@@]12C[C@@H](C)[C@]3([H])[C@H](C[C@@H](O)[C@@]3(C)C[C@]1([H])C(=C)C(=O)O2)OC(C)=O
14409	[H]C([H])=C1C(=O)O[C@@]2([H])C=C(COC(C)=O)[C@@]3([H])CC[C@](C)(O)[C@]3([H])C[C@]12[H]
14410	[H][C@@]12[C@@H]3OC(=O)C(=C)[C@H]3CC[C@@](C)(O)[C@]1(O)[C@@H]1O[C@@H]1[C@]2(C)Cl
14411	[H][C@@]1(C)C(=O)O[C@]2([H])C3=C(C)C(=O)C[C@]3([H])[C@@](C)(C[C@@H](O)[C@@]12[H])OC(C)=O
14412	CC(=O)OCC1=C2C(C[C@@]3(C)O[C@]33CC[C@@](C)(O3)\C=C2\OC1=O)OC(C)=O
14413	COCC1=C2[C@H](C[C@@](C)(OC(C)=O)[C@@]3(O)CC[C@@](C)(O)[C@H](OC(=O)C(C)=C)[C@@]23OC1=O)OC(C)=O
14414	[H][C@@](C)(OC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\1)OC(=O)C(\C)=C/C)[C@@](C)(O)C(=O)O[C@H]1[C@@H]2[C@H](OC(=O)C2=C)\C=C(C)CC\C=C(/C1OC(C)=O)C(=O)OC
14415	[H][C@@](C)(OC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@@H](C\1)OC(=O)C(\C)=C/C)[C@@](C)(O)C(=O)O[C@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2O[C@]2(C)CC\C=C(/C1OC(C)=O)C(=O)OC
14416	COC(=O)C1=C/CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)[C@@]2(C)O[C@H]2C)[C@H]\1OC(C)=O
14417	COC(=O)C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](OC(=O)[C@@]2(C)O[C@H]2C)[C@H]\1OC(C)=O
14418	[H][C@]12[C@@H](OC(C)=O)[C@@]3(C[C@@]11[C@@H]4[C@@H](CC(C)=C([C@@H](C)CCCOC(C)=O)[C@H]4O)OC1=O)[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1CC3=C2C
14419	[H][C@]12[C@@H](OC(C)=O)[C@@]3(CC[C@@H]4[C@H](O)C\C(C)=C\CC\C(C)=C\[C@H]4OC1=O)[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1CC3=C2C
14420	[H][C@]12[C@@H](OC(C)=O)[C@@]3(CC[C@H]4C=C(C(C)CCCO)C(=C)C[C@H]4OC1=O)[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1CC3=C2C

ID	SMILES
14421	[H][C@]1([C@@H](OC(C)=O)[C@@]2(CCC3=CC=C(C)C(=C3)C(C)CCCO)[C@@H](C)C[C@@H]3OC(=O)C(=C)[C@H]3CC2=C1C)C(O)=O
14422	C[C@@H](CCCO)C1=C(C)C[C@H]2OC(=O)C(=C)[C@H]2[C@@H]1O
14424	C[C@@H](CCCOC(C)=O)C1=C(C)C[C@H]2OC(=O)C(=C)[C@H]2[C@@H]1O
14425	C\C1=C/CC\C(C)=C\[C@H]2OC(=O)C(=C)[C@@H]2[C@H](O)C1
14426	CC1C2C(OC1=O)C1C(CO)=CC(=O)C1=C(C)CC2OS(O)(=O)=O
14427	CC1C2C(OC1=O)C1C(CO[C@H]3O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)=CC(=O)C1=C(C)CC2O
14428	[H]CC1=CC(=O)C2=C(C)CC(OS(O)(=O)=O)C3C(C)C(=O)OC3C12
14429	CC1=C2C(C3OC(=O)C(=C)C3C(C1)OC(=O)CC1=CC=C(O)C=C1)C(CO)=CC2=O
14430	[H]C1CC(C)=C2C(C3OC(=O)C(=C)C13)C(CO[C@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)=CC2=O
14431	[H]C1CC(C)=C2C(C3OC(=O)C(=C)C13)C(COS(O)(=O)=O)=CC2=O
14434	[H]C(=O)C1=C2C=CC(C)=C3CCC(=C)C3=C2OC1=O
14435	COC(=O)C1=C2C=CC(C)=C3CCC(=C)C3=C2OC1=O
14436	[H]C(=O)C1=C2C=CC(C)=C3CCC(C)C3=C2OC1=O
14437	COC(=O)C1=C2C=CC(C)=C3CCC(C)C3=C2OC1=O
14439	[H]C(=O)C1=C/CC(O[C@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)\C(C)=C\C2OC(=O)C(C)C2CC\1
14440	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]1(C)[C@@H](CC[C@@]2(C)O)OC(C)=O
14441	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]1(C)[C@H](O)CC[C@]2(C)O
14442	[H][C@]1(O)CCC(=C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]12C
14443	[H][C@@]1(CCC(=C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]12C)OC(C)=O
14444	[H][C@@]1(CC=C(C)[C@]2([H])[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]12C)OC(C)=O
14445	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@@H](OC(C)=O)[C@H]1O[C@@]21C
14446	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@@H](OCC(C)C)[C@H]1O[C@@]21C
14447	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1C(=O)C(Cl)=C2C
14448	[H][C@]12O[C@H](CC(=O)C(C)=O)[C@](C)(C[C@H]3OC(=O)C(=C)[C@@H]3[C@@H]1OC(=O)CC(C)C)O2
14450	[H][C@@]12O[C@@H](CC(=O)C(C)=O)[C@@](C)(C[C@H]3OC(=O)C(=C)[C@@H]3[C@@H]1OC(=O)CC(C)C)O2
14451	CC(C)CC(=O)O[C@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H](O)[C@@H](C)C2=CC(=O)[C@](C)(O2)[C@@H]1O
14452	CC(C)CC(=O)O[C@@H]1C[C@@]2(C)O[C@@H]2C[C@@H](O)\C(C)=C/[C@H]2OC(=O)C(=C)[C@H]12
14453	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](C[C@@]1(C)[C@H](O)C[C@H](OC(=O)OC)C2=C)OC(=O)CC(C)C
14457	CC(C)CC(=O)O[C@H]1[C@@H]2[C@H](OC(=O)C2=C)[C@@H](O)[C@@H](C)C2=CC(=O)[C@@](C)(O2)[C@@H]1O
14458	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](CC(C)=C1C(=O)C=C2CO)OC(=O)C(C)=C
14459	[H][C@@]12[C@H]3OC(=O)\C(=C/C)[C@@H]3[C@@H](CC(C)=C1C(=O)C=C2CO)OC(=O)C(C)=C
14460	[H]C1C[C@@H]2[C@H](OC(=O)\C2=C/C)[C@@]2([H])C(=C)[C@H](C[C@@]2([H])C1=C)O[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O
14461	[H]C1CC(C)=C2C(=O)C=C(CO)[C@]2([H])[C@H]2OC(=O)\C(=C/C)[C@H]12
14462	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(C)=C1C(=O)C=C2CO)OC(C)=O
14463	[H][C@@]12[C@H]3OC(=O)\C(=C/C)[C@@H]3[C@H](CC(C)=C1C(=O)C=C2CO)OC(C)=O
14465	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(C)=C1C(=O)C=C2CO)OC(=O)CC1=CC=C(O)C=C1
14466	[H][C@@]12[C@H]3OC(=O)\C(=C/C)[C@@H]3[C@H](CC(C)=C1C(=O)C=C2CO)OC(=O)CC1=CC=C(O)C=C1
14467	[H]C1CC(C)=C2C(=O)C=C(CO[C@@H]3O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)[C@]2([H])[C@H]2OC(=O)\C(=C/C)[C@H]12
14468	[H][C@@]12[C@H]3OC(=O)\C(=C/C)[C@@H]3[C@@H](O)CC(C)=C1C(=O)C=C2CO[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O
14469	[H][C@@]12[C@H]3OC(=O)\C(=C/C)[C@@H]3[C@H](CC(C)=C1C(=O)C=C2CO[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)OC(=O)C(C)=C
14470	C[C@H]1[C@@H]2CC\C(C=O)=C/C[C@H](O[C@@H]3O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)\C(C)=C\[C@H]2OC1=O
14471	C[C@H]1[C@@H]2CC\C(C)=C\C[C@H](O[C@@H]3O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)\C(C)=C\[C@H]2OC1=O
14472	C[C@H]1[C@@H]2CC\C(CO)=C\C[C@H](O[C@@H]3O[C@H](CO)[C@@H](O)[C@H](O)[C@H]3O)\C(C)=C\[C@H]2OC1=O
14473	[H][C@@]12C[C@H](OC3OC(CO)C(O)C(O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)C(O)C(C)(O)CC
14474	[H][C@@]12C[C@H](OC3OC(CO)C(O)C(O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](O)CC2=C
14475	[H][C@@]12CC(O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)=C(C)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1[C@H](O)CC2=C
14477	[H][C@@]12CC(O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)=C(C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)CC1=CC=C(O)C=C1
14479	[H][C@@]12C[C@H](O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1CCC2=C
14480	[H][C@@]12C[C@H](O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1C[C@@H](O)C2=C
14481	[H][C@@]12C[C@H](O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@H](O)CC2=C
14482	[H][C@@]12C[C@H](O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)[C@@H](C)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1[C@H](O)CC2=C
14483	[H][C@@]12C[C@H](O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)CC1=CC=C(O)C=C1
14484	[H][C@@]12CC(O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)CC1=CC=C(O)C=C1
14485	[H][C@@]12CC(O[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)C(=C)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](CC2=C)OC(=O)C(O)C(C)C
14486	C[C@H]1[C@@H]2CC[C@@]3(C)[C@H](O)CCC(C)=C3[C@H]2OC1=O
14487	[H][C@@]12[C@H]3OC(=O)[C@@H](C)[C@@H]3CC[C@@]1(C)[C@@H](CC=C2C)O[C@@H]1OC(CO)[C@@H](O)[C@H](O)C1O
14488	C\C1=C/CC\C(CO[C@@H]2OC(CO)[C@@H](O)[C@H](O)C2O)=C\[C@H]2OC(=O)C(=C)[C@@H]2CC1
14489	C[C@H]1[C@@H]2CC\C(C)=C\CC\C(CO[C@@H]3OC(CO)[C@@H](O)[C@H](O)C3O)=C\[C@H]2OC1=O
14490	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1C(=O)C(O)=C2C
14491	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1C(=O)C(OC(C)=O)=C2C
14492	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(=C)[C@]1([H])C(=O)C=C2C
14493	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1C(=O)[C@H]1O[C@@]21C
14494	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC=C(C)[C@]1(O)CC=C2C
14495	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(=C)[C@]1(O)CC=C2C
14496	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@@H](O)[C@H]1O[C@@]21C
14497	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@H](O)[C@@H]1O[C@]21C
14498	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@H](OC(C)=O)[C@@H]1O[C@]21C
14499	[H][C@@]12[C@H]3OC(=O)[C@@H](C)[C@@H]3CCC(C)=C1[C@H](O)[C@@H]1O[C@]21C
14500	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@]3(C)O[C@@]13C(=O)C=C2C
14501	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@]3(C)O[C@@]13[C@@H](O)C=C2C
14503	[H][C@@]12[C@H]3OC(=O)[C@@H](C)[C@@H]3CC[C@@]3(C)O[C@]13[C@@H](O)C=C2C
14504	[H][C@@]12CC=C(C)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1CCC2=C
14505	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]3(C)O[C@]13[C@@H](O)C=C2C
14506	[H]C1C=C(C)[C@]2([H])[C@H]3OC(=O)[C@@H](C)[C@@H]3CC[C@@]3(C)O[C@]123
14507	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@@H](OC(=O)C(C)=C)[C@H]1O[C@@]21C
14508	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CCC(C)=C1[C@@H](OC(=O)C(\C)=C\C)[C@H]1O[C@@]21C
14509	CC1C=C(C)C2[C@H]3OC(=O)[C@@H](C)[C@@H]3CC[C@]3(C)O[C@@]123

ID	SMILES
14510	C[C@H]1[C@@H]2CC[C@]3(C)O[C@]33[C@@H](O)C=C(C)C3[C@H]2OC1=O
14511	C[C@H]1[C@@H]2CC[C@]3(C)O[C@]33[C@H](O)C=C(C)C3[C@H]2OC1=O
14514	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(=C)[C@]1([H])[C@@H](O)C=C2C)OC(C)=O
14515	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](C[C@](C)(O)[C@]1([H])C(=O)C=C2C)OC(C)=O
14516	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)CC(C)C)C=C(C)[C@]1(O)C[C@@H](Cl)[C@]2(C)O
14517	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(C)CC)C=C(C)[C@]1(O)C[C@@H](Cl)[C@]2(C)O
14518	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@@H](OC(=O)C(\C)=C/C)C=C(C)[C@]1(O)C[C@@H](Cl)[C@]2(C)O
14519	[H][C@]12[C@@H](C)C(=O)O[C@@H]1[C@]1([H])C(C[C@H](O)[C@]1(C)O)=C(C)C[C@@H]2OC(C)=O
14520	[H][C@]12[C@H](C)C(=O)O[C@@H]1[C@]1([H])C(C[C@@H](O)[C@]1(C)O)=C(C)C[C@@H]2OC(C)=O
14521	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(C)=C1C[C@@H](O)[C@]2(C)O)OC(C)=O
14522	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(C)=C1C[C@H](O)[C@@]2(C)O)OC(=O)C(C)C
14523	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3[C@H](CC(C)=C1C[C@H](O)[C@@]2(C)O)OC(=O)C(\C)=C/C
14524	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@]1(C)[C@H](O)C[C@H](OC(=O)CC(C)C)C2=C
14526	CC(=O)OC\C1=C/CC\C(CO)=C\C2OC(=O)C(=C)C2[C@H](C1)OC(C)=O
14527	C[C@@H]1[C@@H]2CCC(=C)C3C[C@H](O)C(=C)C3[C@H]2OC1=O
14528	CO[C@@]12CC[C@@](C)(O1)\C=C1\OC(=O)C(COC(C)=O)=C1[C@H](C[C@H]2C)OC(=O)C(C)=C
14529	C[C@@H]1C[C@H](OC(=O)C(C)=C)C2=C(COC(C)=O)C(=O)O\C2=C\[C@@]2(C)CC[C@]1(O)O2
14530	[H]C([H])=C1C(=O)O[C@@]2([H])C[C@H](C)[C@]3(O)C([H])([H])[C@@]4([H])O[C@@]4(C)[C@]3([H])C[C@]12[H]
14531	[H]C([H])=C1C(=O)O[C@@]2([H])C[C@H](C)[C@]3([H])CC[C@@](C)(O)[C@]3([H])C[C@]12[H]
14532	[H][C@@]12[C@H]3OC(=O)C(=C)[C@@H]3CC[C@@](C)(O)[C@@]11O[C@H]1[C@H](Cl)[C@@]2(C)O
14533	CO[C@H]1CC(=O)C([C@H]2OC(=O)C(=C)[C@@H]2CCC(C)=O)=C1C
14534	[H][C@@]12CC(=O)C(C)=C1[C@H](O)[C@H]1[C@@H](C[C@@H]2C)OC(=O)C1=C
14535	C[C@@H]1[C@H]2C\C=C(CO)\CC\C=C(CO)/C[C@@H]2OC1=O
14536	[H][C@@]12OC(=O)C(CC[C@@H]3O[C@]3(C)C(=O)\C=C/C1(C)C)=C2
14537	[H][C@@]12OC(=O)[C@@]3(CC\C=C(C)/C(=O)/C=C/C1(C)C)O[C@@H]23
14538	[H][C@@]12OC(=O)C(CC[C@H](O)[C@H](C)C(=O)\C=C/C1(C)C)=C2
14539	[H][C@@]12OC(=O)C(CC[C@H](OCC)[C@H](C)C(=O)\C=C/C1(C)C)=C2
14540	[H][C@@]12C\C=C/[C@@](C)(O)C(=O)[C@@]3([H])CC(C)(C)[C@@]([H])(OC1=O)[C@@]23[H]
14541	[H]C1[C@]2([H])[C@@]3([H])[C@](C)([H])C(C)=C4[C@@]5([H])OC(=O)[C@]([H])(C)[C@]5([H])CC[C@](C)(O)[C@]34[H])C34[C@H]5OC(=O)[C@@]([H])(C)[C@]5([H])CC[C@](C)(O[C@@]13C)[C@]24[H]
14542	[H]C1[C@]2([H])[C@@]3([H])[C@](C)([H])C(C)=C4[C@@]5([H])OC(=O)[C@@]([H])(C)[C@]5([H])CC[C@](C)(O)[C@]34[H])C34[C@H]5OC(=O)[C@]([H])(C)[C@]5([H])CC[C@](C)(O[C@@]13C)[C@]24[H]
14543	[H][C@@]12CC[C@](C)(O)[C@@]3([H])C([C@H]1OC(=O)[C@H]2C)=C(C)[C@@]1([H])[C@]3([H])[C@@]2([H])C(=O)[C@@]([H])(C)C11[C@H]3OC(=O)[C@@H](C)[C@]3([H])CC[C@](C)(O)[C@]21[H]
14544	[H][C@@]1(C)C(=O)O[C@H]2[C@@]1([H])CC[C@]1(C)O[C@@]3(C)[C@@H](O)[C@]4([H])[C@@]5([H])[C@]([H])(C)C(=C6[C@@]7([H])OC(=O)[C@@]([H])(C)[C@]7([H])CC[C@](C)(O)[C@]56[H])C23[C@]14[H]
14545	[H][C@@]1(C)C(=O)O[C@H]2[C@@]1([H])CC[C@]1(C)O[C@@]3(C)[C@@H](O)[C@]4([H])[C@@]5([H])[C@]([H])(C)C(=C6[C@@]7([H])OC(=O)[C@]([H])(C)[C@]7([H])CC[C@@](C)(O)[C@]56[H])C23[C@]14[H]
14546	CC(C)C(=O)OCCC[C@H](C)C1=C(C)C[C@H]2OC(=O)C(=C)[C@H]2[C@@H]1OC(=O)C(C)C
14547	CCC(C)CC(=O)O[C@H]1[C@H]2[C@@H](CC(C)=C1[C@@H](C)CCCOC(C)=O)OC(=O)C2=C
14548	CC(C)C(=O)O[C@@H]1[C@@H]2[C@H](CC(C)=C1[C@H](C)CCCOC(C)=O)OC(=O)[C@]21C[C@]23C(OC(C)=O)[C@H]1C(C)=C2C[C@@H]1[C@@H](C[C@H]3C)OC(=O)C1=C
14549	CCC(C)CC(=O)O[C@@H]1[C@@H]2[C@H](CC(C)=C1[C@H](C)CCOC(C)=O)OC(=O)[C@]21C[C@]23C(OC(C)=O)[C@H]1C(C)=C2C[C@@H]1[C@@H](C[C@H]3C)OC(=O)C1=C
14550	C\C=C(\C)C(=O)O[C@H]1C[C@@H](C)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(CO)=C12
14551	C\C=C(\C)C(=O)O[C@H]1C[C@@H](C)[C@]2(O)CC[C@@](C)(O2)\C=C2\OC(=O)C(COC(C)=O)=C12
14552	C[C@@H]1C[C@H](OC(=O)C(\C)=C\CO)C2=C(CO)C(=O)O\C2=C\[C@@]2(C)CC[C@]1(O)O2
14553	C\C=C(/C)C(=O)OCC1=C2[C@@H](O)C[C@@H](C)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O
14554	[H][C@@]12CC[C@@](C)(O)[C@]1([H])C1OC(=O)C(=C)[C@@H]1CCC2=C
14555	[H][C@@]12[C@H]3OC(=O)[C@H](C)[C@@H]3C[C@@H](OC(C)=O)C(C)=C1C=C[C@@]2(C)O
14556	[H][C@@]12[C@H]3OC(=O)[C@H](C)[C@@H]3C[C@H](OC(C)=O)C(C)=C1C=C[C@@]2(C)O
14557	[H][C@@]12[C@H]3OC(=O)[C@H](C)[C@@H]3C[C@@H](OC(C)=O)C(=C)[C@]1(O)C=C[C@@]2(C)O
14558	[H][C@@]12[C@@H](O)[C@H]3[C@@H](C[C@@]1(C)[C@H](O)CC=C2C)OC(=O)C3=C
14559	[H][C@]12C(C)CC[C@@H](O)[C@]1(C)C[C@H]1OC(=O)C(=C)[C@H]1[C@@H]2O
14560	[H][C@@]12CC[C@H](CO)[C@]1([H])[C@H]1OC(=O)C(=C)[C@@H]1[C@@H](O)CC2=C
14561	[H][C@@]12CC[C@H](CO)[C@]1([H])[C@H]1OC(=O)[C@@H](C)[C@@H]1[C@@H](O)CC2=C
14562	[H][C@@]12[C@H]3OC(=O)C(C)C3[C@@H](O)C[C@@]1(COC(=O)[C@H]2C)C=C
14563	[H][C@]12C[C@@](C)(C3=C1[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1C3)[C@]1(C2)[C@H]2C[C@@]3([H])C(=C)C[C@H](O)C[C@]3(C)C[C@H]2OC1=O
14564	[H][C@]12CC3([C@H]4C[C@@]5([H])C(=C)C[C@H](O)C[C@]5(C)C[C@H]4OC3=O)[C@]3(C1)[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1CC3=C2C
14565	[H][C@]12C[C@@](C)(C3=C1[C@@H](C)C[C@@H]1OC(=O)C(=C)[C@H]1C3)[C@]1(C2)[C@H]2C=C3[C@@H](C)CCC[C@]3(C)C[C@H]2OC1=O
15202	[H][C@@](C)(OC(C)=O)[C@@](C)(O)C(=O)O[C@H]1C[C@@]2(C)[C@@H](CC=C(C)[C@]2([H])[C@H]2OC(O)C(=C)[C@H]12)OC(C)=O
15203	[H][C@@](C)(OC(C)=O)[C@@](C)(OC(C)=O)C(=O)O[C@H]1C[C@@]2(C)[C@H](O)CC=C(C)[C@]2([H])[C@H]2OC(O)C(=C)[C@H]12
15204	[H]O[C@@H]1CC=C(C)[C@]2([H])[C@H]3OC(O)C(=C)[C@@H]3[C@H](C[C@@]12C)OC(=O)[C@](C)(O[H])[C@]([H])(C)OC(C)=O
15205	C\C1=C\CC[C@@]2(C)O[C@H]2[C@H]2OC(=O)C(=C)[C@@H]2CC1
15206	[H][C@@]12C[C@H](O)C(=C)[C@]1([H])[C@H]1OC(=O)[C@@](O)(CO)[C@@H]1CCC2=C
15207	OCC(=C)C(=O)O[C@H]1CC(=C)C2C[C@H](O)[C@](O)(CO)C2[C@H]2OC(=O)C(=C)[C@H]12
15208	[H]C1(C)OC1(C)C(=O)O[C@@H]1CC(C)=C2C(=O)C=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
15209	[H]C1(C)OC1(C)C(=O)O[C@@H]1C[C@](C)(O)[C@]2([H])C(=O)C=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
15210	[H]C(C)(O)C(C)(O)C(=O)O[C@@H]1CC(C)=C2C(=O)C=C(C)[C@]2([H])[C@H]2OC(=O)C(=C)[C@H]12
15211	[H][C@@]12[C@H]3OC(=O)C(C)=C3C=CC(C)=C1C(=O)C(O)=C2C
15266	CCOCC1=C2[C@H](C[C@@](C)(O)[C@]3(O)CC[C@@](C)(O3)\C=C2\OC1=O)OCOC
15275	[H][C@@]12[C@H]3OC(=O)C(=C)C3C(OC(C)=O)C=C(C)[C@]1(O)CC=C2C
15276	[H][C@@]12[C@H]3OC(=O)C(=C)C3C(CC(=C)[C@]1(O)CC=C2C)OC(C)=O
15278	[H][C@@]12[C@H]3OC(=O)C(=C)C3C(O)CC(=C)C1(O)C[C@H]1O[C@@]21C
15280	[H][C@@]12[C@H]3OC(=O)C(=C)C3C(O)C=C(C)C1(O)C[C@H]1O[C@@]21C
15338	[H]C(CC(C)=C)C1[C@H](OC(=O)C1=C)C1=C(C)[C@@H](O)CC1C
15339	[H]C1CC(C)C([C@H]2OC(=O)C(=C)C2[C@@H](O)CC(C)=C)=C1C
15513	[H][C@@]12[C@H]3OC(=O)C(=C)C3CC[C@@]1(C)[C@H](O)C[C@H](O)C2=C
15514	[H][C@@]12[C@H]3OC(=O)C(=C)C3[C@@H](O)CC(C)=C1C(=O)C=C2C
15515	[H][C@@]12[C@H]3OC(=O)[C@H](C)C3[C@@H](O)CC(C)=C1C(=O)C=C2C
15516	[H]C1[C@@H](OC(=O)C(\C)=C\C)[C@@H]2[C@H](OC(=O)C2=O)\C=C(C)\CC\C=C1\C(=O)OC
15517	[H]C1[C@@H](OC(=O)C(C)=C)[C@@H]2[C@H](OC(=O)C2=O)\C=C(C)\CC\C=C1\C(=O)OC
15518	[H]C1[C@@H](OC(=O)C(\C)=C/C)[C@@H]2[C@H](OC(=O)C2=O)\C=C(C)\CC\C=C1\C(=O)OC

ANEXOS II

Produção científica

1. **Artigo: Hybrid Compounds in the Search for Alternative Chemotherapies against Neglected Tropical Diseases**

   Artigo aceito na revista ***Letters in Organic Chemistry***
   ISSN: 1875-6255
   Fator de Impacto: 0.73

2. **Artigo: Ligand-based virtual screening of a benzylisoquinoline alkaloids dataset with potential anti-inflammatory activity.**

   Artigo será submetido na revista ***Molecules***
   ISSN: 1420-3049
   Fator de Impacto: 2.861

3. **Cell-based platforms and in silico methodologies for discovery of new drugs**
   Artigo será submetido na revista ***Expert opinion on Drug Discovery***
   ISSN: 1746-0441
   Fator de Impacto: 3.876

4. **Ligand-based virtual screening of a benzylisoquinoline alkaloids dataset with anti-inflammatory potential activity.**

   Resumo expandido aceito e publicado na WRSAMC01: Workshop in Research Sciences Applied to Medicinal Chemistry, Paraíba, Brasil, 2016
   DOI:10.3390/mol2net-02-03859

5. **In silico Studies to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an Asteraceae in-house Database**

   Certificate of poster presentation in the 6th BCNP Brazilian Conference on Natural Products/XXXII RESEM, Nov. 5th-8th, in Vitoria, ES.

6. **Triagem virtual baseada nas estruturas dos ligantes e receptores para selecionar compostos de um banco de dados de lactonas sesquiterpénicas de Asteraceae com potencial atividade leishmanicida.**

   Resumo do trabalho (pôster) apresentado no 5° encontro de Química da UFPB, Formação dos professionais de Química: Novos Desafios para Novos Tempos, 25 a 28 de setembro de 2017

Hybrid Compounds in the Search for Alternative Chemotherapies against Neglected Tropical Diseases.

Chonny Herrera Acevedo[a], Luciana Scotti*[a], Mateus F. Alves[a], Margareth de F. F. M. Diniz[a], Marcus Tullius Scotti[a]

[a]*Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil*

Received: ***Revised:*** ***Accepted***

Abstract: Neglected tropical diseases (NTDs] affect more than a billion people worldwide, mainly populations living in poverty conditions. More than 56% of annual NTD deaths are caused by Leishmaniasis, Sleeping sickness, and Chagas disease. For these three diseases, many problems have been observed with the chemotherapies commonly used principally: resistance, high toxicity, and low efficacy. In the search for alternatives treatments, approximative hybridization is an interesting approach which generates new molecules by merging two pharmacophores and then looking for improvements in biological activity or reduced compound toxicity. Here, we review various studies which present such hybrid molecules with promising *in vitro* and *in vivo* activities against *Leishmania* and *Trypanosoma* parasites.

Keywords: Hybrid compounds, Leishmaniasis, Chagas disease, Sleeping sickness, Neglected tropical diseases, antiprotozoal activity.

1. INTRODUCTION

Leishmaniasis, Chagas disease and Sleeping sickness are the three principal Neglected Tropical Diseases (NTDs] caused by differing kinetoplastic parasite species: [1] Leishmaniasis is caused by more than 20 protozoan parasite species of the genus *Leishmania*; [2] Trypanosomiasis is caused by the protozoan *Trypanosoma*, in two classes which affect humans, Chagas Disease [American trypanosomiasis] is generated by the etiologic agent *T. cruzi*; [3] Sleeping sickness (Human African Trypanosomiasis- HAT] is caused by *T. brucei*.

Multiple problems with current chemotherapies against these three diseases have been observed. Antimonial compounds were the first and most common treatment against *Leishmania*, however, despite their effectiveness, high toxicity was often observed [4]. Other more recent chemotherapies such as pentamidine, miltefosine and paromomycin retain many adverse effects including nephrotoxicity [5], increased resistance [6-8], teratogenicity and even abortifacient effects [9]. Amphotericin B, although more effective than previous drugs, possesses many side effects such as vomiting, fever and toxicity, which add to its high cost, and difficult availability [10].

*Address correspondence to this author at the Post-Graduate Program in Natural and Synthetic Bioactive at Federal University of Paraíba, 58051-900 João Pessoa, Paraíba, Brazil; Tel.: +55-83-99869-0415; E-mail: Luciana.scotti@gmail.com

Serious difficulties have also been observed with therapeutic products against trypanosomiasis. For Chagas disease, the effectiveness of nifurtimox and benznidazole is close to 100%, but risks of infection reactivation, nausea, abdominal pain, and vomiting, chronic headaches, and fatigue etc., still exist [11, 12].

Pentamidine and suramin are used respectively in the first phase of HAT for treatment against *T. brucei gambiense* and *T. brucei rhodiense*. Yet many side effects are observed such as neuropathy, renal problems, among others [13, 14]. For the second phase, melarsoprol, an arsenic-derived compound, is the only drug available for treatment, despite serious risks involved in its use, including high toxicity and side effects such as diarrhea, fever, and frequent cases of severe encephalopathy [15].

Given that 56% of the deaths generated annually by NTDs are caused by these three parasitic diseases [16], the major challenge for their control and elimination is development and distribution of new and effective therapeutic products with reduced side effects [17]. An interesting approach; hybrid compounds based on combining two covalently linked pharmacophore groups and generating new molecules with greater pharmacological potency than the individual moieties has recently emerged [18].

The advantages of the hybrid approach include optimized biological properties like moiety affinity and selectivity [19], the possibility of designing multitarget compounds with biological activities contributed from each pharmacophore, potentially lower production costs, and lower risk of adverse drug–drug interactions as compared to multicomponent

XXX-XXX/15 $58.00+.00

therapies [20, 21]. Biological activities for such compounds have been studied and from the results, certain properties can be attributed such as: anticancer [22, 23], antimalarial [24, 25], antifungal [26], and anti-Alzheimer [27], among others.

Inspired by the results of previous studies which indicate anti-parasitic activity for diverse structures such as chalcones, quinazonilones, β-carbolines, benzofurans, etc., several research groups have designed new hybrid molecules covalently linking fragments of these compounds, or using anti-parasitic moieties of compounds currently employed in treatments against these diseases (i.e. pentamidine]. In this review, promising *in vitro* and *in vivo* activity results are presented from alternative drug research against Leishmaniasis, Chagas disease and Sleeping sickness, all expressed as percentage inhibition, or in IC_{50} values (half maximal inhibitory concentration] of the synthesized hybrid molecules.

2. LEISHMANIA

Several studies have revealed molecular hybrids with activity against *Leishmania*; alternatives that may well overcome multiple problems with current leishmaniasis chemotherapy (Figure **1**].

Initial examples of this type are two thiosemicarbazones of 3-carboxi-β-carbolines, **1.1** and **1.2**, with *in vitro* activity against *L. donovani* promastigotes, with respective IC_{50} values of 5.0 and 2.5 μM. Pentamidine IC_{50} 0.56μM was used as the control. The mechanism of action for these two hybrids were evaluated such that **1.1** acts inhibiting ribonucleoside diphosphate reductase and therefore DNA biosynthesis; **1.2** inhibits RNA and protein synthesis [28].

Dicationic compounds have also exhibited antileishmanial activity [29]. Pentamidine was used initially to treat visceral leishmaniasis and is currently used in Africa for cutaneous leishmaniasis HIV–co-infections, however many adverse effects exist [3, 30]. Various hybrid compounds have been designed using pentamidine because of its biological activity. Porwal et al, proposed a pentamidine-aplysinopsin hybrid that substitutes the amidinophenoxy function in the pentamidine structure with aplysinopsin, a natural tryptophan-derived marine product that acts on similar biological targets as does the dicationic class of molecules [31]. During synthesis, in preliminary antileishmanial screening with the nine compounds formed, it was observed that only compound **1.3**, with Z-geometry presented both promastigote inhibition (95%], and *in vivo* intracellular *L. donovani* amastigote inhibition (62%]; after treatment, no adverse effects were observed in the animals [32].

Based on the promising results observed for the lead molecule, eight compounds were synthesized by varying: the thiocarbonyl function, the cyano- and phenolic moieties, the chain length and the geometry; however no compound tested presented such promising activity as did compound **1.3**. The cyano- group of **1.3** was then modified attempting to increase its leishmanicidal activity. From the eight compounds obtained, compound **1.4** an E-geometrical isomer containing the 2-thio analogue of aplysinopsin presented 99.3% of inhibition for intracellular amastigotes with an IC_{50} value of 2.0 μM; being 10 times more effective than the control (pentamidine isethionate salt], and less toxic than pentamidine for the KB cell line, a molecule with promise for treating *L. donovani* [32].

A series of hybrid compounds based on pentamidine was also designed, using skeletons of two natural products with previously reported anti-leishmanicidal activity: annonmontine and licochalcone A. Annomontine, is a pyrimidine-beta-carboline alkaloid isolated from *Annona foetida*, that presents an IC_{50} value of 34.8 μM against promastigote forms of *L. braziliensis* [33]. Licochalcone A, is an oxygenated chalcone isolated from the roots of Chinese licorice (*Glycyrrhiza uralensis]* which at a dose of 20 mg/kg weight for six days yielded leishmanicidal activity against *L. major* and *L. donovani* promastigotes, reducing the parasite load in the liver and spleen of hamsters by more than 96%, with respect to untreated animals [34, 35].

Nine pyrimidin-pentamidine hybrids and six chalcone-pentamidine hybrids were synthesized and their *in vitro* activity against *L. donovani* amastigote forms was tested. All of the pyrimidin-pentamidine hybrids presented leishmanicidal activity, with IC_{50} values ranging from 0.30 to 1.72 μM. Compound **1.5**, with an ethyl piperazine substitution, displayed the highest selectivity index (SI: >232] and the lowest inhibition concentration (0.30μM]. The bis-chalcone-pentamidine hybrid **1.6** was the only molecule of the six tested that inhibited amastigote forms with an IC_{50} value of 1.65μM having a selectivity index of (SI: >242]. In all cases, lower concentrations than the IC_{50} values of the three leishmanicidal compounds used as controls were obtained, being: sodium stibogluconate (SSG], miltefosine and pentamidine (IC_{50}: 71.90, 12.5 and 20.43 μM respectively].

Observing results revealing two compounds with good activity against *L. amazonensis* and *L. chagasi* promastigotes from a series of 16 aromatic compounds synthesized through a Morita–Baylis–Hillman reaction [37], seven chalcone-like hybrids were obtained using the same synthesis combining the initial series with methyl-salicylate which possesses anti-inflammatory properties. In all cases, the hybrids were more active than their precursors. Activity against *L. amazonensis* ranged from IC_{50} values of 7.65 to 31.15 μM, for *L. chagasi* they varied between 10.14 and 57.34 μM. **1.7** emerged as a promising leishmanial molecule presenting the highest activity for the two strains studied, this is possibly due to the presence of a characteristic ortho-nitro group in its structure; (IC_{50} 7.65 μM for *L. amazonensis* and 10.14 μM for *L.chagasi*] [38].

Triazines, a compound group with potential antileishmanial activity which acts inhibiting leishmanial dihydrofolate reductase [39], was the basis for the synthesis of a series of [1,2,4]triazin[5,6-b]indol-3-ylthio-1,3,5-triazines, and [1,2,4]triazin[5,6-b]indol-3-ylthiopyrimidine hybrids. After antileishmanial activity screening, compound **1.8** a triazin[5,6-b]indol-3-ylthio-1,3,5-triazin derivative (with N-ethylpiperazine], appeared as the most active compound against *L. donovani* amastigotes (IC_{50} 4.01μM] with even lower values in respect to the SSG and pentamidine controls (IC_{50} 71.90 and 20.43 μM respectively]. Thus, compound **1.8** is highlighted as a promising leishmanial compound due its selectivity that presents a CC_{50} value of 227.04 μM (J774A.1 cell line] and an SI = 56.61 [40] .

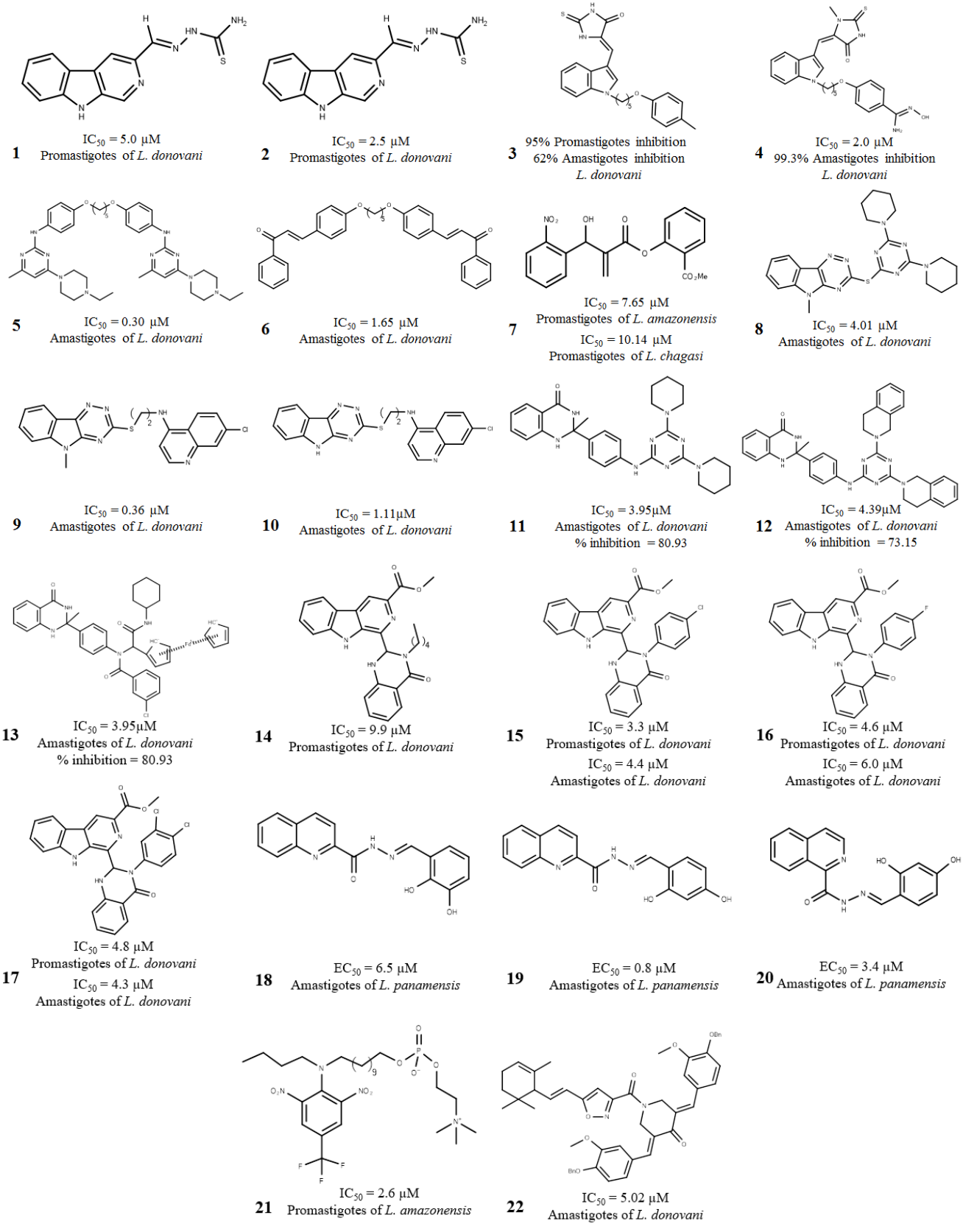


Figure 1: Active hybrids against different parasitic forms and species of *Leishmania* with reported inhibitory percentages and/or IC_{50} values

Using this type of compound, a series of 19 hybrids covalently linking 1,2,4-triazino[5,6b] indol-3-thione with 7-chloro-4-aminoquinoline, antileishmanial activity was tested for the whole series using extracellular and intracellular *L. donovani* promastigotes, presenting respective ranges for their IC_{50} values of from 1.95 to 24.14 μM, and 0.36 to 29.48 μM. A cytotoxicity assay was performed with Vero cells. From this, compound **1.9** was promising against *Leishmania*, since it showed a lower value IC_{50} of (0.36μM] with better selectivity (SI: > 1111] than miltefosine (IC_{50}= 8.10 μM, SI: 54.76], or SSG (IC_{50}= 54.6 μM, SI: > 400], the two leishmanial compounds used as controls. Lipophilic substituents are fundamental factor in the biological activity of this type of hybrid since other compounds of the series (like **1.10]**, having high structural similarity with **1.9**, also present activity against amastigotes IC_{50} =1.11μM and SI = 67 [41].

The same research group proposed certain triazine hybrids which presented activity against *Leishmania*, in a study where four series of molecules were synthesized by approximative hybridization of quinazonilone, an interesting pharmacophore which possesses a wide range of biological activities, linked with four scaffolds: pyrimidine, triazine, tetrazole, and peptide; *L. donovani* anti-amastigote and antipromastigote activity was tested for the whole series and cytotoxicity assays were performed using J774A.1 and Vero cell lines [42].

The results revealed nine compounds with high *in vitro* antileishmanial activity [an IC_{50} range for the intracellular amastigote form from: 0.73 to 7.76 μM] all displaying lower concentrations than the controls miltefosine and SSG, IC_{50} = 8.4 and 46.70 μM, yet without cytotoxicity. *In vivo* tests were performed using a golden hamster model; with inoculations by intraperitoneal route for five days, a dose of 50 mg/Kg/day for the nine most active compounds. Two quinazolinone-triazine hybrids **1.11** and **1.12** reached respective inhibitions of 80.93% (IC_{50} = 3.95 μM] and 73.15% (IC_{50} = 4.39 μM] for *L. donovani*, meanwhile **1.13** a quinazolinone-peptide molecule blocked the parasite multiplication by 51.42% (IC_{50} = 0.73 μM]. In murine macrophage cells (J774A.1], the anti-leishmanial activity mediated by **1.11** and **1.12** is associated with Th1 cytokines (IL-12 and TNF-α] activation, and down-regulation of Th2 cytokines (IL-10 and TGF-β] [42].

Another series of hybrid compounds was obtained linking the quinazolinone scaffold with β-carbolines, which in previous studies showed antileishmanial activity [43, 44]. The authors evaluated *L. donovani* trypanothione reductase (TR] (a flavoprotein oxidoreductase, which is crucial for *Leishmania* survival [45]] inhibition capacity for the fifteen hybrids synthesized. Six compounds were able to inhibit TR from 67% to 99% at 100μM, and from 56% to 90% at 50μM. However, after biological studies with *L. donovani* amastigotes and promastigotes it was observed that compound **1.14** with the lowest inhibition constant (Ki =0.8 μM], only possesses *in vitro* activity against the promastigote form (IC_{50} =9.9 μM]. Then, compounds **1.15**, **1.16,** and **1.17** appeared as promising leishmanial hybrid molecules because of activity against *L. donovani* promastigotes (IC_{50} = 3.3, 4.6, and 4.8 μM] and amastigotes (IC_{50} = 4.4, 6.0, and 4.3 μM], with high SI, using Vero cells in the cytotoxicity assay. According to the results, a clear correlation exists between activity and TR inhibition. Thus, the leishmanial activity of these three β-carboline–quinazolinone hybrids is related to their efficient competitive TR inhibition (Ki 7.2, 3.9, and 3.7 μM, respectively] [46].

Multiple biological activities have been observed for the quinoline core, [47-49], highlighting the antiparasitic properties of chloroquine, (a 4-aminoquinoline which is currently used in antimalarial treatment], and certain derivatives with antileishmanial activity [50].

Based on these properties and the considerable antimycobacterial activity of isoxyl and phenylthiourea derivatives, Nava-Suazo et al, synthesized a series of eight hybrid compounds from chloroquine, ethambutol, and isoxyl; testing their *in vitro* activity against five protozoans, among these *L. mexicana.* The hybrid compound (**1.17**], which had in its structure the 4-butoxyphenyl substituent, was the most active against the promastigote form IC_{50} = 9.90 μM, even more active than pentamidine, the leishmanial compound used for control: IC_{50} = 13.32 μM [51].

In research of new biologically active hybrid molecules with the quinolone core [a compound which presents antileishmanial activity against several species of *Leishmania*]; ten synthesized compounds (containing this pharmacophore] were covalently linked to hydrazone, [52-54]. From *in vitro* tests against the intracellular amastigote form of *L. panamensis* for the whole series, **1.18**, **1.19** and **1.20** appeared as interesting scaffolds, respectively presenting high inhibition percentages: 75.7%, 64.7% and 62.6%, and low EC_{50} values 6.5, 0.8 and 3.4 μM. Meglumine antimoniane (EC_{50}= 6.3 μM and 79.4% inhibition] and Amphotericin B (EC_{50}= 0.04 μM and 76.0% inhibition] were used as controls. The activity of these three quinolone-hydrazone hybrids is associated with the isoquinolinic core, and a hydroxyl group in positions 3 and 4 of the aromatic ring. However, the high cytotoxicity levels observed imply further modifications of such templates, with the aim of finding future leishmanicidal drugs [55].

A promising hybrid TC95 molecule, **1.21**, was synthesized by linking trifuralin, (a herbicide which presents *in vivo* activity against *L. mexicana* and *L. major;* reducing the size of cutaneous lesions on BALB/c mice as a topical ointment [56, 57]], and miltefosine, (an alkyl phospholipid compound yielding high efficacy against promastigotes and amastigotes of several species of *Leishmania*, and is currently used as chemotherapy for visceral and cutaneous leishmaniasis in spite of its side effects [9, 58]]. The results revealed the high *in vitro* activity of **1.21** against promastigotes and intracellular amastigotes of *L. amazonensis* with respective IC_{50} values of 2.6 and 1.2 μM, being in both cases more active than miltefosine (IC_{50} = 25 μM] with an SI = 24, thus suggesting lower toxicity. Using several techniques, it was possible to determine that **1.21** generates changes in the cytoskeleton, accumulation of lipid bodies, and cell cycle arrest (mainly in G1 phase], among other actions [59].

Finally, seven hybrids combining hetero-retinoids with bisbenzylidine ketones were synthesized; and their *in vitro* activities against intracellular amastigotes of *L. donovani* were tested. Five of the hetero-retinoid-bisbenzylidine hybrid compounds presented strong activity against this parasitic form, in a range of IC_{50} values from 1.83 to 6.10

μM, being more active than the antileishmanial drugs used as controls, SSG and miltefosine, which reached IC_{50} values of 53.12 and 8.10 μM. The activity of this class of molecules is related to the type of substituent present and its position on the phenyl ring. From cytotoxicity assays performed with Vero cells, it was observed that the hybrid molecule, **1.22** (SI > 79.68], which presented an IC_{50} value (5.02 μM], was ten and twelve times more selective than the controls, emerging as a promising molecule and allowing the identification of a new type of compound with potential activity against *Leishmania* [60].

3. TRYPANOSOMIASIS

Approximative hybridization has also been used in the search for alternative chemotherapies for the two types of trypanosomiasis which affect humans: Chagas disease transmitted by the parasite *Trypanosoma cruzi* (Figure **2**], and Human African Trypamosomiasis, caused by *Trypanosoma brucei* infection (Figure **3**].

3.1 Chagas disease

In view of previous results indicating that the N-oxide moiety is important to the trypanocidal activity of benzofuran derivatives; hydrazone-benzofuran hybrids were synthesized, such as vinylsulfonylbenzofuroxans **2.1** and 5-arylethenylbenzofuroxans **2.2** [61-63]. Along with cytotoxic activity measurements using mouse macrophages (J-774.1], the inhibitory capacity of thiosemicarbazones, N-hydrazones and amidinohydrazones containing nitrofurans and naphthoquinones against cruzipain (CP] and TR; (two enzymes with vital roles for *Trypanosoma*], and *in vitro* anti-trypanosomal activity of these and their intermediates using two strains of *T. cruzi* epimastigotes were determined [64].

Compound **2.3**, a thiosemicarbazone, presented good inhibitory capacity against *T. cruzi* growth (70%], and antitrypanosomal activity against *T. cruzi* strains: Tulahuen 2 (IC_{50} = 15μM] and CL Brener (IC_{50} = 19.8 μM], being selective for the parasite (SI = 26.7]. Compound **2.3** was also highlighted as a lead structure for CP (cruzipain] inhibition with an inhibitory concentration of 43 μM. The authors suggested that this hybrid molecule acts on *T. cruzi* producing oxidative stress inside the parasite, since in ESR spectroscopy using DMPO, compound **2.3** was able to produce free radicals in the presence of a *T. cruzi* microsomal fraction [64].

The same authors studied a series of benzofuroxan hybrids, evaluating their activity against two strains of *T. cruzi*. Respectively, the results revealed compounds **2.4** (*E*-configuration, IC_{50} = 1.0μM,] and **2.5** (*Z*-configuration, IC_{50} = 1.6μM] as the most active vinylthiobenzofuroxan derivatives against epimastigote forms of the Tulahuen 2 strain and CL Brener clone. However, almost all of the tested hybrids were active against *T. cruzi*. A cytotoxicity assay for certain hybrid compounds was performed using J-774 mouse macrophages, and the results showed that the chalcone-like hybrids **2.6** (SI= 9.4] and **2.7** (SI= 13.3] have low cytotoxicity and high selective indexes. Studying their mechanisms of action through enzymatic assay, compound **2.6** partially inhibited *T. cruzi* mitochondrial dehydrogenases; the capacity to produce oxidative oxygen species in the parasite was observed by means of ESR spectra [65].

Based on the proven trypanocidal action of nitroheterocyclic compounds, a new approach yielded a series of naphthoquinone derivatives which was synthesized from nor-lapachones and lapachones. Activity against trypomastigote forms of the *T. cruzi* Y strain was tested. Results indicated two interesting scaffolds for development of antitrypanosomal drugs: compound **2.8** (IC_{50} = 86.3 μM] and **2.9** (IC_{50} = 88.2 μM], which presented more activity than benznidazole (IC_{50} = 103.6 μM] the trypanocidal compound employed as control [66]. From a whole series of nor-β-lapachone triazole derivatives, naphthoquinoidal [1,2,3]-triazole, presented *in vitro* activity against *T. cruzi* trypomastigotes in a range of IC_{50} values from 17.3 to 348.1 μM, with higher anti-Chagas activity than their precursor nor-lapachol (having an IC_{50} value of 1281μM]. The hybrid molecule 2,2-dimethyl-3-(4-phenyl-[1,2,3]triazol-1-yl]-2,3-dihydro-naphtho[1,2-b]furan-4,5-dione (compound **2.10**] had the lowest IC_{50} value (17.3μM], being more active than the benznidazole control. From these results, although no cytotoxic assays were reported, such compounds emerge as a new structural antitrypanosomal moieties [67].

Derivatives of the monoterpene limonene have been developed by linking heterocyclic fragments which possess antitrypanosomal activity: benzofuroxan, benzimidazole dioxide, 5-nitrofuran, and 5-nitroindazole. From *in vitro* assays using the epimastigote form of the *T. cruzi*, Tulahuen 2 strain, the 50% inhibitory dose (ID_{50}] for all of the synthesized compounds and their precursors were determined. Four known trypanocidal compounds were used as controls (nifurtimox, benznidazole, terbinafine and ketoconazole]. Certain nitro-derivative hybrids **2.11-2.14**, displayed good trypanocidal activity in an ID_{50} range of between 5.6 to 7.7 μM, being more active than terbinafine and ketoconazole (ID_{50} =17μM and 10μM] and their 5-nitrofuran and 5-nitroindazole parents (ID_{50} = 8.2μM and 16.5μM]. The four compounds have inhibition activity similar to nifurtimox and benznidazole (ID_{50} =7.7μM and 7.4μM]. The positive results recommend these hybrids as starting points to generate new antitrypanosomal agents [68].

Cruzain is a *T. cruzi* cysteine-protease that plays an essential role in the survival of the parasite within the host cell, and is therefore an interesting target for trypanocidal drug design. Among the structures having cruzain inhibitory capacity, one of the most interesting is N-acylhydrazone [69]. Using the furoxan moiety, which is an NO donor, a series of forty hybrid molecules was synthesized. Compound (**2.15**] (E]-2-methyl-N'-(4-phenyl-3-furoxanylmethylidene]-4H-imidazo[1,2-α]pyridine-3-carbohydrazide stood out as a potential drug lead due to its *in vitro* anti-*T. cruzi* activity and selectivity. The inhibitory concentration of compound **2.15** in axenic epimastigotes of the Tulahuen 2 strain (IC_{50} = 5μM], and Tulahuen C4 strain amastigotes (IC_{50} = 0.91μM], was (in both cases] lower than the trypanosomal drug, nifurtimox, which presented respective IC_{50} values for epimastigotes and amastigotes of 7.7 and 10.4 μM. The results of the cytotoxicity assay using murine macrophages (J-774.1] suggest that compound **2.15** with an IC_{50} value of 56.2 μM is not only highly potent against *T. cruzi*, but is also one of the most selective compounds of the series (SI = 62] [70].

Another study performed with these pharmacophores revealed ten compounds from a set of fifteen molecules with anti-Chagas activity in a range of IC_{50} values from 2.88 to

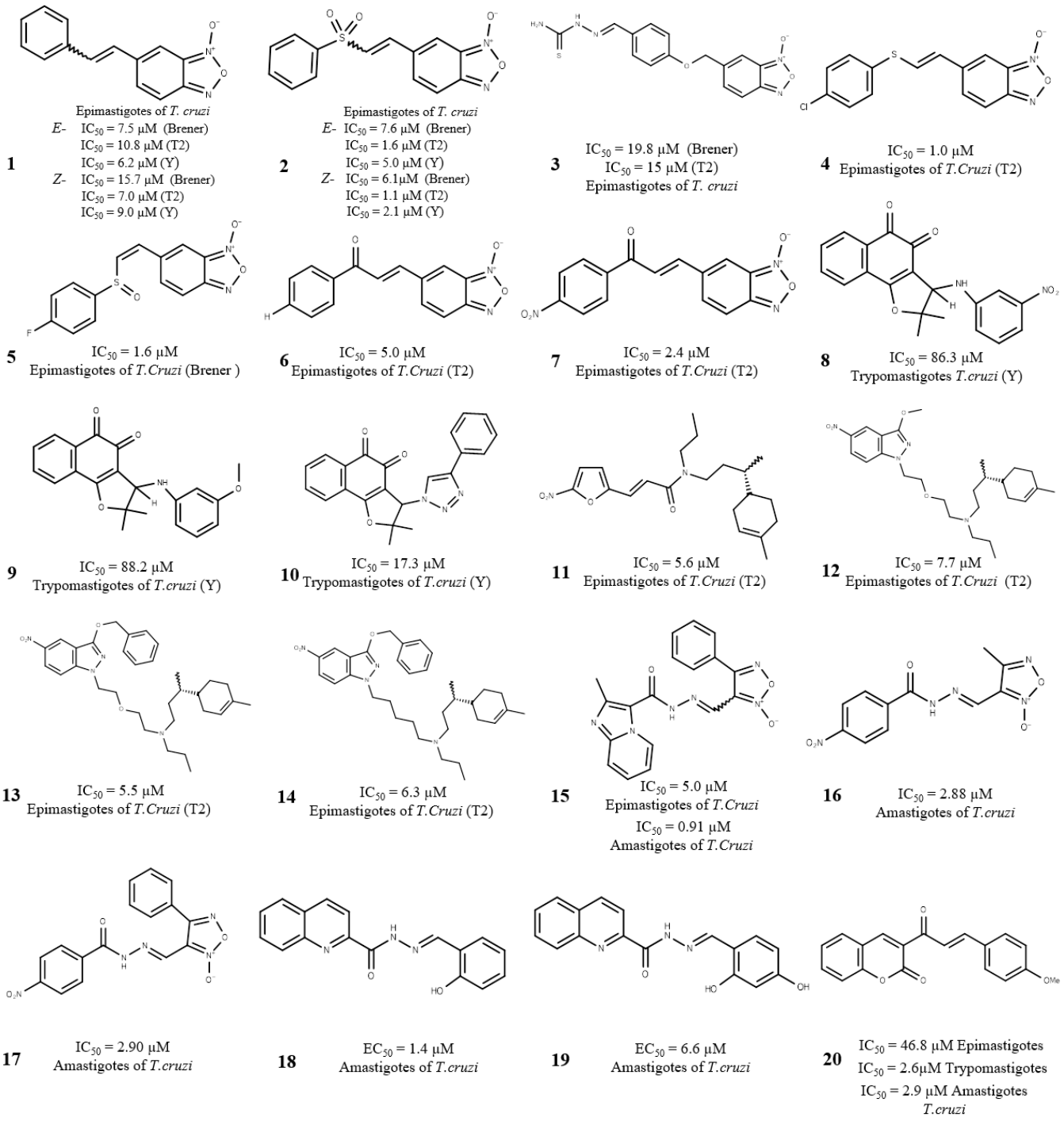


Figure 2: Hybrids active against differing *Trypanosoma cruzi* forms/strains with reported EC_{50} or IC_{50} values. Strains: Brener = CL Brener, T2 = Tulahuen 2, and Y strain.

27.43 μM against the amastigote form of *T. cruzi*. Compounds **2.16** (IC_{50} = 2.88 μM] and **2.17** (IC_{50} = 2.90 μM] were the most active compounds, with respectively high SI values (28.09 and 15.69,] in results obtained from a cytotoxic assay using the human hepatocellular carcinoma cell line HepG2. Compound **2.16** was more selective and almost three times less cytotoxic (CC_{50}= 80.9 μM] than benznidazole, the anti-Chagas drug used as the control, and which presented an SI = 18.7 (CC_{50}= 28.1 μM]. Structurally, these two active compounds are the only two of the series, which have a nitro group in the *para* position of their benzene ring. Looking for the mechanism of action of these hybrids, the authors found by cruzain inhibition assay that this enzyme is the biological target of both compounds **2.16** (IC_{50} = 11.2 μM] and **2.17** (IC_{50} = 15.1 μM]. It was shown by nitric oxide releasing tests that there is no relationship between NO releasing potential and trypanocidal activity for these two molecules [71] .

Among the multiple biological activities observed for the quinoline nucleus, (such as the previously mentioned leishmanicidal activity], this much studied class of molecules and their derivatives have been included in the search for

promising anti-Chagas compounds [72-74]. An example is the research of Coa et al., where in addition to the leishmanicidal compounds **1.18**, **1.19** and **1.20**; tests using *T. cruzi* intracellular amastigotes, showed that quinoline-hydrazone hybrids **2.18** and **2.19** have *in vitro* antitrypanosomal activity, with EC_{50} values of 1.4 and 6.6 μM, respectively, having in both cases lower active concentrations than benznidazole (EC_{50} =10.5μM] which was used as the positive control. In cytotoxicity assays with the U937 cell line it was observed that compound **2.18** was less cytotoxic (LC_{50} 56.0μM], and therefore more selective (SI =11.6], with respect to compound **2.19**, which had an LC_{50} value of 3.6 μM and SI = 0.5. Thus, the hybrid **2.18** shows promising *in vitro* results and can be considered as a hit molecule for developing new drugs against Chagas disease [55].

In addition to nitroheterocyclic compounds, certain non-nitrogen hybrids have also shown promising activity against *T. cruzi*. The trypanosomal activity of a series of five coumarin–chalcone hybrids was tested using the epimastigote form of *T. cruzi* clone Dm28c at two concentrations (10μM and 100μM]. Four of five synthesized molecules were active, reaching percentages of trypanocidal activity ranging from 9.5 to 23.8% (at 10μM], and from 24.1 to 47.9% (at 100μM]. Nifurtimox was used as the control, and presented trypanocidal activities of 52.5 and 100% [75]. Further study with these hybrids permited proposal of a good coumarin-chalcone scaffold, **2.20**, which yields *T. cruzi* activity. Percentage values for non-viable parasites of 78%, 96%, and 100% were respectively obtained for epimastigotes, trypomastigotes, and amastigotes in an *in vitro* test with a parasite concentration of 100μM. Further, the respective IC_{50} values of compound **2.20** were determined at 46.8, 2.6, and 2.9 μM for these three parasitic forms. This was more active than nifurtimox against the trypomastigote and amastigote forms (IC_{50} = 10.0 and 18.6μM]. However, a high cytotoxicity was observed for this hybrid both in RAW 264.7 and VERO cell lines. Based on the observed activity and its possible relationship with structural key features of methoxy substitution at positions 2' and 5', this coumarin-chalcone might well be used as a starting point for developing new anti-*T. cruzi* compounds [76].

3.2 Sleeping sickness

Studying the synergistic effects of benzoxaborole and chalcone scaffolds, Qiao et al. synthesized a series of hybrids containing these skeletons and later evaluated their activity against the *T. brucei* 427 strain. Compounds **3.1**, **3.2,** and **3.3** showed excellent *in vitro* potency against the trypomastigote forms with respective inhibitory growth concentrations (IC_{50}] of 0.086, 0.075, and 0.032 μM, and with **3.3** as the most potent anti-*T. brucei* compound of the whole series tested. In assays with the L929 cell line, **3.1** and **3.2** did not present cytotoxicity, unlike **3.3,** which had a CC_{50} of 1.45 μM. Despite this, the SI of this hybrid was greater than 100. *In vivo* assays were also performed using BALB/c mice for **3.2** and **3.3**; the results indicated 100% survival and zero parasitemia in mice. Therefore, one may consider these compounds as leads, with clear synergistic effects for **3.3** substituents 4-NH_2 and 3-OMe in its trypanosomal action [77].

Compound **3.4**, is a 3-substituted quinolinone derivative, which presents good activity against the *T. brucei* CMP strain; having a Minimum Effective Concentration (MEC] of 12.5 uM, and being interesting for its observed low cytotoxicity in the MCF7 cancer cell line [% viability = 84] [78]. From this, together with the trypanosomal properties of the chalcone core, Roussaki et al., designed a new series of quinolinone-chalcone hybrids and tested their anti-parasitic activity. Anti-*T. brucei* compounds **3.5** and **3.6** yielded the lowest *in vitro* IC_{50} test values, 2.6 and 3.3μM. Also, **3.7** and **3.8**, presented good activity against the same parasite (IC_{50} = 4.9μM]. According to the authors, the activity of these hybrids is related to its structural characteristics; mainly the substituents on ring B of the chalcone core, as well as an amidic nitrogen on the heterocyclic ring of the quinolinone motif. Assays performed with THP-1 cells showed that two compounds of this group (**3.6** and **3.7**] were cytotoxic with CC_{50} values of 26.2 and 23.3μM. In this same study, structural modification of the chalcone α,β-unsaturated carbonyl system were performed, from among these, two new pyrazoline analogues with antitrypanosomal potential were found, of these two, **3.9** was highlighted. In addition to its antiparasitic activity at an IC_{50} of 1.46 μM, it was more potent than all of the synthesized active hybrids without a pyrazoline ring, being non-cytotoxic for mammalian cells (THP-1] [79].

From the above results, the pirazoline scaffold turns out to be an interesting group for finding anti-*T. brucei* lead molecules, thus, others studies have been performed evaluating biological activities of hybrids containing this moiety. Havrylyuk, et al. synthesized a series of 4-thiazolidinone-pyrazoline compounds, testing their *in vitro* antiparasitic activity using the bloodstream form of *T. brucei gambiense*; testing as well the cytotoxicity of these molecules in the rat myoblast-derived cell line (L-6]. Of all the hybrids tested, compounds, **3.10** and **3.11**, are promising for having the lowest values of IC_{50} (0.6 and 0.7 μM] and high selectivity for the parasite (292 and 57.1]. The authors associated the activity of these compounds with N3 position substitutions in thiazolidinone; being that the presence of methyl and aryl groups leads to an increase in antitrypanosomal activity. In both cases, **3.10** and **3.11**, had lower IC_{50} values (4.4μm and 17.8, respectively] as well as higher selectivity than nifurtimox, which was used as the control [80].

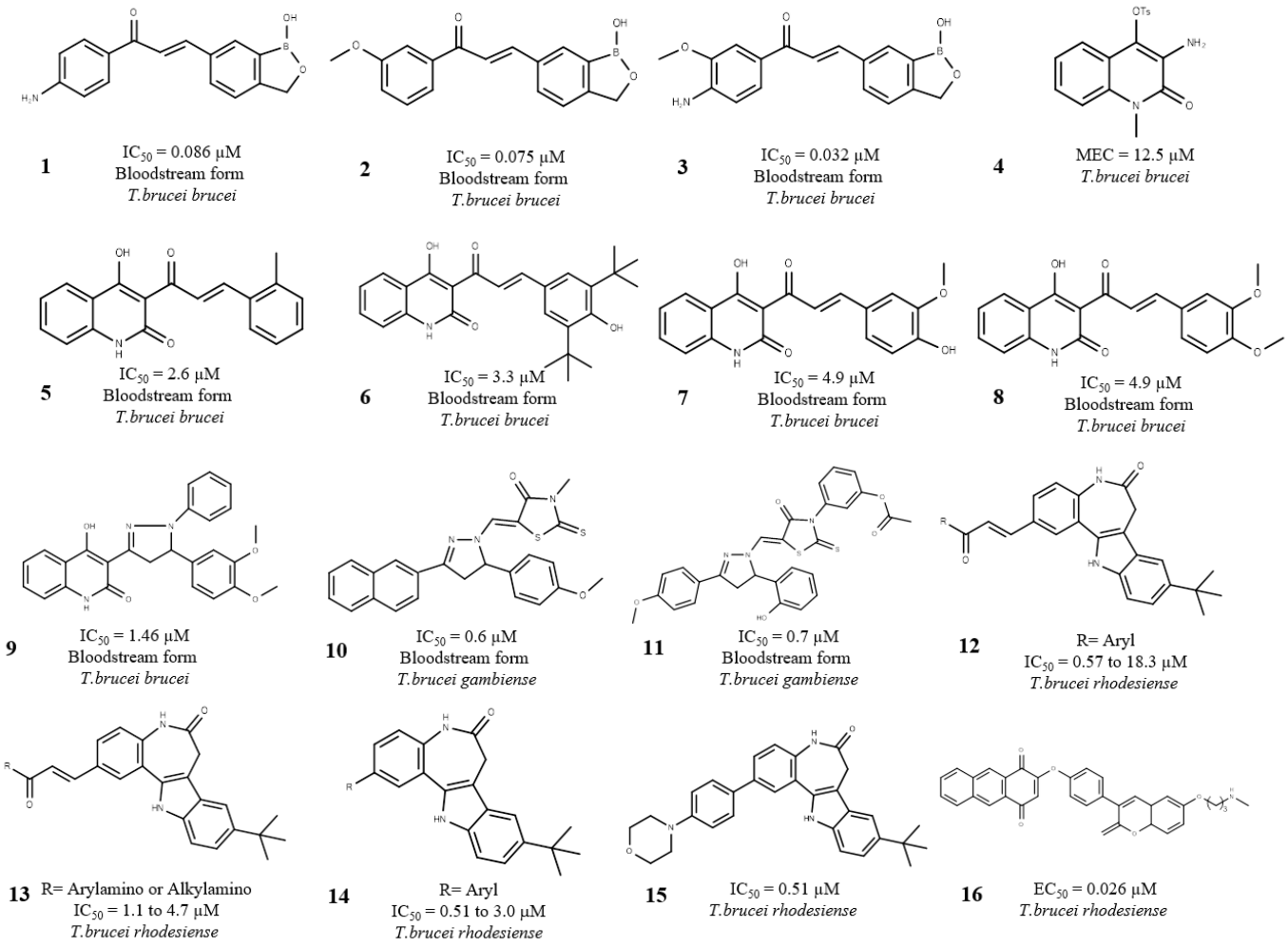


Figure 3: Hybrids active against different subspecies of *Trypanosoma brucei* with their reported Maximum effective concentration (MEC], EC_{50} or IC_{50} values.

In an interesting structure-activity relationship (SAR] study, several series of hybrids containing paullones, (a class of molecules which have in their structure the 7, 12-dihydroindolo[3,2-d][1] benzazepin-6(5H]-one scaffold] [81], and previously presenting antiparasitic activity [82], were synthesized. Initially, through covalent linking with chalcone **3.12** certain hybrid compounds were obtained. Trypanosomal inhibition screening was performed for these compounds using 5μM of *T. brucei rhodesiense* strain STIB900 bloodstream trypomastigotes. The results showed an inhibition range of from 35 to 100% with IC_{50} values of between 0.57 and 18.3μM. However these compounds displayed considerable toxicity in tests using the THP-1 cell line ($CC_{50:}$ 1.3-70μM], this behavior was related to the electrophilic properties of α,β-unsaturated carbonyl compounds. Looking to decrease these cytotoxic levels, a new group of acceptor modified **3.13** molecules was synthesized (by Michael reaction], generating an α,β-unsaturated amide group, being less electrophilic than their corresponding ketones. Nine of the ten compounds from this group had IC_{50} values of lower than 5μM, nevertheless cytotoxicity for these **3.13** structures was still high ($CC_{50:}$ 0.34-13.7μM] [83].

Being observed that the presence of an α,β-unsaturated system is related to toxicity for these derivatives, a third group of 2aryl-paullones was synthesized. Compound **3.14** was synthesized without an α,β-unsaturated partial structure. High anti-*T. brucei* activity for the whole series was observed with IC_{50} values ranging from 0.51 to 3.0 μM, as well as a considerable reduction in cytotoxicity levels for the human macrophage cell line. Compound **3.15,** 9-tert-butyl-2-(4-morpholinophenyl] paullone was the most active molecule in this study with an IC_{50} value of 0.51 μM, and high selectivity (SI =156.9]. These results indicated **3.14** as a promising group in the search for new trypanosomal drugs, structurally allowing the conclusion that the enone system in this type of compound is not related to activity, yet is important to cytotoxic effects [83].

Finally, in a study performed with two series of quinone-coumarin hybrids, **3.16,** was the most effective molecule against the *T. brucei rhodesiense* bloodstream stage, STIB 900 strain (EC_{50} = 0.026 μM], being highly selective against *Trypanosoma* (SI= 301.8]. The anti-*T. brucei* activity of this compound is similar to that observed for melarsoprol (EC_{50} = 0.01 μM], the reference drug used in the study. The low toxicity shown for **3.16** in mammalian L6 cells (CC_{50} = 7.95 μM] was associated by the authors, at least in part, to the human glutathione reductase inhibition test results. Since interference with this enzyme was not observed [84].

CONCLUSION

In the search for safer and more effective chemotherapies against the varied parasitic *Leishmania* and *Trypanosoma* species that cause these three principal NTDs, the hybrid approach has emerged as an excellent way to design and develop biologically active molecules. *In vitro* and *in vivo* activity testing against these diverse parasitic forms has revealed various hybrid compounds with promising antiparasitic activity, and in some cases even lower IC_{50} values than current treatments; an excellent starting point.

However, in many cases, such lead molecules are highly toxic to the mammalian cell, for this reason and since only certain targets such as Cruzipain and Trypanothione reductase have been identified, Structure-activity relationship (SAR] studies are still necessary to optimize active molecules, to reveal their physicochemical characteristics, and to describe their mechanisms of action,

LIST OF ABBREVIATIONS

$CC_{50:}$ 50% cytotoxic concentration
CP: Cruzipain
MEC: Maximum effective concentration
DMPO: 5,5-dimethyl-1-pirroline-N-oxide
DNA: Deoxyribonucleic acid
EC_{50}: 50% Effective concentration
HAT: Human African Trypanosomiasis
HIV: Human immunodeficiency virus
$IC_{50:}$ 50% Inhibitory concentration
$ID_{50:}$ 50% Inhibitory dose
IL: Interleukin
NTD: Neglected Tropical Disease
RNA: Ribonucleic acid
SI: Selective index (CC_{50}/IC_{50}]
SSG: Sodium stibogluconate
TGF-β: Transforming growth factor beta
TNF-α: Tumor necrosis factor alpha
TR: Trypanothione reductase

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank both CNPq and Capes for financial Support, and the Graduate Student Agreement Program—PEC-PG of CNPq—Brazil.

REFERENCES

1. Khare S, Nagle AS, Biggart A, Lai YH, Liang F, Davis LC, et al. Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. Nature. 2016;537(7619]:229-33.
2. Pearson RD, Sousa AQ. Clinical spectrum of Leishmaniasis. Clin Infect Dis. 1996;22(1]:1-13.
3. Herrera Acevedo C, Scotti L, Feitosa Alves M, Formiga Melo Diniz MDF, Scotti MT. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. Molecules. 2017;22(1]:79.
4. Burrows JN, Elliott RL, Kaneko T, Mowbray CE, Waterson D. The role of modern drug discovery in the fight against neglected and tropical diseases. Med Chem Comm. 2014;5(6]:688-700.
5. Clementi A, Battaglia G, Floris M, Castellino P, Ronco C, Cruz DN. Renal involvement in leishmaniasis: a review of the literature. NDT plus. 2011;4(3]:147-52.
6. Shaw CD, Lonchamp J, Downing T, Imamura H, Freeman TM, Cotton JA, et al. In vitro selection of miltefosine resistance in promastigotes of Leishmania donovani from Nepal: genomic and metabolomic characterization. (1365-2958 (Electronic]].
7. Perez-Victoria FJ, Castanys S Fau - Gamarro F, Gamarro F. Leishmania donovani resistance to miltefosine involves a defective inward translocation of the drug. (0066-4804 (Print]].
8. Hendrickx S, Inocêncio da Luz RA, Bhandari V, Kuypers K, Shaw CD, Lonchamp J, et al. Experimental Induction of Paromomycin Resistance in Antimony-Resistant Strains of L. donovani: Outcome Dependent on In Vitro Selection Protocol. PLoS Neglected Tropical Diseases. 2012;6(5]:e1664.
9. Singh N, Kumar M, Singh RK. Leishmaniasis: Current status of available drugs and new potential drug targets. Asian Pacific Journal of Tropical Medicine. 2012;5(6]:485-97.
10. McGwire BS, Satoskar AR. Leishmaniasis: clinical syndromes and treatment. QJM. 2014;107(1]:7-14.
11. Apt W. Current and developing therapeutic agents in the treatment of Chagas disease. Drug Des Dev Ther. 2010;4:243-53.
12. Viotti R, Vigliano C, Lococo B, Alvarez MG, Petti M, Bertocchi G, et al. Side effects of benznidazole as treatment in chronic Chagas disease: fears and realities. Expert review of anti-infective therapy. 2009;7(2]:157-63.
13. Nok AJ. Arsenicals (melarsoprol], pentamidine and suramin in the treatment of human African trypanosomiasis. Parasitol Res. 2003;90(1]:71-9.
14. Babokhov P, Sanyaolu AO, Oyibo WA, Fagbenro-Beyioku AF, Iriemenam NC. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. Pathog Glob Health. 2013;107(5]:242-52.
15. Marsiccobetre S, Rodríguez-Acosta A, Lang F, Figarella K, Uzcátegui NL. Aquaglyceroporins Are the Entry Pathway of Boric Acid in Trypanosoma brucei. Biochimica et Biophysica Acta (BBA] - Biomembranes. 2017;1859(5]:679-85.
16. Collaborators GMaCoD. Global, regional, and national age–sex specifi c all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;385(9963]:117-71.
17. Strub-Wourgaft N. NEGLECTED INFECTIOUS DISEASES. BMJ Global Health. 2017;2(Suppl 2]:A4-A.
18. Muller-Schiffmann A, Sticht H Fau - Korth C, Korth C. Hybrid compounds: from simple combinations to nanomachines. (1179-190X (Electronic]].
19. Decker M. Hybrid molecules incorporating natural products: applications in cancer therapy, neurodegenerative disorders and beyond. Current Medicinal Chemistry. 2011;18(10]:1464-75.
20. Muregi FW, Ishih A. Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design. Drug development research. 2010;71(1]:20-32.
21. Walsh J, Bell A. Hybrid drugs for malaria. Current pharmaceutical design. 2009;15(25]:2970-85.
22. Sashidhara KV, Kumar A, Kumar M, Sarkar J, Sinha S. Synthesis and in vitro evaluation of novel coumarin–chalcone hybrids as potential anticancer agents. Bioorganic & medicinal chemistry letters. 2010;20(24]:7205-11.
23. Belluti F, Fontana G, Dal Bo L, Carenini N, Giommarelli C, Zunino F. Design, synthesis and anticancer activities of stilbene-coumarin hybrid compounds: Identification of novel proapoptotic agents. Bioorganic & medicinal chemistry. 2010;18(10]:3543-50.
24. Dechy-Cabaret O, Benoit-Vical F, Robert A, Meunier B. Preparation and Antimalarial Activities of "Trioxaquines", New Modular Molecules with a Trioxane Skeleton Linked to a 4-Aminoquinoline. ChemBioChem. 2000;1(4]:281-3.
25. Walsh JJ, Coughlan D, Heneghan N, Gaynor C, Bell A. A novel artemisinin–quinine hybrid with potent antimalarial activity. Bioorganic & medicinal chemistry letters. 2007;17(13]:3599-602.
26. Wang Y, Damu GL, Lv J-S, Geng R-X, Yang D-C, Zhou C-H. Design, synthesis and evaluation of clinafloxacin triazole hybrids as a new type of antibacterial and antifungal agents. Bioorganic & medicinal chemistry letters. 2012;22(17]:5363-6.
27. Michalska P. Novel multitarget hybrid compounds for the treatment of Alzheimer's disease. Current Topics in Medicinal Chemistry. 2016;16:1-17.
28. Dodd RH, Ouannes C, Robert-Gero M, Potier P. Hybrid molecules: Growth inhibition of leishmania donovani promastigotes by

thiosemicarbazones of 3-carboxy-. beta.-carbolines. Journal of medicinal chemistry. 1989;32(6]:1272-6.
29. Bell CA, Hall JE, Kyle DE, Grogl M, Ohemeng KA, Allen MA, et al. Structure-activity relationships of analogs of pentamidine against Plasmodium falciparum and Leishmania mexicana amazonensis. Antimicrobial Agents and Chemotherapy. 1990;34(7]:1381-6.
30. Sundar S, Singh A. Recent developments and future prospects in the treatment of visceral leishmaniasis. Therapeutic Advances in Infectious Disease. 2016;3(3-4]:98-109.
31. Bialonska D, Zjawiony JK. Aplysinopsins - Marine Indole Alkaloids: Chemistry, Bioactivity and Ecological Significance. Marine Drugs. 2009;7(2]:166-83.
32. Porwal S, Chauhan Ss Fau - Chauhan PMS, Chauhan Pm Fau - Shakya N, Shakya N Fau - Verma A, Verma A Fau - Gupta S, Gupta S. Discovery of novel antileishmanial agents in an attempt to synthesize pentamidine-aplysinopsin hybrid molecule. (1520-4804 (Electronic]].
33. Costa EV, Pinheiro MLB, Xavier CM, Silva JR, Amaral ACF, Souza AD, et al. A Pyrimidine-β-carboline and Other Alkaloids from Annona f oetida with Antileishmanial Activity. Journal of natural products. 2006;69(2]:292-4.
34. Chen M, Christensen SB, Blom J, Lemmich E, Nadelmann L, Fich K, et al. Licochalcone A, a novel antiparasitic agent with potent activity against human pathogenic protozoan species of Leishmania. Antimicrobial Agents and Chemotherapy. 1993;37(12]:2550-6.
35. Chen M, Christensen SB, Theander TG, Kharazmi A. Antileishmanial activity of licochalcone A in mice infected with Leishmania major and in hamsters infected with Leishmania donovani. Antimicrobial Agents and Chemotherapy. 1994;38(6]:1339-44.
36. Tyagi V, Khan S, Shivahare R, Srivastava K, Gupta S, Kidwai S, et al. A natural product inspired hybrid approach towards the synthesis of novel pentamidine based scaffolds as potential anti-parasitic agents. Bioorganic & Medicinal Chemistry Letters. 2013;23(1]:291-6.
37. Junior CGL, de Assis PAC, Silva FPL, Sousa SCO, de Andrade NG, Barbosa TP, et al. Efficient synthesis of 16 aromatic Morita–Baylis–Hillman adducts: Biological evaluation on Leishmania amazonensis and Leishmania chagasi. Bioorganic Chemistry. 2010;38(6]:279-84.
38. Barbosa TP, Sousa SC, Amorim FM, Rodrigues YK, de Assis PA, Caldas JP, et al. Design, synthesis and antileishmanial in vitro activity of new series of chalcones-like compounds: a molecular hybridization approach. Bioorganic & medicinal chemistry. 2011;19(14]:4250-6.
39. Sunduru N, Agarwal A, Katiyar SB, Nishi, Goyal N, Gupta S, et al. Synthesis of 2,4,6-trisubstituted pyrimidine and triazine heterocycles as antileishmanial agents. Bioorganic & Medicinal Chemistry. 2006;14(23]:7706-15.
40. Gupta L, Sunduru N Fau - Verma A, Verma A Fau - Srivastava S, Srivastava S Fau - Gupta S, Gupta S Fau - Goyal N, Goyal N Fau - Chauhan PMS, et al. Synthesis and biological evaluation of new [1,2,4]triazino[5,6-b]indol-3-ylthio-1,3,5-triazines and [1,2,4]triazino[5,6-b]indol-3-ylthio-pyrimidines against Leishmania donovani. (1768-3254 (Electronic]].
41. Sharma R, Pandey AK, Shivahare R, Srivastava K, Gupta S, Chauhan PM. Triazino indole-quinoline hybrid: a novel approach to antileishmanial agents. (1464-3405 (Electronic]].
42. Sharma M, Chauhan K, Shivahare R, Vishwakarma P, Suthar MK, Sharma A, et al. Discovery of a New Class of Natural Product-Inspired Quinazolinone Hybrid as Potent Antileishmanial agents. Journal of Medicinal Chemistry. 2013;56(11]:4374-92.
43. Kumar R, Khan S, Verma A, Srivastava S, Viswakarma P, Gupta S, et al. Synthesis of 2-(pyrimidin-2-yl]-1-phenyl-2,3,4,9-tetrahydro-1H-beta-carbolines as antileishmanial agents. European journal of medicinal chemistry. 2010;45(8]:3274-80.
44. Chauhan SS, Gupta L, Mittal M, Vishwakarma P, Gupta S, Chauhan PM. Synthesis and biological evaluation of indolyl glyoxylamides as a new class of antileishmanial agents. Bioorg Med Chem Lett. 2010;20(21]:6191-4.
45. Tovar J, Wilkinson S, Mottram JC, Fairlamb AH. Evidence that trypanothione reductase is an essential enzyme in Leishmania by targeted replacement of the tryA gene locus. Molecular microbiology. 1998;29(2]:653-60.
46. Chauhan SS, Pandey S, Shivahare R, Ramalingam K, Krishna S, Vishwakarma P, et al. Novel [small beta]-carboline-quinazolinone hybrid as an inhibitor of Leishmania donovani trypanothione reductase: Synthesis, molecular docking and bioevaluation. MedChemComm. 2015;6(2]:351-6.
47. Kumar S, Bawa S, Gupta H. Biological Activities of Quinoline Derivatives. Mini Reviews in Medicinal Chemistry. 2009;9(14]:1648-54.
48. Michael JP. Quinoline, quinazoline and acridone alkaloids. Natural Product Reports. 2008;25(1]:166-87.
49. Supong K, Thawai C, Supothina S, Auncharoen P, Pittayakhajonwut P. Antimicrobial and anti-oxidant activities of quinoline alkaloids from Pseudomonas aeruginosa BCC76810. Phytochemistry Letters. 2016;17:100-6.
50. Faruk Khan MO, Levi MS, Tekwani BL, Wilson NH, Borne RF. Synthesis of isoquinuclidine analogs of chloroquine: Antimalarial and antileishmanial activity. Bioorganic & Medicinal Chemistry. 2007;15(11]:3919-25.
51. Nava-Zuazo C, Estrada-Soto S Fau - Guerrero-Alvarez J, Guerrero-Alvarez J Fau - Leon-Rivera I, Leon-Rivera I Fau - Molina-Salinas GM, Molina-Salinas Gm Fau - Said-Fernandez S, Said-Fernandez S Fau - Chan-Bacab MJ, et al. Design, synthesis, and in vitro antiprotozoal, antimycobacterial activities of N-{2-[(7-chloroquinolin-4-yl]amino]ethyl}ureas. (1464-3391 (Electronic]].
52. Rando DG, Avery Ma Fau - Tekwani BL, Tekwani Bl Fau - Khan SI, Khan Si Fau - Ferreira EI, Ferreira EI. Antileishmanial activity screening of 5-nitro-2-heterocyclic benzylidene hydrazides. (1464-3391 (Electronic]].
53. Khan KM, Rasheed MFAUZ-U, Zia-Ullah Fau - Hayat S, Hayat S Fau - Kaukab F, Kaukab F Fau - Choudhary MI, Choudhary MIFAUA-u-R, et al. Synthesis and in vitro leishmanicidal activity of some hydrazides and their analogues. (0968-0896 (Print]].
54. Bernardino AMR, Gomes AO, Charret KS, Freitas ACC, Machado GMC, Canto-Cavalheiro MM, et al. Synthesis and leishmanicidal activities of 1-(4-X-phenyl]-N′-[(4-Y-phenyl]methylene]-1H-pyrazole-4-carbohydrazides. European journal of medicinal chemistry. 2006;41(1]:80-7.
55. Coa JC, Castrillon W, Cardona W, Carda M, Ospina V, Munoz JA, et al. Synthesis, leishmanicidal, trypanocidal and cytotoxic activity of quinoline-hydrazone hybrids. (1768-3254 (Electronic]].
56. Chan MM, Grogl M, Chen CC, Bienen EJ, Fong D. Herbicides to curb human parasitic infections: in vitro and in vivo effects of trifluralin on the trypanosomatid protozoans. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(12]:5657-61.
57. Chan MM, Grogl M, Callahan H, Fong D. Efficacy of the herbicide trifluralin against four P-glycoprotein-expressing strains of Leishmania. Antimicrobial Agents and Chemotherapy. 1995;39(7]:1609-11.
58. Sundar S, Olliaro PL. Miltefosine in the treatment of leishmaniasis: Clinical evidence for informed clinical risk management. Therapeutics and Clinical Risk Management. 2007;3(5]:733-40.
59. Godinho JL, Georgikopoulou K Fau - Calogeropoulou T, Calogeropoulou T Fau - de Souza W, de Souza W Fau - Rodrigues JCF, Rodrigues JC. A novel alkyl phosphocholine-dinitroaniline hybrid molecule exhibits biological activity in vitro against Leishmania amazonensis. (1090-2449 (Electronic]].
60. Tiwari A, Kumar S, Shivahare R, Kant P, Gupta S, Suryawanshi SN. Chemotherapy of leishmaniasis part XIII: design and synthesis of novel heteroretinoid-bisbenzylidine ketone hybrids as antileishmanial agents. (1464-3405 (Electronic]].
61. Porcal W, Hernandez P Fau - Boiani M, Boiani M Fau - Aguirre G, Aguirre G Fau - Boiani L, Boiani L Fau - Chidichimo A, Chidichimo A Fau - Cazzulo JJ, et al. In vivo anti-Chagas vinylthio-, vinylsulfinyl-, and vinylsulfonylbenzofuroxan derivatives. (0022-2623 (Print]].
62. Porcal W, Hernandez P Fau - Aguirre G, Aguirre G Fau - Boiani L, Boiani L Fau - Boiani M, Boiani M Fau - Merlino A, Merlino A Fau - Ferreira A, et al. Second generation of 5-ethenylbenzofuroxan derivatives as inhibitors of Trypanosoma cruzi growth: synthesis, biological evaluation, and structure-activity relationships. (0968-0896 (Print]].
63. Aguirre G, Boiani L, Boiani M, Cerecetto H, Di Maio R, González M, et al. New potent 5-substituted benzofuroxans as inhibitors of Trypanosoma cruzi growth: Quantitative structure–activity relationship studies. Bioorganic & medicinal chemistry. 2005;13(23]:6336-46.
64. Porcal W, Hernández P, Boiani L, Boiani M, Ferreira A, Chidichimo A, et al. New trypanocidal hybrid compounds from the association of hydrazone moieties and benzofuroxan heterocycle. Bioorganic & Medicinal Chemistry. 2008;16(14]:6995-7004.
65. Castro D, Boiani L, Benitez D, Hernández P, Merlino A, Gil C, et al. Anti-trypanosomatid benzofuroxans and deoxygenated analogues: Synthesis using polymer-supported triphenylphosphine, biological evaluation and mechanism of action studies. European journal of medicinal chemistry. 2009;44(12]:5055-65.
66. da Silva Júnior EN, de Souza MCB, Fernandes MC, Menna-Barreto RF, Maria do Carmo F, de Assis Lopes F, et al. Synthesis and anti-

Trypanosoma cruzi activity of derivatives from nor-lapachones and lapachones. Bioorganic & medicinal chemistry. 2008;16(9):5030-8.
67. da Silva EN, Menna-Barreto RF, Maria do Carmo F, Silva RS, Teixeira DV, de Souza MCB, et al. Naphthoquinoidal [1, 2, 3]-triazole, a new structural moiety active against Trypanosoma cruzi. European journal of medicinal chemistry. 2008;43(8]:1774-80.
68. Alvarez G, Gerpe A, Benitez D, Garibotto F, Zacchino S, Graebin CS, et al. New Limonene-Hybrid Derivatives with Anti-T. cruzi Activity. Letters in Drug Design & Discovery. 2010;7(6]:452-60.
69. Romeiro NC, Aguirre G Fau - Hernandez P, Hernandez P Fau - Gonzalez M, Gonzalez M Fau - Cerecetto H, Cerecetto H Fau - Aldana I, Aldana I Fau - Perez-Silanes S, et al. Synthesis, trypanocidal activity and docking studies of novel quinoxaline-N-acylhydrazones, designed as cruzain inhibitors candidates. (1464-3391 (Electronic]].
70. Hernández P, Rojas R, Gilman RH, Sauvain M, Lima LM, Barreiro EJ, et al. Hybrid furoxanyl N-acylhydrazone derivatives as hits for the development of neglected diseases drug candidates. European journal of medicinal chemistry. 2013;59:64-74.
71. Massarico Serafim RA, Gonçalves JE, de Souza FP, de Melo Loureiro AP, Storpirtis S, Krogh R, et al. Design, synthesis and biological evaluation of hybrid bioisoster derivatives of N-acylhydrazone and furoxan groups with potential and selective anti-Trypanosoma cruzi activity. European journal of medicinal chemistry. 2014;82:418-25.
72. Marella A, Tanwar OP, Saha R, Ali MR, Srivastava S, Akhter M, et al. Quinoline: A versatile heterocyclic. Saudi Pharmaceutical Journal. 2013;21(1]:1-12.
73. Cretton S, Breant L, Pourrez L, Ambuehl C, Marcourt L, Ebrahimi SN, et al. Antitrypanosomal Quinoline Alkaloids from the Roots of Waltheria indica. Journal of Natural Products. 2014;77(10]:2304-11.
74. Lechuga GC, Borges JC, Calvet CM, de Araújo HP, Zuma AA, do Nascimento SB, et al. Interactions between 4-aminoquinoline and heme: Promising mechanism against Trypanosoma cruzi. International Journal for Parasitology: Drugs and Drug Resistance. 2016;6(3]:154-64.
75. Vazquez-Rodriguez S, Figueroa-Guinez R, Matos MJ, Santana L, Uriarte E, Lapier M, et al. Synthesis of coumarin-chalcone hybrids and evaluation of their antioxidant and trypanocidal properties. MedChemComm. 2013;4(6]:993-1000.
76. Vazquez-Rodriguez S, Guinez RF, Matos M, Olea-Azar C, Maya J, Uriarte E, et al. Synthesis and trypanocidal properties of new coumarin–chalcone derivatives. Med Chem. 2015;5:173-7.
77. Qiao Z, Wang Q, Zhang F, Wang Z, Bowling T, Nare B, et al. Chalcone–Benzoxaborole Hybrid Molecules as Potent Antitrypanosomal Agents. Journal of Medicinal Chemistry. 2012;55(7]:3553-7.
78. Audisio D, Messaoudi S, Cojean S, Peyrat J-F, Brion J-D, Bories C, et al. Synthesis and antikinetoplastid activities of 3-substituted quinolinones derivatives. European journal of medicinal chemistry. 2012;52:44-50.
79. Roussaki M, Hall B Fau - Lima SC, Lima Sc Fau - da Silva AC, da Silva Ac Fau - Wilkinson S, Wilkinson S Fau - Detsi A, Detsi A. Synthesis and anti-parasitic activity of a novel quinolinone-chalcone series. (1464-3405 (Electronic]].
80. Havrylyuk D, Zimenkovsky B, Karpenko O, Grellier P, Lesyk R. Synthesis of pyrazoline–thiazolidinone hybrids with trypanocidal activity. European journal of medicinal chemistry. 2014;85:245-54.
81. Tolle N, Kunick C. Paullones as inhibitors of protein kinases. Current topics in medicinal chemistry. 2011;11(11]:1320-32.
82. Reichwald C, Shimony O, Dunkel U, Sacerdoti-Sierra N, Jaffe CL, Kunick C. 2-(3-Aryl-3-oxopropen-1-yl]-9-tert-butyl-paullones: A New Antileishmanial Chemotype. Journal of Medicinal Chemistry. 2008;51(3]:659-65.
83. Ryczak J, Papini Ma, Lader A, Nasereddin A, Kopelyanskiy D, Preu L, et al. 2-Arylpaullones are selective antitrypanosomal agents. European journal of medicinal chemistry. 2013;64:396-400.
84. Belluti F, Uliassi E, Veronesi G, Bergamini C, Kaiser M, Brun R, et al. Toward the Development of Dual-Targeted Glyceraldehyde-3-phosphate Dehydrogenase/Trypanothione Reductase Inhibitors against Trypanosoma brucei and Trypanosoma cruzi. ChemMedChem. 2014;9(2]:371-82.

Marcus Scotti
para mí

25 oct.

inglés › español Traducir mensaje Desactivar para: inglés ×

---------- Forwarded message ----------
From: **Noureen** <noureenazhar@benthamscience.org>
Date: Tue, Oct 24, 2017 at 4:27 PM
Subject: FW: Query Regarding Graphics Enhancement (GD/2017/LOC-6949)
To: luciana.scotti@gmail.com
Cc: loc@benthamscience.org, Editorial Office <editorial@benthamscience.org>

Ref # 48011

Dear Dr. Scotti,

Thank you for your email in connection with your manuscript entitled **"Hybrid Compounds in the Search for Alternative Chemotherapies against Neglected Tropical Diseases"** submitted for publication in **"Letters in Organic Chemistry"**. We are pleased to inform you that the provided figure(s) being as per the publication standard will be duly proceeded for publication.

We appreciate your kind cooperation in this respect.

Regards,

Noureen Azher
Manager
Graphics Dept.

molecules

MDPI

Article

Ligand-based virtual screening of a benzylisoquinoline alkaloids dataset with potential anti-inflammatory activity.

Chonny Herrera-Acevedo[1], Mateus Feitosa Alves[1], Luciana Scotti[1], Margareth F.F.M. Diniz[1] and Marcus Tullius Scotti[1, *]

[1] Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil; caherrera@ltf.ufpb.br (C.H) ; mateusfalves@gmail.com (M.F); luciana.scotti@gmail.com (L.S); margareth@ltf.ufpb.br (M.D) ; mtscotti@gmail.com (M.S)

* Correspondence: mtscotti@gmail.com; Tel.: +55-83-99869-0415

Abstract: Inhibitor of nuclear factor kappa B kinase beta subunit (IKK-β) and extracellular signal-regulated kinase 1 (ERK1) are two proteins involved in cytokine intracellular signaling pathways, which have great importance due to their anti-inflammatory role. In this work, from the ChEMBL database 775 and 48 structures with activity against IKK-β and ERK1, respectively, were obtained. The compounds were classified using values of pIC_{50}, presenting a range of 4.29 (from 5.01 to 9.30) for IKK-β and 3.10 (from 5.05 to 8.15) for ERK1. From SMILES codes, two-dimensional (2-D) structures were generated and then 2-D molecular descriptors were calculated. QSAR models were performed using a Random Forest (RF) algorithm. Models were evaluated through cross-validation (leave-one-out), Q^2_{LOO} = 0.69 and 0.66 as well as external test, Q^2_{ext} = 0.74 and 0.58 for IKK-β and ERK1, respectively. Finally, the pIC_{50} values of 179 benzylisoquinoline alkaloids (BIAs) were predicted in the QSAR models and four compounds were found with the highest activity for each protein studied, which through a molecular docking analysis showed characteristic interactions such as a hydrogen bond between residue lysine 44 of human IKK-β with hydroxyl group of best-ranked BIAs for this target, as well as the steric interaction present for the four most active BIAs for ERK1 with alanine 52

Keywords: benzylisoquinoline alkaloids, IKK-β, ERK1, anti-inflammatory activity, virtual screening, molecular docking.

1. Introduction

The inflammatory process is a nonspecific complex, stereotype, coordinated response of tissues to injury. Inflammation is evident in infectious diseases, but in recent times research has increasingly shown that several noninfectious diseases may present inflammatory conditions. This fact becomes important as new anti-inflammatory drugs emerge with different targets for the treatment of diseases such as allergy, asthma, rheumatic diseases and atherosclerosis [1-5].

In an inflammatory process, the production of cytokines by activated mast cells is a consequence of the genetic transcription of recently induced cytokines. The protein kinase family of mitogen-activated kinases (MAPK) are normally activated during inflammation, where three families are characterized: JNK (c-Jun *N*-terminal kinase), p38 and ERK (extracellular signal-regulated kinase) [6,7]. Pathways leading to cytokine gene expression require the nucleotide exchange factors guanylin, VAV and SOS, shifting the RAS balance from inactive to active. RAS regulates the RAF pathway positively, leading to phosphorylation and activation of MAP kinases (MAPKs) regulated by ERK1 and ERK2 [8-10]. Extracellular signal-regulated kinase 1 (ERK1, E.C. 2.7.11.24) is one of the two isoforms of ERK described. It is present in the RAS/RAF/MERK/ERK signaling pathway, which has

vital importance because many essential cell processes are involved. In ~30% of cancers and cognitive disorders, an abnormal activation exists [11].

Another way to phosphorylate and activate MAPKs is through the JNKs, phosphorylated by VAV, promoting activation of the p38 pathway [12]. All activate cytokine transcription factors and promote T-cell (NFAT) and nuclear factor-κB (NF-κB) activation, in addition to AP-1 activation. These factors stimulate transcription of various cytokines and tumor necrosis factor [13].

NF-κB is a transcriptional factor, which plays a key role in numerous physiological processes, including inflammatory processes [14]. NF-κB activation is stimulated by a kinase complex, IκB kinase (IKK), which is composed of three core proteins: IKK1/IKK-α, IKK2/IKK-β and NEMO/IKK-g [14,15], IKK-α and IKK-β being two catalytic subunits that are structurally related kinases. IKK-β (E.C. 2.7.11.10), is an interesting target due to its role in inflammation-induced tumor growth and progression, as well as an important modulator of tumor surveillance and rejection [15].

Secondary metabolites, such as alkaloids, can be found with different biological activities: antipyretic activities, anti-inflammatory, antiallergic, bronchodilator and immunomodulatory activities [16]. Benzylisoquinoline alkaloids (BIA) are metabolites, which present a great diversity (~2,500 BIAs are known today) and several pharmacological activities, such as antimicrobial agents, muscle relaxants, and potential anticancer drugs [17,18].

In this work, 775 and 48 structures with activity against IKK-β and ERK1, respectively, were obtained from the ChEMBL database. After calculating two-dimensional (2-D) molecular descriptors, QSAR models for each target using a Random Forest (RF) algorithm were performed to predict the anti-inflammatory activity of 179 BIA structures. Finally, structural characteristics of the best-ranked structures were evaluated through molecular docking.

2. Results and discussion

The calculation of 128 descriptors by the Dragon 7 program for each structure dataset took approximately 30 minutes using a computer with an i7-4790 processor, running at 3.6 GHz and equipped with 16 GB of RAM. From the 3,839 Dragon 7 2-D descriptors calculated for each compound of the ChEMBL dataset for the two proteins, those descriptors with constant or only one different value, standard deviation values larger than 4σ and that showed intercorrelation coefficients higher than 0.95 were excluded. Finally, for IKK-β 1062 (27.7%) 2-D descriptors while for ERK1 665 (17.3%) 2-D descriptors were used in the QSAR models' construction.

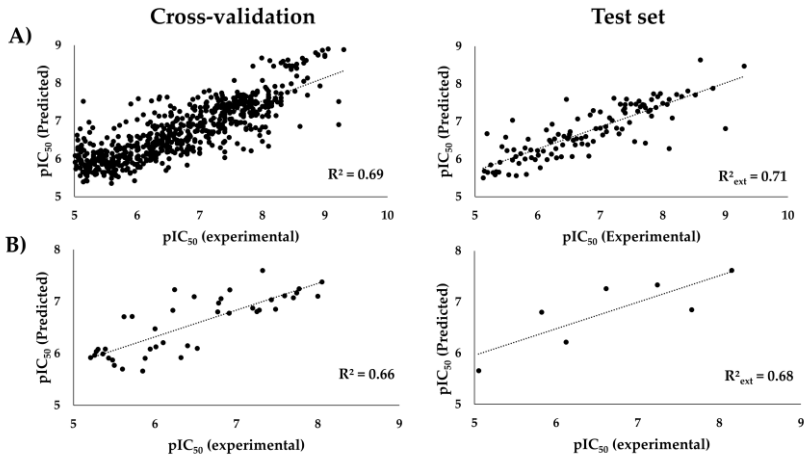
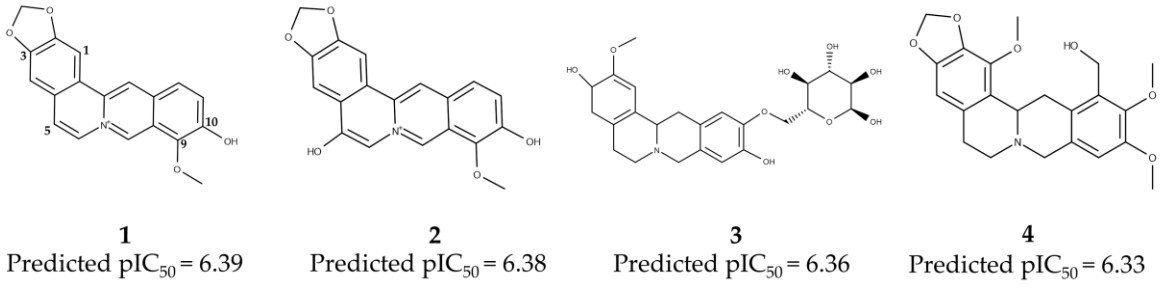


Figure 1: Plot of experimental vs predicted values of pIC_{50} for cross-validation and test set generated by A) IKK-β and B) ERK1 QSAR models

The 2-D descriptors, together with their dependent variable (pIC_{50} value) were used as input data in the Knime program for generating the RF models. QSAR models were evaluated through the Q^2 parameter both cross-validation (leave-one-out) and external test, obtaining Q^2_{LOO} values for internal validation of 0.69 and 0.66 as well as Q^2_{ext} for external test 0.74 and 0.58 for IKK-β and ERK1, respectively. Figure 1 presents the experimental versus predicted values of pIC_{50} and coefficient of determination for the two QSAR models. IC_{50} is the concentration that inhibits 50% of enzymatic activity.

The domain of applicability was used to check the samples in the test set regarding the reliability of predictions, 88.9% of compounds in the test set of the ERK1 model (one compound) and 96.1% in the test set of the IKK-β model (six compounds) were classified as unreliable. Then, pIC_{50} values of 179 BIAs were predicted against IKK-β and ERK1 using the two QSAR models.

For the dataset of 179 BIAs, 16% calculated 2-D-descriptors were used, the other descriptors were excluded under the criteria mentioned above. Subsequently, these results were imported into Knime to predict the pIC_{50} values. For IKK-β, predicted values presented a range between 5.77 and 6.39, while for ERK1 the pIC_{50} values calculated by the QSAR model were between 5.62 and 6.17. target Docking studies were performed for the four structures that presented the highest anti-inflammatory activity for each using human IKK-β and ERK1 structures obtained from the protein data bank (PDB) database.



1 Predicted pIC_{50} = 6.39 | **2** Predicted pIC_{50} = 6.38 | **3** Predicted pIC_{50} = 6.36 | **4** Predicted pIC_{50} = 6.33

Figure 2: BIAs with the highest anti-inflammatory activities for IKK-β

The two BIAs (**1** and **2**) with the highest anti-inflammatory activity for IKK-β present a structural similarity regarding the presence of methoxy and hydroxyl groups at positions 9 and 10, respectively (Figure 2) being almost equal in their predicted pIC_{50} values. Structure **3** presented a pIC_{50} value of 6.36, being structurally different with respect to **1** and **2**, having in its structure a glycosylation and no presence of the 1,3 benzodioxole moiety. In the same way, molecule **4** (pIC_{50} = 6.33) is less aromatic structure than **1** and **2**, with methyl groups in the 9 and 10 positions.

Molecular docking studies were performed to evaluate the interactions of these four molecules with the human IKK-β (PDB ID: 4kik). Figure 3 shows the less energetic poses for the best-ranked molecules, confirming the importance of the hydroxyl moiety in 9 and 10 positions for the inhibitory capacity of BIAs. This pharmacophore interacts with two residues (donor and acceptor of hydrogen bonds): lysine 44 and aspartate 166 (Figure 3A, B). Only in **4** is this double interaction not present due to methylation in the two positions, which may be related to the decrease of pIC_{50} value (Figure 3D). Lysine 44 is present in all cases evaluated, being considered a critical residue in the interaction of IKK-β with BIAs.

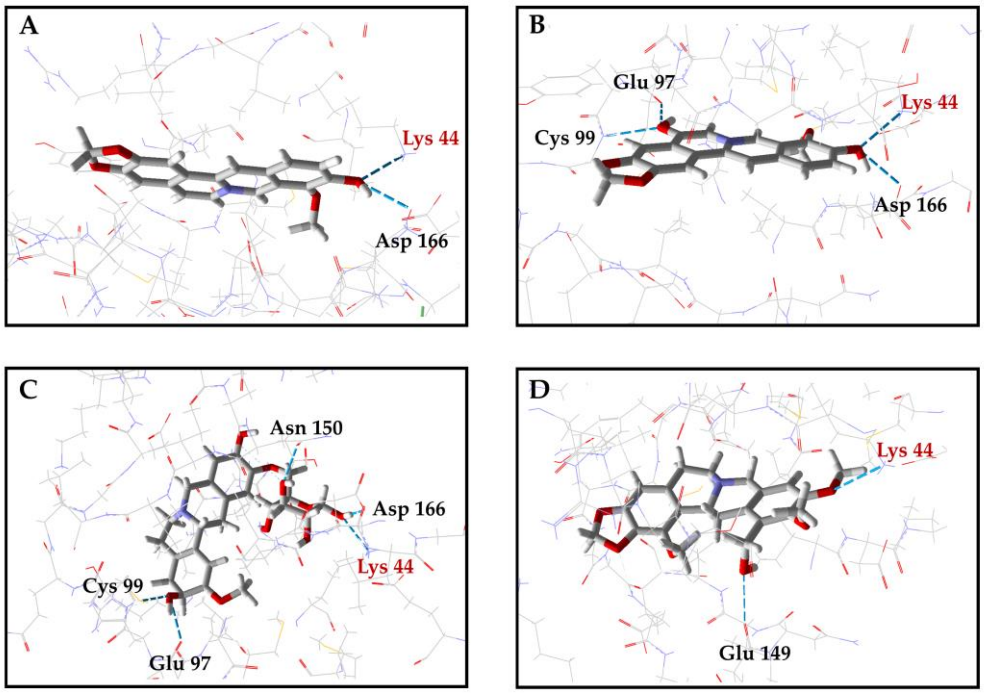
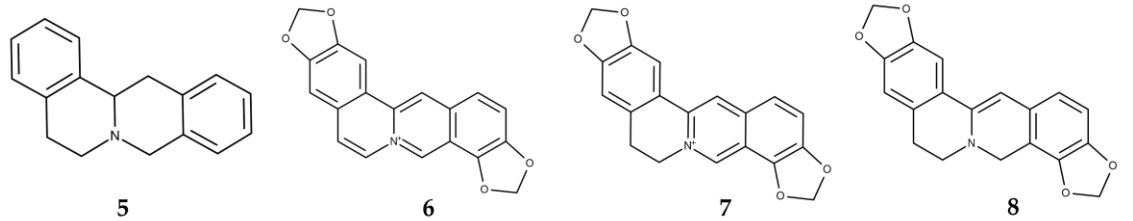


Figure 3: Molecular docking of the four BIAs with the highest anti-inflammatory activities for IKK-β. Structures A) **1**, B) **2**, C) **3** and D) **4**. Lys44, a common interaction residue, is highlighted in red.

The best-ranked molecules, **1** and **2**, (Figure 3A, B) have a similar spatial disposition in the pocket of the protein, these are very flat molecules compared with the other two. Interestingly, molecule **2** has a hydroxyl moiety in position 5 that allows the interaction of this with glutamate 97 and cysteine 99. These characteristic interactions also are observed in molecule **3** (Figure 3C), a more flexible molecule that has a different conformation in the IKK-β pocket. In this case, a hydroxyl group of glucose is interacting with Lys44 and Asp166, facilitating the hydrogen-bond (H-bond) interaction of the –OH group present in position 3 with the residues Glu97 and Cys99. These two H-bonds are also observed in the interaction of K-252A (PDB ID: ksa), a reported ligand for IKK-β, in which the carbonyl group of the 2-pyrrolidone ring makes a H-bond with Cys99, while nitrogen interacts with Glu97 [19].



5 Predicted pIC_{50} = 6.17 | **6** Predicted pIC_{50} = 6.13 | **7** Predicted pIC_{50} = 6.12 | **8** Predicted pIC_{50} = 6.11

Figure 4: BIAs with the highest anti-inflammatory activities for ERK1

Meanwhile, for ERK1, protoberberine skeleton **5** presents the highest predicted value of pIC_{50}, therefore, no structural modification regarding the core of this structure present in our database seems to increase the activity (Figure 4). Another three BIAs (**6**, **7** and **8**) with high structural similarity presented high predicted values of pIC_{50} between the 179 structures predicted by the QSAR models, with pIC_{50} values obtained being very close among them.

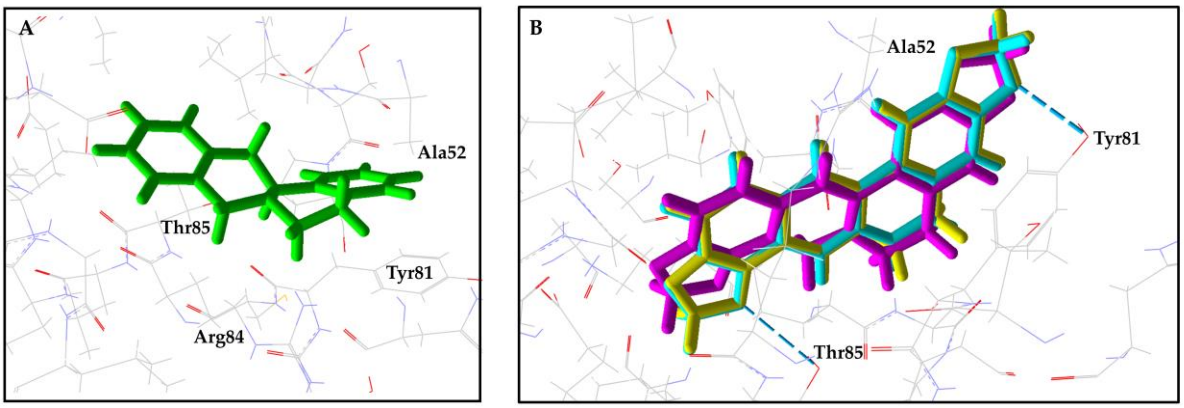


Figure 5: Molecular docking of the four BIAs with the highest anti-inflammatory activities for ERK1. A) Docking pose of protoberberine **5** in the pocket of ERK1. B) Docking orientations of structures **6** (blue), **7** (yellow) and **8** (pink) in ERK 1. Dashed lines in blue represent the H-bond interactions of **6** and **7** with Tyr81 and Thr85 of ERK1.

Subsequently, molecular docking of these four BIAs (Figure 4) was performed, using Molegro software. Figure 5 shows the less energetic poses for these molecules in the pocket of human ERK1 protein (PDB ID: 4qtb). For protoberberine **5** (Figure 5A), no H-bond interactions were observed, however, due to its nonpolar skeleton, steric and hydrophobic interactions can be related to the predicted activity, a steric interaction being highlighted with Ala52 of ERK1, which is present for the four BIAs as well as in the ligand (PDB ID: 38z) published as an inhibitor for this target [20].

For the structures **6**, **7** and **8**, inhibitory activity is related to the common core for these molecules, a high structural similarity is observed (Figure 5B) and the spatial arrangement of the three molecules is almost identical in the active site of human ERK1. Molecule **8**, is slightly more flexible than **6** and **7** because it is the least aromatic structure among them, a small modification in its orientation that prevents the interaction of the oxygen of the dioxolane ring (carbon 9) with the residue Thr85 of ERK1 (Figure 5B, pink) being observed. Finally, a common H-bond is observed for the oxygen of the dioxolane moiety (carbon 3) of these three structures with Tyr81, being important for the correct position of these structures in the ERK1 pocket.

3. Materials and Methods

4.1. Dataset

From the ChEMBL database, 775 and 48 structures with activity against IKK-β (ChEMBL1991) and ERK1 (ChEMBL3385), respectively, were obtained (https://www.ebi.ac.uk/chembl/). The compounds were classified using values of pIC_{50} (–log IC_{50}), presenting a range of 4.29 (from 5.01 to 9.30) for IKK-β and 3.10 (from 5.05 to 8.15) for ERK1. In this case, IC_{50} represents the concentration required for 50% inhibition of enzymatic activity. From smiles codes, 2-D structures were generated in Standardizer software that canonized structures, added hydrogens and performed aromatic form conversions [JChem 14.9.1.0, 2014; ChemAxon (http://www.chemaxon.com)].

4.2. Molecular descriptors

Using Dragon 7 software (Kode srl, Dragon (software for molecular descriptor calculation) version 7.0.4, 2016, https://chm.kode-solutions.net), 3,839 2-D molecular descriptors were calculated. For ERK1, 1,839 descriptors were primarily excluded (192- near constant, 1,627- constant and 20- one missing), meanwhile, for IKK-β 2,313 descriptors (51- near constant, 1,460- constant and 802- one missing). In the same way, correlated descriptors above 0.95 were also excluded, 665 descriptors for ERK1 and 1,062 descriptors, for IKK-β being finally exported. For the BIAs dataset, 2-D-descriptors

also were calculated and the same exclusion criteria were applied. Preliminarily, 2021 descriptors were eliminated, 614 descriptors being exported.

4.3. Random forest models

Descriptors were imported to Knime 3.1.0 software (www.knime.org) [21]. All variables were subjected to autoscaling and then were partitioned to generate two groups, a training group composed of 80% of the whole molecules set and a test group composed of the remaining 20%. (Q)SAR models were performed using an RF regression algorithm. The parameters selected for RF regression included 100 models, using a static random seed. Models were evaluated through cross-validation (leave-one-out), Q^2_{LOO}, as well as external test, Q^2_{ext}.

The applicability domain (APD) based on the Euclidean distances was used to identify compounds in the test set for which predictions may be unreliable if the values are higher than APD = $d + Z\sigma$, where d and σ are average Euclidian distance and standard deviation of the set of samples in the training set that have lower Euclidian distance average values of all samples in the training set. The parameter Z is an empirical cut off value, 0.5 was used as the default [22].

4.4. Molecular docking

The structure of human ERK1 (PDB ID: 4qtb) and IKK-β (PDB ID: 4kik) proteins with their respective ligands, (3*R*)-1-(2-oxo-2-{4-[4-(pyrimidin-2-yl)phenyl]piperazin-1-yl}ethyl)-*N*-[3-(pyridin-4-yl)-2H-indazol- 5-yl]pyrrolidine-3-carboxamide (PDB ID: 38z) and K252-A (PDB ID: ksa), were downloaded from the PDB [19,20]. All the water molecules were deleted from the enzyme structure, and the enzyme and compound structures were prepared using the same default parameter settings in the same software package (score function: Moldock score; ligand evaluation: internal ES internal Hbond, sp^2–sp^2 torsions, all checked; number of runs: 10 runs; algorithm: Moldock SE; maximum interactions: 1500; max. Population size: 50; max. Steps: 300; neighbor distance factor: 1.00; max. Number of poses returned: 5). The docking procedure was performed using a grid of 15 Å in radius and 0.30 Å in resolution to cover the ligand-binding site of the two enzymes' structures.

4. Conclusions

From a ligand-based model using RF and 2-D molecular descriptors, selected BIAs with methoxy and hydroxyl groups at positions 9 and 10 were determined as the most potentially active structures against IKK-β from an in-house database of BIAs. For ERK1, the RF model selected the protoberberine core without any substitution as the most potentially active compound.

Molecular docking analysis allowed the identification of key interactions in the active site for each of the two proteins studied: hydrogen bond between residue Lys44 of human IKK-β with hydroxyl group of best-ranked BIAs for this target, and a steric interaction present for the four most active BIAs for ERK1 with the residue Ala52; these two interactions may be related to the predicted anti-inflammatory activity for these structures.

Supplementary Materials: The following are available online at www.mdpi.com/link, Table S1: Molecules_BIAs_Supplementary_Material.xlsx

Acknowledgments: We would like to thank the CNPq and Capes for financial Support, and the Student Agreement Program of Graduate—PEC-PG of CNPq—Brazil.

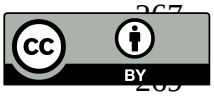
Author Contributions: L.S, M.F built database; C.H performed all calculus; and C.H, M.F, M.S and M.D wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Hunter, P. The inflammation theory of disease. *EMBO Rep* **2012**, *13*, 968-970. DOI: 10.1038/embor.2012.142.
2. Wu, S.; Yang, R.; Wang, G. Anti-asthmatic effect of pitavastatin through aerosol inhalation is associated with cd4+ cd25+ foxp3+ t cells in an asthma mouse model. *Sci Rep* **2017**, *7*. DOI: 10.1038/s41598-017-06476-6.
3. Rauber, S.; Luber, M.; Weber, S.; Maul, L.; Soare, A.; Wohlfahrt, T.; Lin, N.-Y.; Dietel, K.; Bozec, A.; Herrmann, M. Resolution of inflammation by interleukin-9-producing type 2 innate lymphoid cells. *Nat Med* **2017**, *23*, 938-944. DOI: 10.1038/nm.4373.
4. Libby, P. Inflammation in atherosclerosis. *Arterioscler, Thromb, Vasc Biol* **2012**, *32*, 2045-2051. DOI: 10.1161/hc0902.104353.
5. Khalaj, K.; Ahn, S.H.; Bidarimath, M.; Nasirzadeh, Y.; Singh, S.S.; Fazleabas, A.T.; Young, S.L.; Lessey, B.A.; Koti, M.; Tayade, C. A balancing act: Rna binding protein hur/ttp axis in endometriosis patients. *Sci Rep* **2017**, *7*, 5883. DOI: 10.1038/s41598-017-06081-7.
6. Wu, H.-M.; Fang, L.; Shen, Q.-Y.; Liu, R.-Y. Sp600125 promotes resolution of allergic airway inflammation via tlr9 in an ova-induced murine acute asthma model. *Mol Immunol* **2015**, *67*, 311-316. DOI: 10.1016/j.molimm.2015.06.016.
7. Verhein, K.C.; Salituro, F.G.; Ledeboer, M.W.; Fryer, A.D.; Jacoby, D.B. Dual p38/jnk mitogen activated protein kinase inhibitors prevent ozone-induced airway hyperreactivity in guinea pigs. *PloS one* **2013**, *8*, e75351. DOI: 10.1371/journal.pone.0075351.
8. Manetz, T.S.; Gonzalez-Espinosa, C.; Arudchandran, R.; Xirasagar, S.; Tybulewicz, V.; Rivera, J. Vav1 regulates phospholipase cγ activation and calcium responses in mast cells. *Mol Cell Biol* **2001**, *21*, 3763-3774. DOI: 10.1128/MCB.21.11.3763-3774.2001.
9. Jabril-Cuenod, B.; Zhang, C.; Scharenberg, A.M.; Paolini, R.; Numerof, R.; Beaven, M.A.; Kinet, J.-P. Syk-dependent phosphorylation of shc a potential link between fcεri and the ras/mitogen-activated protein kinase signaling pathway through sos and grb2. *J Biol Chem* **1996**, *271*, 16268-16272. DOI: 10.1074/jbc.271.27.16268.
10. Rossman, K.L.; Der, C.J.; Sondek, J. Gef means go: Turning on rho gtpases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* **2005**, *6*, 167-180. DOI: 10.1038/nrm1587.
11. Christen, R.; Lovern, M.; Clifford, A.M.; Yue, J.-X.; Goss, G.G.; Pouysségur, J.; Lenormand, P. Erk1 and erk2 present functional redundancy in tetrapods despite higher evolution rate of erk1. *BMC Evol Biol* **2015**, *15*, 179. DOI: 10.1186/s12862-015-0450-x.
12. Gilfillan, A.M.; Tkaczyk, C. Integrated signalling pathways for mast-cell activation. *Nat Rev Immunol* **2006**, *6*, 218-230. DOI: 10.1038/nri1782.
13. Marquardt, D.L.; Walker, L.L. Dependence of mast cell ige-mediated cytokine production on nuclear factor-κb activity. *J Allergy Clin Immunol* **2000**, *105*, 500-505. DOI: 10.1067/mai.2000.104942.
14. Senegas, A.; Gautheron, J.; Maurin, A.G.D.; Courtois, G. Ikk-related genetic diseases: Probing nf-κb functions in humans and other matters. *Cell Mol Life Sci* **2015**, *72*, 1275-1287. DOI: 10.1007/s00018-014-1793-y.
15. Li, T.; Wong, V.K.W.; Jiang, Z.H.; Jiang, S.P.; Liu, Y.; Wang, T.Y.; Yao, X.J.; Su, X.H.; Yan, F.G.; Liu, J. Mutation of cysteine 46 in IKK-βeta increases inflammatory responses. *Oncotarget* **2015**, *6*, 31805-31819. DOI: 10.18632/oncotarget.5567.
16. Alves, M.; Scotti, M.; Scotti, L.; Mendonça, J.F.; de Melo, S.; Dos Santos, S.; de Fátima, F.M.D.M. Secondary metabolites from cissampelos, a possible source for new leads with anti-inflammatory activity. *Curr Med Chem* **2016**, *24*, 1629-1644. DOI: 10.2174/0929867323666161227123411.
17. Beaudoin, G.A.; Facchini, P.J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **2014**, *240*, 19-32. DOI: 10.1007/s00425-014-2056-8.
18. Hawkins, K.M.; Smolke, C.D. Production of benzylisoquinoline alkaloids in saccharomyces cerevisiae. *Nat Chem Biol* **2008**, *4*, 564-573. DOI: 10.1038/nchembio.105.
19. Liu, S.; Misquitta, Y.R.; Olland, A.; Johnson, M.A.; Kelleher, K.S.; Kriz, R.; Lin, L.L.; Stahl, M.; Mosyak, L. Crystal structure of a human iκb kinase β asymmetric dimer. *J Biol Chem* **2013**, *288*, 22758-22767. DOI: 10.1074/jbc.M113.482596.
20. Chaikuad, A.; Tacconi, E.M.; Zimmer, J.; Liang, Y.; Gray, N.S.; Tarsounas, M.; Knapp, S. A unique inhibitor binding site in erk1/2 is associated with slow binding kinetics. *Nat Chem Biol* **2014**, *10*, 853-860. DOI: 10.1038/nchembio.1629.
21. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. Knime-the konstanz information miner: Version 2.0 and beyond. *SIGKDD Explor* **2009**, *11*, 26-31. DOI: 10.1145/1656274.1656280.

22. Scotti, M.T.; Scotti, L.; Ishiki, H.M.; Peron, L.M.; de Rezende, L.; do Amaral, A.T. Variable-selection approaches to generate qsar models for a set of antichagasic semicarbazones and analogues. *Chemom Intell Lab Syst* **2016**, *154*, 137-149. DOI: 10.1016/j.chemolab.2016.03.023.



Cell-based platforms and *in silico* methodologies for discovery of new drugs

Mateus Feitosa Alves[1], Chonny Herrera Acevedo[1], Margareth de Fátima Formiga Melo Diniz[1], Marcus Tullius Scotti[1], Luciana Scotti[1]*

[1]Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil

*Corresponding author: Luciana Scotti, Program in Natural and Synthetic Bioactive at Federal University of Paraíba, 58051-900 João Pessoa, Paraíba, Brazil; Tel.: +55-83-99869-0415; E-mail: luciana.scotti@gmail.com

ABSTRACT

Introduction: Cell-based assays have emerged as an important alternative for the development of new drugs due to multiple difficulties presented by current treatments against different diseases such as resistance, high toxicity, low effectiveness among others, as well as the fact that highly active molecules have been obtained in the laboratory tests, many of them have not presented any positive results in patients.

Areas covered: This review describes the cell-based assays: High-throughput screening threading (HTS) and high-content screening (HCS), highlighting the NCI60 platform and some studies performed in this area using several methodologies, looking for promissory molecules against cancer, bacteria, neglected tropical diseases and Alzheimer's disease.

Expert opinion: The authors suggest that cell-based assays despite their multiples successful results still is a drug-design area in constant development which can be improved due to their huge potential. For this, *in silico* studies are proposed a promising complementary area, since it possesses multiple advantages, which would help to achieve more robustness results.

Keywords: Cell-based platforms, high-throughput screening, high-content screening, NCI60, lead compounds, drug discovery, *in silico*.

Article Highlights

- Cell-based platforms have emerged as an important alternative for the development of new drugs due among many others to its versatility and capacity to predict clinical efficacy of active molecules.
- HTS and HCS are two of the main cell-based methodologies which have allowed the discovery of new molecules with activity against several diseases.
- The NCI60, a cell-based platform developed in 1990's has been a fundamental tool in this research area and is still used to have the largest database of test compounds in cells for assessing anticancer activity.
- Some studies of cell-based assays have resulted in anticancer, anti-neglected tropical diseases (NTD), antibacterial and anti-Alzheimer promissory molecules.
- *In silico* approaches are powerful tools that complement HTS and HCS studies in order to get robust knowledge about lead activities.

1. INTRODUCTION

Cell-based assays have been a key tool in the development of new drugs since, among their many advantages, they allow us to evaluate the interaction of the drug and its target, which facilitates the understanding of related biological processes and the prediction of clinical efficacy of new agents [1, 2]. Biological and technological advances have allowed for a growth in this type of experiment, with the revolution of molecular biology as a determinant factor in this development [3, 4, 5]. For example, the search for new anticancer drugs has evolved in the last six decades from the use of animal cell cultures in the 1950s to the historic platform National Cancer Institute 60 (NCI60) in the 1990s, which allowed the discovery and development of many anticancer therapeutic products [6, 7], and more modern developments in the 21st century such as the Center for Molecular Therapeutics 1000 (CMT1000) which uses more than 1200 cell lines [2].

High-throughput screening (HTS) is cell-based methodologies for discovering new drugs by identifying substances [8], for the technology and test methodology are constantly evolving [9]; A multidisciplinary effort has been required to implement these two platforms (chemists, biologists, engineers, computer scientists) [10]. Among the main features of these cell-based assays are the wide variety of detection forms that have been linked, in particular: radiolabeling,

surface plasmon resonance (SPR), fluorescent ligands, mass spectrometry (MS), among many others [1, 11], these examples decrease the cost by reducing the equipment needs and allowing for a wider array of experiments [12].

A large number of results produced by HTS, which are sometimes considered underutilized, may be employed using *in silico* methodologies that aid not only in the interpretation of the data, but may also be used to predict the activities of new compounds, may be used in agrochemicals, pharmacology, toxicity, among others [13, 14, 15].

In this context, this review presents a description of the HTS and HCS (High-content screening) platforms, commonly used in cell-based assays for drug development and there are several examples of these platforms that can be clustered into four main groups: anticancer (including NCI60 as fundamental tool in this research area), anti-neglected tropical diseases (NTD), antibacterial and anti-Alzheimer. This demonstrates the great diversity of methodologies recently used, as well as a set of promissory molecules found from these studies.

2. CELL-BASED PLATFORMS

High-throughput screening (HTS) is a scientific experimentation method used for drug discovery. This method uses robotics, data processing and control software, liquid handling devices, and sensitive detectors [16]. HTS allows a researcher to quickly conduct millions of chemical, genetic, or pharmacological tests. Through this process, it is possible to readily identify active compounds, antibodies or genes that modulate a particular biomolecular pathway [17]. The results of these experiments provide starting points for drug design and for the understanding of the interaction or role of a particular biochemical process in biology [18]. Successful replication of this technique requires devices that require minimal experience to operate, are amenable to many different cell-based assays, can easily generate quantifiable readings, and are inexpensive to manufacture and operate [12].

One of the key foundations for the operation of HTS equipment is the microliter plate, a disposable plastic container having 384, 1536 or 3456 wells, all multiples of 96. The wells may contain test items as well as, cells or enzymes and substances that can be used as experimental controls. A library should be cataloged with the content contained on the boards in order to avoid experimental errors [19].

The HTS technique is suitable for the application of machine learning methods for the development of a model that predicts *in silico* compound activity bringing benefits to the pharmaceutical industry, beyond other areas (figure 1). The US Environmental Protection

Agency [6] conducts studies involving HTS in order to generate biological data generating models that predict compound toxicity [20]. Techniques such as Support Vector Machine (SVM), Neural Networks (NN), Classification And Regression Tree (CART), Linear Discriminant Analysis [9], Partial Least Square (PLS), and k-Nearest Neighbor (kNN) have been presented with high sensitivity, with SVM being the most outstanding [21].

HTS may involve a multidisciplinary field involving other techniques bringing advantages to the discovery of new drugs. The Nuclear Magnetic Resonance (NMR) technique applied to HTS has generated important structure-activity information acquired from the NMR analysis, which can potentiate the library design, providing important information from the HTS. Another technique that may be associated with HTS are fluorescence-based assays which include the scintillation proximity assay (SPA), time resolved energy transfer, fluorescence anisotropy, fluorescence correlation spectroscopy, and fluorescence flotation spectroscopy. These techniques have enormous benefits because they are simple and quick to implement [19, 22, 23, 24].

HTS had a major boost following the completion of the Human Genome Project because of the significant increase in targets of therapeutic interest. From this, it was possible to conduct successful research using the discovery of new drugs. The drug Maraviroc (Pfizer), a chemokine receptor antagonist, is an example that HTS, together with medicinal chemistry, can be used to optimize pharmacological and pharmacokinetic properties. Optimization of the molecule improved parameters (binding potency against the receptor, antiviral activity and absorption, and pharmacokinetics) that were responsible for the selection of Maraviroc for clinical trials among the 1000 molecules analyzed [25].

Through HTS, inhibitors of the Hepatitis C virus (HCV) NS5A were discovered because of the ability of this technique to identify compounds that modulate biological activity without the need for prior knowledge of the mechanism of action or the nature of the targets [25]. From the optimization, we selected the molecule BMS-790052 for clinical trials by exhibiting enhanced antiviral potency, extending inhibitory activity to encompass the HCV 1a genotype and to optimize oral bioavailability and sustained pharmacokinetic properties [25, 26].

HCS is also a platform for discovering new drugs by identifying substances like genes, proteins, and other biomolecules in normal and abnormal cells [8]. The term "high content" is used to refer to processes that are spatially and temporally defined in the context of the structural and functional integrity of each individual cell within an array of cells, and can evaluate

individual cells, organelle morphology, and complex phenotypes, allowing the evaluation of cell subpopulations and multiple simultaneous measurements by cells [27, 28].

The differences between HTS and HCS are small, since HCS shares many basic techniques with high throughput screening. The main difference between these techniques is that the HCS obtains multiple readings from the cellular analysis. In summary, HTS involves a part of biological or biochemical mechanisms, while HCS analyzes the change in the whole cell phenotype [29].

This technique is widely used in the preclinical phase of new studies for the discovery of new drugs, as well as predicting *in vivo* toxicity. As a result, it is used to support clinical trials [30]. Compared to HTS, HCS has additional advantages by providing an evaluation of effects such as cytotoxicity or fluorescence of test compounds [31].

HCS uses automated screening based on cellular morphology images allowing the identification of small compounds that alter cellular phenotypes and modify the cellular function. Selection of molecules based on a cellular phenotype does not require prior knowledge of the biochemical targets that are affected by the compounds [32]. For this reason, a limiting step for HCS is target identification [33].

In some cases, HCS has been called high content analysis (HCA), high content imaging (HCI) or image cytometry (IC) [31]. Image-based HTS assays require solutions for image segmentation and coding of cell compartments to track events, such as those reported by signaling pathways and protein translocations [34] This feature can speed up the training process, successfully discover rare phenotypes, and improve the accuracy of the analysis [35]. Images of cellular events can be used to monitor, through HTS, the activation of the co-receptor recipients of a G-protein or other receptors, the internalization in a cell or changes in its quantity facilitating the understanding of targets of future drugs [36]. The image-based HCS assay succeeded in identifying that the alkaloid ellipticine increased the nuclear localization of endogenous p53 in HCT116 colon cancer cells with a resulting increase in transactivation of the p21 promoter [37]. Researchers showed that the vast majority (60% to 80%) of the studies utilizing HCS used univariate analysis tools and only one or two image-based features in the analysis. However, advanced data analysis methods that allow multiparametric data will allow the HCS to reach its potential, turning more powerful implementations [38].

All HCS assays produce toxicity results, since this technique measures physiological responses of a cell to the stimulus, whether environmental or chemical. From cell counts to

checking how healthy the cell is, HCS can help define the toxicity status. In *in vitro* toxicity studies, HCS showed a high sensitivity and specificity, with a low false positive rate [39, 40, 41]. The results of HCS have a global acceptance and, from them, the toxicity of various compounds, nanoparticles and toxins of fungi in food can be evaluated [42].

As a popular technique in the area of oncology, HCS has emerged as a point of interest in the field of neurobiology in the search of new drugs for the treatment of diseases such as Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, and brain cancer. The use of HTS allows new molecules to be tested in cultures of neuronal cells in order to know their rate of growth and neuronal protection [43].

3. APPLICATIONS

3.1 Anticancer

The development of effective drugs for cancer has difficulties due to the few preclinical models present for cancer [44, 45]. Human cell lines have been widely used, although there are some controversies among researchers. However, these models continue to be considered as the best tools for the identification of new drugs [2].

New technologies have been developed representing new progress. With these techniques, it is possible to create autonomous devices that are easy to use [12]. The National Cancer Institute (USA) has developed a platform with 60 human tumor cell lines (NCI60), with nine types of cancer represented: melanoma, lung cancer, colon, brain, ovary, prostate, kidney, among others [2, 46]. From this platform, it was possible to identify the broad chemotherapeutic resistance of several cell lines in NCI60 was strongly correlated to the expression of multidrug resistance [2, 47]

NCI 60 allows us to show cell line sensitivity profiles to provide information on the mechanisms of growth inhibition and cell death from the analysis of compounds. These, in turn, produced similar patterns of activity leading to the observation and development of data analysis through the mean graph, which represents a sensitivity profile through the analysis of 50% growth inhibition (GI_{50}), total growth inhibition (TGI) and 50% lethal concentration (LC_{50}), as well as the "COMPARE" algorithm that compares these profiles with all or parts of the screening database [48]. With this, new anticancer agents were identified by the NCI60 137 screening program [47].

In the 1990s, through NCI60, the first class of compounds, *ellipticinium* derivatives, that had selectivity for central nervous system (CNS) tumors were selected [48, 49]. With the growth of research in natural products and the development of the NCI60, a collection of data about natural products was cataloged [48]. In 2005, approximately 88,000 pure compounds and 34,000 crude extracts were tested on NCI60 [2, 50]. Compounds were also isolated, for example, *salicylahalamides* which were obtained from a crude extract producing a unique profile on NCI60 and shown to function as an inhibitor of vacuolar ATPase [48, 51].

The NCI60 platform can be used for the purpose of capturing a profile of anticancer compounds to predict patterns of inhibition of growth of compounds by interleaving a cell line profile as well as chemical information. Recently, Cortés-Ciriano et al. have shown that predicted drug associations and growth inhibition patterns are largely consistent with experimental data, which also suggests the possibility of identifying genomic drug sensitivity markers for novel compounds in novel cell lines [52].

Despite its restricted use, the NCI60 is still used tool with the largest database of test compounds in cells for assessing anticancer activity [53, 54, 55, 56].

Otherwise, using HTS, nineteen compounds are identified as inhibitors of breast cancer stem cells through a cell-based screening of more than 2500 compounds in spheres of malignant human breast gland-derived cells (HMLER-shEcad cells). HMLER-shEcad cells are incubated for three days with all compounds of the library at concentration of 10 µM, after which, cell viability was tested. The authors found 152 compounds which reduce 50% of cell-viability, and identified three groups of the most active compounds. The first one is composed of benztropine and deptropine, two tropine-derivatives. The second group have two steroids, adrenosteron and tomatidine, a steroidal alkaloid of spirosolane type. Ketoconazole and aripiprazole, which contain a piperazine core as a common structural characteristic, constitute the third group. Results of the performed *in vitro* inhibition tests performed showed an IC_{50} value for benztropine mesylate **1** (Figure 2)**,** close to 5 µM, as well through *in vivo* tests which were observed as this compound reduces the frequency of cancer stem cells and inhibits the tumor-initiating [57].

Research shows that HTS based on phenotypic cells are used to identify small molecules that inhibit cell apoptosis and, in addition, structure-activity relationship studies can be used to identify active compounds for *in vivo* testing [58]. In a genomic approach, compounds were initially selected post HTS proving to be effective at the known target. Second, they were

predicted through an *in silico* approach of the 14 candidates 8 showed antiproliferative activity against glioblastoma: ivermectin, trifluridine, astemizole, amlodipine, maprotiline, apomorphine, mometasone and nortriptyline. The seven best compounds in the *in silico* evaluation was positive and when compared to HTS corresponded to a twenty times greater enrichment [59].

SPiDER (self-organizing map–based prediction of drug equivalence relationships) is a prediction algorithm which allows the prediction of macromolecular targets through a self-organizing maps (SOMs) consensus [60], establishing in this case the potential mechanism of action for benztropine mesylate. From molecular representations and based on the Euclidian distance between reference ligands and interesting molecules it is possible to calculate the probability in function of the distance to generate a final score which is the sum of the results of the SOM algorithm for each representation [61]. The ranking generated by this *in silico* tool for benztropine mesylate demonstrates the possible importance of neurotransmitter receptors for inhibitory effects of benztropine on breast cancer stem cells, since nicotinic, acetylcholine receptor agonist, histamine receptors H1 & H2, and the muscarinic acetylcholine receptor antagonist presented low values of probability *p* between 0.005 and 0.023 [57].

Mitogen activated protein kinases (MAPKs) originally called extracellular signal-regulated kinases (ERKs) are proteins which are related to cellular differentiation, movement, division, and cell death [62, 63, 64]. The MAP/ERK kinase pathway is a target of interest in cancer treatment, since a relationship exists between this pathway and tumor development. A mammalian cell-based assay was performed with the objective to inhibit this signalization way, through the development of a new high throughput model. Using rat chondrosarcoma (RCS) cells, two Mitogen-activated protein kinase kinase (MEK) inhibitors were tested and it was found that only one of these **2** (Figure 2), can inhibit the ERK pathway at 20 μM. In the same way, a fibroblast growth factor receptors (FGFR)-inhibitor was able to inhibit successfully this way at low concentrations (< 5 μM), concluding that this new platform is effective in the search of promissory molecules with antitumoral activity [65].

From PTC Library, which has more than 150,000 small molecules, indolic alkaloid, 6-bromo-1-(1H-pyrrol-1-yl)-2,3,4,9-tetrahydro-1H-carbazole [66] and (5R,11aR)-2-cyclohexyl-5-(4-methoxyphenyl)-3-thioxo-2,3,5,6,11,11a-hexahydro-1H-imidazo[1',5':1,6]pyrido[3,4-b]indol-1-one (**4)** (Figure 2) have emerged as potent inhibitors of hypoxia induced by the VEGF, an important factor in tumor growth being that increased levels of VEGF are found in

almost all common solid tumors [67]. Situations found in tumor tissues, such as hypoxia, can lead to increased levels of this protein leading to angiogenesis, supporting tumor growth. Pharmacological properties have been improved more than 200-fold with respect to **4** through molecular modeling. Compound **5** (a bromine-derivate) is a better inhibitor of VEFG with an IC_{50} value of about of 6 nM (Figure 2), whereas **4** has an IC_{50} value of 681 nM [68].

Destruction of tumor cells by different agents is in many cases mediated by programmed cell death or apoptosis in target cells [9]. Caspase-3 is a protease which is active during this process and is commonly used as a biosensor in the search of new anticancer agents. Tian et al. used a cyan fluorescent protein – yellow fluorescent protein (CFP-YFP) pair, and performed HTS based on the fluorescence resonance energy transfer (FRET) methodology in the search of potential anticancer molecules. The HeLa cell line – C3, which contains a Capsase-3 biosensor, was cultured in 96-well plates within several different compounds, and CFP and YPF emission signals were monitored at specific times. A decrease in the YFP/CFP ratio (an effect of FRET) is observed in these compounds by action of Caspase-3, indicating apoptosis in HeLa cells and the compounds can therefore be considered as lead molecules [69].

Initially, it was observed that HeLa-C3 cells treated with anticancer drugs (paclitaxel, vincristine, etoposide, and hydroxyurea) present a significant reduction in the YFP/CFP ratio. After, a screen performed with compounds of *Podophyllum* at three different concentrations (10, 20 and 40 nM) showed an apoptotic induction of Podophyllotoxin (**6)** at 20 nM being 5000-fold lower in concentration with respect to etoposide reaching a similar effect (Figure 2). Etoposide was used as the control due to its structural similarity, since the molecule also was obtain from the same genus [69].

3.2 Neglected tropical diseases

Neglected tropical diseases (NTDs) affect one billion people, mainly those living in poverty conditions from tropical and subtropical regions of more than 150 countries, causing approximately 142,400 deaths per year [66, 70]. One of the main problems of control and elimination of these diseases are the chemotherapies treatments used, due to problems such as low effectiveness, high toxicity, resistance, among others [71]. In this context, many approximations have been used in drug design looking for alternative drugs against NTDs [72, 73, 74, 75]. The following describes some examples of cell-based assays, which have provided molecules with promissory activity.

The viral genome of Dengue consists of a positive-sense RNA of 11kb, which encodes for 3 structural and 7 non-structural proteins. The pre-membrane protein (PrM) is one of the three proteins which have an important role for proper folding and secretion of viral envelope. Cleavage of PrM by host enzymes is fundamental in the virus maturation process. Then, PrM blockage is an interesting approximation in the search of new drugs against *Flavivirus* [76].

Stolp et al. performed a cell-based assay monitoring of PrM cleavage using HA and FLAG antibodies and flow cytometry. After, the experiment was miniaturized in 96-well plates, followed by a screening with a compound library. The Prestwick Chemical Library® was used for HTS and fluorescence was observed when using anti-FLAG-FITC as a cutoff parameter. From the 1,249 small molecules tested (31 were discarded by autofluorescent or toxicity), three were selected as potential drugs due to their blocking cleavage capacity of PrM (anti-FLAG-FITC fluorescence values above of 15%). Two of the three selected compounds are not specific against Dengue virus PrM processing, and also affect the HIV envelope, where only a specific compound is desired, identified as thiostrepton **7** (Figure 2), a thiopeptide antibiotic produced by *Streptomyces* [77].

Thus, thiostrepton emerged as a lead for the Dengue virus through this approximation. Its dose-response curve was obtained as well as its half-maximal inhibition concentration (IC_{50}) using percentage fluorescence and staining with anti-FLAG-FITC, which was analyzed by flow cytometry and a predicted value of 4.94 was obtained through nonlinear regression.[77].

Huang et al., using Baby hamster kidney-21 (BHK-21), designed a microfluidic platform for evaluating in real-time, anti-dengue effect of carrageenan (sulfated polysaccharides) **8** (Figure 2), monitoring cellular oxygen consumption rates (OCRs), an important parameter to evaluate the metabolic activity of a virus. Using three groups of studies, the pre-treated, virucide group (cells infected simultaneously treated with carrageenan) and post-treated group, authors initially found that after 10 h of treatment, carrageenan is not cytotoxic against BHK-21 since at different concentrations of polysaccharide variations in the OCRs are not observed. Following this, by using the platform the OCR of BHK-21 cells was measured with or without pre-treatment with 1 mg/mL of carrageenan, and it was determined that carrageenan presents an inhibitory effect on the virus infection. This inhibitory effect is stronger in the virucide group (82%) compared to the pre-treatment group and post-treatment group (70% and 50%, respectively). The effect of carrageenan on the Dengue virus is proportional to the dose used to inhibit the virus infection within 7 h in all studied groups [78].

A library of 2157 compounds was used to perform a HTS to find growth inhibitory compounds of *L. amazonensis* promastigotes. Auranofin **9** (Figure 2), a gold(I) coordination complex used as a treatment for rheumatoid arthritis [79] presented an effective concentration (EC50) of 0.30 µM, being lower than amphotericin B, (EC_{50} = 0.38 µM), an antileishmanial agent used as a control. In the same way, this compound turns out to be less cytotoxic with respect to amphotericin B; in an assay performed with CellToxTM® Green, another two antileishmanial agents, sodium stibogluconate and paromomycin, are less effective in growth inhibition of promastigotes and are not cytotoxic [80].

Cell-based amastigotes assays using RAW 264.7 macrophage cell line were performed and results show that for both *L. major* and *L. amazonensis* auranofin (EC_{50} = 0.07 µM and 0.27 µM) has an effect similar to that which is observed with amphotericin B (EC_{50} = 0.03 µM and 0.03 µM), being in all cases more potent than paromomycin and sodium stibogluconate, with a highly selective index. Caspase 3/7-like activity is observed in promastigotes treated with **9** (0-50 µM), relating its antileishmanial activity with the apoptotic process. Finally, *in vivo* tests show that auranofin produced a significant reduction in the lesion area as well as partial lesion suppression. Using a murine model, this compound emerges as an interesting lead in the search for new antileishmanial agents [80].

Yeast cell–based systems are promissory technologies for the development of new antiparasitic drugs, which are flexible systems that can assure a wide variety of targets allowing a robust and cost-effective HTS test [81]. Bilsland et al. performed a primary automated HTS from a library of 14400 compounds using a fluorescent-yeast system (*S. cerevisiae*), with the aim to obtain new antiparasitic leads. The inhibitory capacity of a set of molecules was evaluated against three fundamental parasitic enzymes (dihydrofolate reductases, N-myristoyltransferases, and phosphoglycerate kinases) from different parasites (*P. falciparum, P. vivax, T. cruzi, T. brucei,* and *L. major*) selected as targets [82].

For the validation process 36 hits compounds were selected to evaluate their activity against yeast strains codifying target enzymes. Using 10 µM of *T. brucei* Lister 427 bloodstream form parasites, the authors found that 50% of molecules were able to kill the parasite; furthermore, another 5 hits were able to drastically inhibit the parasite growth. *In silico* validations were also performed based on structural similarity. Initially a model was created where structures were codified using a fingerprint. From this, the Tanimoto index was calculated and hierarchical clusters were generated comparing structures with their activities.

Moreover, clusters were analyzed in relation to their structural complexity, using a maximal common subgraph (MCS), and comparing the library used by the authors with the ChemBL databank to verify the hits. Results show that only the *P. vivax* PKG (phosphoglycerate kinases) was a hit with a Tanimoto distance coefficients (TDC) ≤ 0.5, and a great number of hits were obtained with promissory activity against this parasitic disease [82].

Another approximation is the search of novel mosquitocidal toxins being safe to humans. Using *Anopheles, Aedes, Drosophila,* and three human cells lines, a cell-based platform was developed and through screening of 200,000 compounds targeted mosquito-specific cytotoxins were identified. Results show three lead compounds (**10**, **11** and **12 -** Figure 2) that the results show to be cytotoxic against the 4A3A cell line of *Anopheles* with IC_{50} values of 1.42, 2.86 and 14.6 µM without cytotoxicity against TPH1 (IC_{50} values above of 120 µM) and HCT116 (IC_{50} values above of 35 µM) human cell lines, as well as for *Drosophila* cell lines (Kc and S2R+) used in this study. In the same way, the whole animal lethality of compound **10** was evaluated using mosquito larva, showing a killing percentage close to 80% for *Anopheles* larvae, meanwhile for *Aedes* and *Drosophila* the values were found to be less than 20%. As a result, compound **10** is an interesting selective lead and the platform emerges as a new approximation for the vectors control [83].

3.3 Antibacterial

Antibiotic resistance is actually considered to be one of the major problems in human health [84, 85]. An urgent need is arising to develop new antibacterial drugs to face these problematic. For this, libraries of natural, recombinant or chemically synthesized compounds have been tested against bacterial targets by several approximations [86], such as cell-based assays and promissory antibacterial molecules have emerged.

An example of this, is the study performed by Farha et al., where a sequential HTS of 1600 compounds searched for inhibitors of bacterial cell wall biogenesis using *S. aureus*. The study found that clomiphene **13** (Figure 2), a non-steroidal fertility medicine derivate of triphenylethylene, appears as lead for the development of antibacterial drugs. Through mechanistic characterization, the authors concluded that undecaprenyl diphosphate synthase (UppS) is the target of clomiphene. In the same way, a dose–response curve of clomiphene against recombinant UppS was performed and a non-linear regression of this curve showed an IC_{50} value of 7.5 µM [87].

UppS is an interesting target for the development of new antibacterial drugs due to its importance in the biosynthesis of bacterial cell walls, catalyzing process to obtain undecaprenyl diphosphate (UPP) [88, 89]. The same group performed another HTS, now with an initial set of 142,000 small molecules from different libraries, with the aim to find molecules with an inhibitory effect on UppS. From the initial set, five compounds were selected which were active against gram-positive *B. subtilis*, as well as a result antagonist of targocil (a late-stage inhibitor of wall teichoic acid synthesis) both in liquid and solid medium. The five compounds demonstrated *in vitro* inhibition of the UppS recombinant of *B. subtilis* and dose-dependence inhibition of *B. subtilis*, *S. aureus*, and *E. coli* UppS [90].

Compound **14** (Figure 2) presents the lowest IC_{50} value against *B. subtilis and S. aureus*, 0.05 and 1.60 µM, respectively, meanwhile compound **15** (Figure 2), is the most active molecule against *E. coli* (IC_{50} value = 1.39 µM) having also activity against *S. aureus* and *B. subtilis* (IC_{50}= 39.9 and 6.22 µM), emerging as interesting antibacterial leads through the blocking of UppS [90].

Recently, using cell-based kinetic dose response curves, MacNair, et al. performed a HTS looking for antibacterial compounds. The *E. coli* K-12 strain were used for dose response assays, defining a minimal inhibition concentration (MIC) for all tested compounds. Growth curves of each antimicrobial agent were constructed in relation of its MIC, being chosen three concentrations (half- quarter and eighth- MIC). From the ten characteristics obtained from the growth curves, a linear discriminant analysis [9] was performed which generates a fingerprint of the antimicrobial agents. Tested molecules were clustered according their mechanism of action (50S ribosome, 30S ribosome, DNA replication, cell wall and others) [91].

From structure-activity relationship (SAR) studies and knowing that antibiotics generate a unique kinetic dose response it is possible to predict off-target antibacterial effects. Analogues of 2-[2-nitrophenyl] acetohydrazide (a biotin A inhibitor) **16**, with an off-target antimicrobial effect were identified through the SAR study, using as a variable the absence or presence of biotin. Principal component analysis (PCA) quantified the growth curve changes between **16** and their analogues. At 1µM, **16** (Figure 2) presented 100% inhibition, meanwhile analogs structurally similar with chlorine or methyl replacing the nitro group reached an inhibition percentage of 63 and 81% respectively, being relating antibacterial properties with the presence of benzene ring [91].

3.4 Alzheimer's disease

Alzheimer's disease is a neurodegenerative disorder which is linked to two proteins: amyloid β (Aβ) in senile plaques and tau protein aggregation in neurofibrillary tangles, as well as neuronal loss [92, 93]. Cleavage of the amyloid-β precursor protein [94] by the BACE 1 (beta-site APP cleaving enzyme-1, β-secretase) and γ-secretase is a fundamental event in this proteopathy.

BACE-1 inhibition is fundamental in the treatment of Alzheimer's disease, since it limits the quantity of amyloid β peptide, preventing the cleavage process. A cell-based study with flavonols (myricetin, quercetin, kaempherol, and morin) and flavones (apigenin) as BACE-1 non-peptidic inhibitors was performed, and promissory molecules were found for the future development of therapies against Alzheimer's disease [95].

Initially, it was observed that all of the tested compounds inhibited BACE-1 with IC_{50} values from 2.8 to 38.5 μM, with myricetin **17** (2.8 μM) being the most active compound (Figure 2). The cell-based assay using a fetal rat cerebral cortex (E18) and β-secretase activity assay kit, revelated a significant reduction in the BACE-1 activity by two flavonols (myricetin, quercetin **18** – Figure 2). Docking studies show that myricetin form two hydrogen bonds to Asp32 and Try198 with hydroxyl groups presents in the C and B ring respectively, as well as an interaction of Gln73 with the A ring of this flavonol through hydrogen bonds. The docking score was found to be 4.65, the highest value with respect to the other studied compounds, which have fewer numbers of interactions with BACE-1: kaempherol did not interact with Gln73, Morin had a slightly different pose and interacted with Asp228 via the B ring, and apigenin did not interact with asparagine residues. Based in these results and knowing that flavonoids did not affect neuronal cell viability, myricetin emerges as an interesting lead for the development of drugs for the treatment of Alzheimer's disease [95].

Calcium channels of the P/Q type may be involved in a number of neurological diseases, including Alzheimer's disease, where from the blockage of these channels the current flow can be normalized leading to relief of symptoms [94, 96, 97]. To that end, researches are being developed involving the use of Fluorescence-Based HTS with blockers of these channels. For this research, the HEK293 cell line, which stably expresses the P/Q-type calcium channel alpha 1A subunit under the control of a tetracycline promoter, was used. HTS containing a library of 150,000 compounds led to the identification of 3262 accesses that succeeded in inhibiting the fluorescence signal. From the use of Hit-to-Lead (HTL), 12.400 analogues were identified. In

the end, 27 specimens of different chemotypes were selected and validated, and could serve as a new therapeutic source for Alzheimer's disease and other neurological diseases [94].

Amyloid precursor protein [94] is important in Alzheimer's disease because from its cleavage the C-terminal 31-amino acid APP-C31 toxic peptide is released along with the production of Delta C31 APP [98, 99]. Animal studies have already shown that the APP mutation prevents this type of caspase cleavage by improving cognitive development. Poksay et al. (2017) demonstrated that compounds screened in HTS were able to inhibit the cleavage of APP-D720 and APP Delta C31 [99]. In another study, HTS was used to identify drugs that protect the integrity of the blood-brain barrier, since it maintains brain homeostasis and prevents the entry of toxins and drugs. Alzheimer's disease plays an important role in disturbing the blood-brain barrier [100]. From the HTS analysis it was possible to identify 50 compounds that had a potential to maintain membrane integrity. Of these, seven were drugs already registered in the FDS, where five of them were able to maintain membrane integrity when tested on human endothelial cells hCMEC/D3. Finally, the drugs etodolac, granisetron and beclomethasone presented the most promising effect as a protector of the blood-brain barrier [101].

HCS is also widely used in cases of discovery of new drugs that act on CNS lesions and neurodegenerative diseases [102] . Through the HCA method, 743 plasmids encoding developmentally regulated genes were tested in neurite growth assays using postnatal cortical neurons where growth inhibitors and growth enhancers were identified through a small participation via GTPase signaling [103].

4. CONCLUSIONS

In this review, the principal cell-based assays, HTS and HCS, were described. These methodologies have been highlighted as interesting alternatives for drug discovery for several diseases and the methods hold advantages in the search of promissory molecules against different diseases such as cancer, leishmaniasis, dengue, among others.

Various examples have been presented here for different diseases and targets, exposing the importance and variety of these types of studies. In particular, HTS has contributed a great number of lead compounds, designing cell-based platforms based in a great number of techniques which range from fluorescence to *in silico* tools such as SOMs. Similarly, we showed the main characteristics and applications of NCI60, one of the most important cell-based platforms for anticancer drug development.

Therefore, from the multiple promissory results obtained and the advantages offered by these methodologies, cell-based assays continue to be an important topic of research and it is assumed that through support from other knowledge areas, will continue to develop more alternative treatments against a great number of diseases.

5. EXPERT OPINION

Cell-based platforms are used for testing large numbers of molecules with potential pharmacological activity. In addition, they are used to know target-drug interactions in order to evaluate predictions that serve as a basis for preclinical study, reducing the use of animals. One of the main characteristics of these methodologies is the versatility. As described in the present review a wide variety of detection methods have emerged as well as an increase in the number of cell lines (for cancer, toxicity, viral diseases, among others), that can be evaluated in the same assay. In the same way, *in silico* studies have been aided in predicting data from active molecules as well as in confirming the results obtained by cell-based assays, giving more robustness of the results.

However, even greater potential of cell-based platforms, there are some difficulties that still limit their application, as the difficulty to optimize the function of the nature of the complex of the hits as well as a low coverage of the chemical space of the molecules. In addition, cell-based platforms have a high cost having to be handled by qualified professionals in the kind of studies.

Thus, in this area of the drug-discovery exists a need for the frequent and intensive use of data banks of structures and biological activity to obtain structural patterns with potential biological activity as well as the search of algorithms for obtaining information from bulky and complex data banks. In the same way, for better results, it is important to consider the combination of High Throughput Screening (HTS) studies with virtual screening, results of in vitro studies with in silico studies for prediction of bioactive structures.

In addition to some studies with structure-based design using data from the Protein Data Bank (PDB) receiver structure through docking, there are no deeper studies in this area. The authors suggest that in addition to these, it is fundamental to perform molecular dynamics studies (despite their higher computational cost), which have already been used in other systems of studies in medicinal chemistry, but still shy for studies linked to cell-based platforms.

The advantages of computational methodologies are low cost, standardization, minimum equipment requirements and short execution time. It has been verified that there are potentially

large methodologies using ligand-based design combining with Structure-based and cell-based associated in the same study. The databases used by the cell-based platform may be essential in assisting in the execution of *in silico* studies. Likewise, as in several areas of medicinal chemistry, for example, in the synthesis of bioactive compounds, identification and isolation of secondary metabolites and studies of prediction of pharmacological properties such as absorption, distribution, metabolism, excretion and toxicity (ADMET), it is possible to perceive that *in silico* studies may decrease the costs associated with assays based on cell platforms

Moreover, cell-based assays are already consolidated in research of new active compounds for negligible, bacterial and anticancer diseases. However, there are few studies related to neurodegenerative diseases, such as Alzheimer's models, which should be further investigated for faster and more robust search for new biologically active compounds in this area. Additionally, Alzheimer's models have known targets and it is known that the effect of the inhibition of these enzymes to this disease obtains effective biological responses which are compatible with this approach.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank the CNPq and Capes for financial Support, and the Student Agreement Program of Graduate—PEC-PG of CNPq—Brazil.

REFERENCES

1. Fang Y. Ligand–receptor interaction platforms and their applications for drug discovery. Expert opinion on drug discovery. 2012;7(10):969-988.
2. Sharma SV, Haber DA, Settleman J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nature reviews cancer. 2010;10(4):241-253.
3. Schenone M, Dančík V, Wagner BK, et al. Target identification and mechanism of action in chemical biology and drug discovery. Nature chemical biology. 2013;9(4):232-240.
4. Buckley K, Ryder AG. Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review. Applied spectroscopy. 2017;71(6):1085-1116.
5. Kaisar MA, Sajja RK, Prasad S, et al. New experimental models of the blood-brain barrier for CNS drug discovery. Expert opinion on drug discovery. 2017;12(1):89-103.

6. Grever MR, Schepartz SA, Chabner BA, editors. The National Cancer Institute: cancer drug discovery and development program. Seminars in oncology; 1992: WB SAUNDERS CO INDEPENDENCE SQUARE WEST CURTIS CENTER, STE 300, PHILADELPHIA, PA 19106-3399.
7. Hendriks HR, Govaerts A-S, Fichtner I, et al. Pharmacologically directed strategies in academic anticancer drug discovery based on the European NCI compounds initiative. British Journal of Cancer. 2017.
8. Giuliano KA, Haskins JR, Taylor DL. Advances in high content screening for drug discovery. Assay and drug development technologies. 2003;1(4):565-577.
9. Fulda S, Debatin K-M. Caspase activation in cancer therapy. 2013.
10. Macarrón R, Hertzberg RP. Design and implementation of high throughput screening assays. Molecular biotechnology. 2011;47(3):270-285.
11. Rohman M, Wingfield J. High-Throughput Screening Using Mass Spectrometry within Drug Discovery. High Throughput Screening: Methods and Protocols. 2016:47-63.
12. Kwon CH, Wheeldon I, Kachouie NN, et al. Drug-eluting microarrays for cell-based screening of chemical-induced apoptosis. Analytical chemistry. 2011;83(11):4118.
13. Bender A. Bayesian methods in virtual screening and chemical biology. Chemoinformatics and Computational Chemical Biology. 2011:175-196.
14. Wicht KJ, Combrinck JM, Smith PJ, et al. Bayesian models trained with HTS data for predicting β-haematin inhibition and in vitro antimalarial activity. Bioorganic & medicinal chemistry. 2015;23(16):5210-5217.
15. Chiddarwar RK, Rohrer SG, Wolf A, et al. In silico target prediction for elucidating the mode of action of herbicides including prospective validation. Journal of Molecular Graphics and Modelling. 2017;71:70-79.
16. Liebert M. Roundtable Discussion: High-Throughput Screening Challenges. Genetic Engineering & Biotechnology News. 2008;28(14):26-27.
17. Caraus I, Alsuwailem AA, Nadon R, et al. Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions. Briefings in bioinformatics. 2015:bbv004.
18. Zhang XD. Optimal high-throughput screening: practical experimental design and data analysis for genome-scale RNAi research. Cambridge University Press; 2011.
19. Liu B, Li S, Hu J. Technological advances in high-throughput screening. American Journal of Pharmacogenomics. 2004;4(4):263-276.
20. Zang Q, Mansouri K, Williams AJ, et al. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. Journal of chemical information and modeling. 2017;57(1):36-49.
21. Mballo C, Makarenkov V. Using machine learning methods to predict experimental high throughput screening data. Combinatorial chemistry & high throughput screening. 2010;13(5):430-441.
22. Fernandes PB. Technological advances in high-throughput screening. Current opinion in chemical biology. 1998;2(5):597-603.

23. Fouda A, Tahsini M, Khodayarian F, et al. A Fluorescence-based Lymphocyte Assay Suitable for High-throughput Screening of Small Molecules. JoVE (Journal of Visualized Experiments). 2017 (121):e55199-e55199.
24. Meleza C, Thomasson B, Ramachandran C, et al. Development of a scintillation proximity binding assay for high-throughput screening of hematopoietic prostaglandin D 2 synthase. Analytical biochemistry. 2016;511:17-23.
25. Macarron R, Banks MN, Bojanic D, et al. Impact of high-throughput screening in biomedical research. Nature reviews Drug discovery. 2011;10(3):188-195.
26. Gao M, Nettles RE, Belema M, et al. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature. 2010;465(7294):96-100.
27. Abraham VC, Taylor DL, Haskins JR. High content screening applied to large-scale cell biology. Trends in biotechnology. 2004;22(1):15-22.
28. Tolosa L, Gómez-Lechón MJ, Donato MT. High-content screening technology for studying drug-induced hepatotoxicity in cell models. Archives of toxicology. 2015;89(7):1007-1022.
29. Bellomo F, Medina D, De Leo E, et al. High-content drug screening for rare diseases. Journal of Inherited Metabolic Disease. 2017:1-7.
30. Nichols A. High content screening as a screening tool in drug discovery. High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery. 2006:379-387.
31. Buchser W, Collins M, Garyantes T, et al. Assay development guidelines for image-based high content screening, high content analysis and high content imaging. 2014.
32. Fraietta I, Gasparri F. The development of high-content screening (HCS) technology and its importance to drug discovery. Expert opinion on drug discovery. 2016;11(5):501-514.
33. Eggert US, Mitchison TJ. Small molecule screening by imaging. Current opinion in chemical biology. 2006;10(3):232-237.
34. Edward R. Use of DNA-specific anthraquinone dyes to directly reveal cytoplasmic and nuclear boundaries in live and fixed cells. Molecules and cells. 2009;27(4):391-396.
35. Piccinini F, Balassa T, Szkalisity A, et al. Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. Cell Systems. 2017;4(6):651-655. e5.
36. Milligan G. High-content assays for ligand regulation of G-protein-coupled receptors. Drug discovery today. 2003;8(13):579-585.
37. Xu GW, Mawji IA, Macrae CJ, et al. A high-content chemical screen identifies ellipticine as a modulator of p53 nuclear localization. Apoptosis. 2008;13(3):413-422.
38. Singh S, Carpenter AE, Genovesio A. Increasing the content of high-content screening an overview. Journal of biomolecular screening. 2014;19(5):640-650.
39. Dykens JA, Jamieson JD, Marroquin LD, et al. In vitro assessment of mitochondrial dysfunction and cytotoxicity of nefazodone, trazodone, and buspirone. Toxicological Sciences. 2008;103(2):335-345.
40. Bova MP, Tam D, McMahon G, et al. Troglitazone induces a rapid drop of mitochondrial membrane potential in liver HepG2 cells. Toxicology letters. 2005;155(1):41-50.

41. Domingos MC, Davies AM, O'brien PJ. Application of High-Content Analysis in Clinical Cytology for Translational Safety Biomarkers of Drug-Induced Toxicity for Lymphoma Chemotherapy. Basic & Clinical Pharmacology & Toxicology. 2014;115(1):145-153.
42. Boncler M, Golanski J, Lukasiak M, et al. A new approach for the assessment of the toxicity of polyphenol-rich compounds with the use of high content screening analysis. PloS one. 2017;12(6):e0180022.
43. Zock JM. Applications of high content screening in life science research. Combinatorial chemistry & high throughput screening. 2009;12(9):870-876.
44. Aggarwal BB, Danda D, Gupta S, et al. Models for prevention and treatment of cancer: problems vs promises. Biochemical pharmacology. 2009;78(9):1083-1094.
45. Dragu DL, Necula LG, Bleotu C, et al. Therapies targeting cancer stem cells: Current trends and future challenges. World journal of stem cells. 2015;7(9):1185.
46. Stinson S, Alley M, Kopp W, et al. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. Anticancer research. 1991;12(4):1035-1053.
47. Alvarez M, Paull K, Monks A, et al. Generation of a drug resistance profile by quantitation of mdr-1/P-glycoprotein in the cell lines of the National Cancer Institute Anticancer Drug Screen. Journal of Clinical Investigation. 1995;95(5):2205.
48. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nature Reviews Cancer. 2006;6(10):813-823.
49. Acton EM, Narayanan VL, Risbood PA, et al. Anticancer specificity of some ellipticinium salts against human brain tumors in vitro. Journal of medicinal chemistry. 1994;37(14):2185-2189.
50. Holbeck S. Update on NCI in vitro drug screen utilities. European Journal of Cancer. 2004;40(6):785-793.
51. Boyd MR, Farina C, Belfiore P, et al. Discovery of a novel antitumor benzolactone enamide class that selectively inhibits mammalian vacuolar-type (H+)-ATPases. Journal of Pharmacology and Experimental Therapeutics. 2001;297(1):114-120.
52. Cortés-Ciriano I, van Westen GJ, Bouvier G, et al. Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. Bioinformatics. 2016;32(1):85-95.
53. Nogales V, Moutinho C, Martinez-Cardús A, et al. Discovery of biomarkers for antitumor drug resistance using 450K methylation data of NCI60. AACR; 2014.
54. Abaan OD, Davis SR, Bilke S, et al. The exomes of the NCI60 and their implications for cancer pharmacogenomics. AACR; 2012.
55. Nguyen D, Tomaszewski JE, Hewitt SM, et al. CD40 expression in human solid tumors and NCI60 tumor cell lines. AACR; 2013.
56. Strittmatter N, Lovrics A, Sessler J, et al. Shotgun Lipidomic Profiling of the NCI60 Cell Line Panel Using Rapid Evaporative Ionization Mass Spectrometry. Analytical Chemistry. 2016;88(15):7507-7514.
57. Cui J, Hollmén M, Li L, et al. New use of an old drug: inhibition of breast cancer stem cells by benztropine mesylate. Oncotarget. 2017;8(1):1007.

58. Xiong CS, Drewe J, Kasibhatla S. A chemical genetics approach for the discovery of apoptosis inducers: from phenotypic cell based HTS assay and structure-activity relationship studies, to identification of potential anticancer agents and molecular targets. Current medicinal chemistry. 2006;13(22):2627-2644.
59. Lee H, Kang S, Kim W. Drug repositioning for cancer therapy based on large-scale drug-induced transcriptional signatures. PloS one. 2016;11(3):e0150460.
60. Kohonen T. The self-organizing map. Neurocomputing. 1998;21(1):1-6.
61. Reker D, Rodrigues T, Schneider P, et al. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proceedings of the National Academy of Sciences. 2014;111(11):4067-4072.
62. Schaeffer HJ, Weber MJ. Mitogen-activated protein kinases: specific messages from ubiquitous messengers. Molecular and cellular biology. 1999;19(4):2435-2444.
63. Severin S, Ghevaert C, Mazharian A. The mitogen-activated protein kinase signaling pathways: role in megakaryocyte differentiation. Journal of Thrombosis and Haemostasis. 2010;8(1):17-26.
64. Mazalouskas MD, Godoy-Ruiz R, Weber DJ, et al. Small G proteins Rac1 and Ras regulate serine/threonine protein phosphatase 5 (PP5)· extracellular signal-regulated kinase (ERK) complexes involved in the feedback regulation of Raf1. Journal of Biological Chemistry. 2014;289(7):4219-4232.
65. Krejci P, Pejchalova K, Wilcox WR. Simple, mammalian cell-based assay for identification of inhibitors of the Erk MAP kinase pathway. Investigational new drugs. 2007;25(4):391-395.
66. Collaborators GMaCoD. Global, regional, and national age–sex specifi c all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;385(9963):117-171.
67. Ferrara N. VEGF as a therapeutic target in cancer. Oncology. 2005;69(Suppl. 3):11-16.
68. Cao L, Weetall M, Bombard J, et al. Discovery of Novel Small Molecule Inhibitors of VEGF Expression in Tumor Cells Using a Cell-Based High Throughput Screening Platform. PloS one. 2016;11(12):e0168366.
69. Tian H, Ip L, Luo H, et al. A high throughput drug screen based on fluorescence resonance energy transfer (FRET) for anticancer activity of compounds from herbal medicine. British journal of pharmacology. 2007;150(3):321-334.
70. WHO. Neglected Tropical Diseases [cited 2016 2016/09/10]. Available from: http://www.who.int/neglected_diseases/diseases/en/
71. Herrera Acevedo C, Scotti L, Feitosa Alves M, et al. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. Molecules. 2017;22(1):79.
72. Scotti L, Tullius Scotti M, Barros Silva V, et al. Chemometric studies on potential larvicidal compounds against Aedes aegypti. Medicinal Chemistry. 2014;10(2):201-210.
73. Ribeiro FF, Junior F, da Silva MS, et al. Computational and Investigative Study of Flavonoids Active Against Typanosoma cruzi and Leishmania spp. Natural product communications. 2015;10(6):917-920.
74. Scotti L, JBM Filho F, O de Moura R, et al. Multi-Target Drugs for Neglected Diseases. Current pharmaceutical design. 2016;22(21):3135-3163.

75. Tullius Scotti M, Scotti L, Ishiki H, et al. Natural Products as a Source for Antileishmanial and Antitrypanosomal Agents. Combinatorial Chemistry & High Throughput Screening. 2016;19(7):537-553.
76. Gao F, Duan X, Lu X, et al. Novel binding between pre-membrane protein and claudin-1 is required for efficient dengue virus entry. Biochemical and biophysical research communications. 2010;391(1):952-957.
77. Stolp ZD, Smurthwaite CA, Reed C, et al. A multiplexed cell-based assay for the identification of modulators of pre-membrane processing as a target against dengue virus. Journal of biomolecular screening. 2015:1087057115571247.
78. Huang S-H, Lin Y-S, Wu C-W, et al. Assessment of the inhibition of Dengue virus infection by carrageenan via real-time monitoring of cellular oxygen consumption rates within a microfluidic device. Biomicrofluidics. 2014;8(2):024110.
79. Shaw CF. Gold-based therapeutic agents. Chemical reviews. 1999;99(9):2589-2600.
80. Sharlow ER, Leimgruber S, Murray S, et al. Auranofin is an apoptosis-simulating agent with in vitro and in vivo anti-leishmanial activity. ACS chemical biology. 2013;9(3):663-672.
81. Denny P, Steel P. Yeast as a potential vehicle for neglected tropical disease drug discovery. Journal of biomolecular screening. 2015;20(1):56-63.
82. Bilsland E, Sparkes A, Williams K, et al. Yeast-based automated high-throughput screens to identify anti-parasitic lead compounds. Open biology. 2013;3(2):120158.
83. O'Neal MA, Posner BA, Coates CJ, et al. A cell-based screening platform identifies novel mosquitocidal toxins. Journal of biomolecular screening. 2013;18(6):688-694.
84. Sharma VK, Johnson N, Cizmas L, et al. A review of the influence of treatment strategies on antibiotic resistant bacteria and antibiotic resistance genes. Chemosphere. 2016 5//;150:702-714. doi: http://doi.org/10.1016/j.chemosphere.2015.12.084.
85. Hiltunen T, Virta M, Laine A-L. Antibiotic resistance in the wild: an eco-evolutionary perspective. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372(1712). doi: 10.1098/rstb.2016.0039.
86. Coates AR, Hu Y. New strategies for antibacterial drug design: targeting non-multiplying latent bacteria. Drugs R D 2006 20060606 DCOM- 20060727;7(3):133-151. eng.
87. Farha MA, Czarny TL, Myers CL, et al. Antagonism screen for inhibitors of bacterial cell wall biogenesis uncovers an inhibitor of undecaprenyl diphosphate synthase. Proceedings of the National Academy of Sciences. 2015;112(35):11048-11053.
88. Fujihashi M, Zhang Y-W, Higuchi Y, et al. Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. Proceedings of the National Academy of Sciences. 2001;98(8):4337-4342.
89. Concha N, Huang J, Bai X, et al. Discovery and characterization of a class of pyrazole inhibitors of bacterial undecaprenyl pyrophosphate synthase. Journal of medicinal chemistry. 2016;59(15):7299-7304.
90. Czarny TL, Brown ED. A Small-Molecule Screening Platform for the Discovery of Inhibitors of Undecaprenyl Diphosphate Synthase. ACS Infectious Diseases. 2016;2(7):489-499.

91. MacNair CR, Stokes JM, French S, et al. A cell-based approach to characterize antimicrobial compounds through kinetic dose response. Bioorganic & Medicinal Chemistry. 2016;24(24):6315-6319.
92. Caraci F, Leggio GM, Salomone S, et al. New drugs in psychiatry: focus on new pharmacological targets. F1000Research. 2017;6.
93. Jucker M, Walker LC. Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. Annals of neurology. 2011;70(4):532-540.
94. Mezler M, Hermann D, M Swensen A, et al. Development and validation of a fluorescence-based HTS assay for the identification of P/Q-type calcium channel blockers. Combinatorial chemistry & high throughput screening. 2012;15(5):372-385.
95. Shimmyo Y, Kihara T, Akaike A, et al. Flavonols and flavones as BACE-1 inhibitors: structure–activity relationship in cell-free, cell-based and in silico studies reveal novel pharmacophore features. Biochimica et Biophysica Acta (BBA)-General Subjects. 2008;1780(5):819-825.
96. MacManus A, Ramsden M, Murray M, et al. Enhancement of 45Ca2+ Influx and Voltage-dependent Ca2+ Channel Activity by β-Amyloid-(1–40) in Rat Cortical Synaptosomes and Cultured Cortical Neurons MODULATION BY THE PROINFLAMMATORY CYTOKINE INTERLEUKIN-1β. Journal of Biological Chemistry. 2000;275(7):4713-4718.
97. Ramsden M, Henderson Z, Pearson HA. Modulation of Ca 2+ channel currents in primary cultures of rat cortical neurones by amyloid β protein (1–40) is dependent on solubility status. Brain research. 2002;956(2):254-261.
98. Banwait S, Galvan V, Zhang J, et al. C-terminal cleavage of the amyloid-β protein precursor at Asp664: a switch associated with Alzheimer's disease. Journal of Alzheimer's Disease. 2008;13(1):1-16.
99. Poksay KS, Sheffler DJ, Spilman P, et al. Screening for small molecule inhibitors of statin-induced APP C-terminal toxic fragment production. Frontiers in pharmacology. 2017;8.
100. Bell RD, Zlokovic BV. Neurovascular mechanisms and blood–brain barrier disorder in Alzheimer's disease. Acta neuropathologica. 2009;118(1):103-113.
101. Qosa H, Mohamed LA, Al Rihani SB, et al. High-throughput screening for identification of blood-brain barrier integrity enhancers: a drug repurposing opportunity to rectify vascular amyloid toxicity. Journal of Alzheimer's Disease. 2016;53(4):1499-1516.
102. Cooper DJ, Zunino G, Bixby JL, et al. Phenotypic screening with primary neurons to identify drug targets for regeneration and degeneration. Molecular and Cellular Neuroscience. 2017;80:161-169.
103. Blackmore MG, Moore DL, Smith RP, et al. High content screening of cortical neurons identifies novel regulators of axon growth. Molecular and Cellular Neuroscience. 2010;44(1):43-54.

FIGURE CAPTIONS

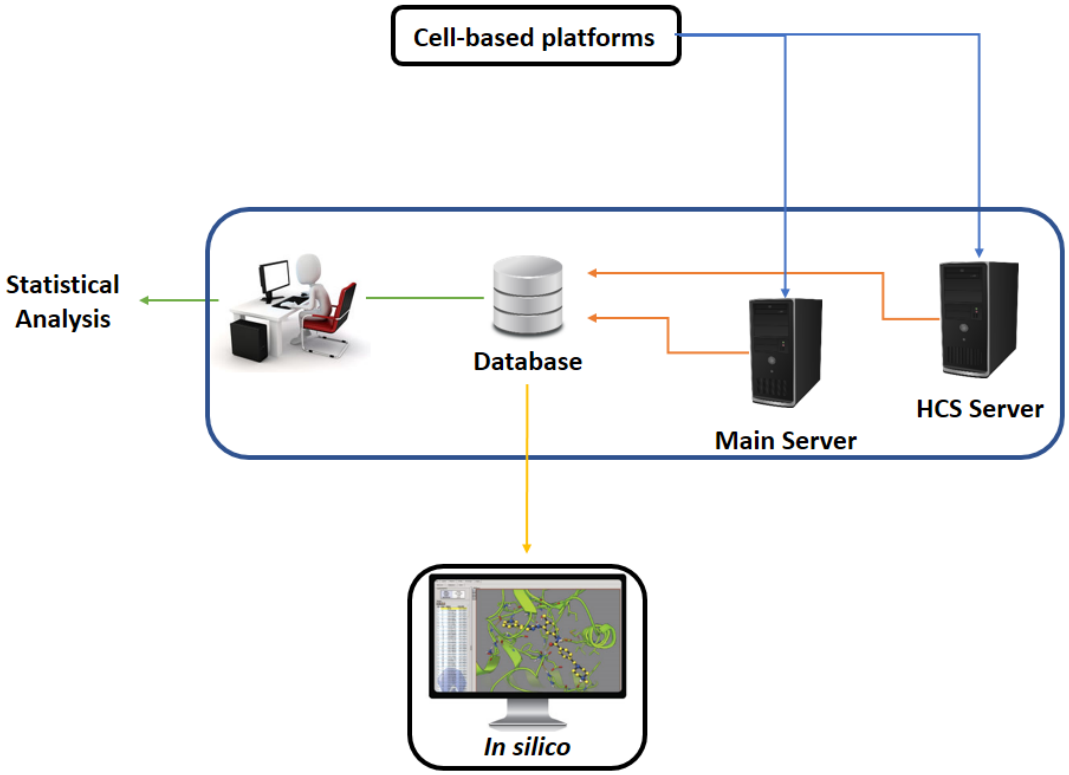


Figure 1: HTS data used as input data with in silico approaches to predict biological activities, minimizing costs.

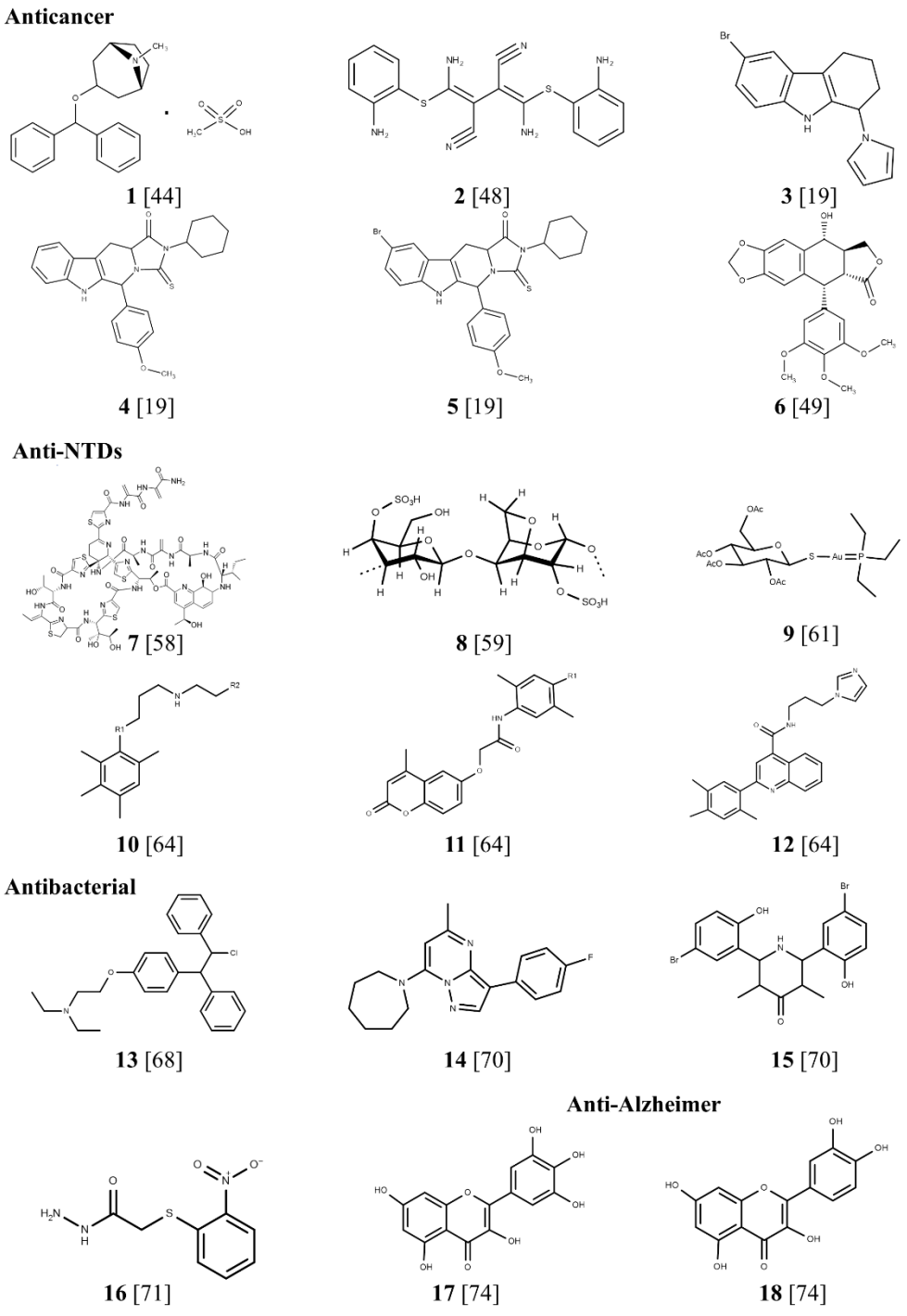


Figure 2: Structures with promissory activity against several types of diseases obtained from Cell-based assays. **1-6**: Anticancer, **7-12**: Anti-NTDs, **13-16**: Antibacterial and **17-18**: Anti-Alzheimer.

SciForum
MOL2NET

Ligand-based virtual screening of a benzylisoquinoline alkaloids dataset with anti-inflammatory potential activity.

Chonny Herrera Acevedo[1], Luciana Scotti[1], Mateus Feitosa Alves[1], Margareth F.F.M. Diniz[1] and Marcus Tullius Scotti[1, *]

[1] Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil; caherrera@ltf.ufpb.br (C.H) ; luciana.scotti@gmail.com (L.S) ; mateusfalves@gmail.com (M.F); margareth@ltf.ufpb.br (M.D) ; mtscotti@gmail.com (M.S)
***** Correspondence: mtscotti@gmail.com; Tel.: +55-83-99869-0415

Received: / Accepted: / Published:

Abstract:

Inhibitor of nuclear factor kappa B kinase beta subunit (IKK-B) and extracellular signal-regulated kinase 1 (ERK1) are two proteins involved in cytokine intracellular signaling pathways, which have a great importance due to their anti-inflammatory role. In this work, from the ChemBL database were obtained 775 and 48 structures with activity against IKK-B (CHEMBL1991) and ERK1 (CHEMBL3385) respectively. The compounds were classified using values of $pIC_{50,}$ presenting a range of 4.29 (from 5.01 to 9.30) for IKK-B and 3.10 (From 5.05 to 8.15) for ERK-1. From SMILES codes, two-dimensional (2D) structures were generated in Standardizer and after calculated 1064 two-dimensional molecular descriptors in Dragon 7 software. Obtained results were imported to Knime 3.1.0 software. All variables were submitted to autoscaling and after were partitioned to generate two groups, a training group composed by the 80% of the whole molecules set and a test group composed by the remaining 20%. (Q)SAR models was performed using a Random Forest (RF) algorithm. Models were evaluated through cross validation (leave-one-out), Q^2_{LOO} = 0.69 and 0.66 as well as external test, Q^2_{ext} = 0.74 and 0.58 for IKK-B and ERK1 respectively. Finally, pIC50 value of 179 benzylisoquinoline alkaloids were predicted in the (Q)SAR models found 4 compound with the highest activity for each one protein studied.

Keywords: benzylisoquinoline alkaloids, IKK-B, ERK1, anti-inflammatory activity, Virtual screening

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*
YouTube link: *please, paste here the link to your personal YouTube video, if any.*

1. Introduction

The inflammatory process is a nonspecific complex, stereotype, coordinated response of tissues to injury [1]. Several proteins with specific roles are present in the signaling pathways involved in these processes.

Nuclear factor-κB (NF-κB), is a transcriptional factor, which plays a key role in numerous physiological, these include inflammatory processes [2]. NF-κB activation is stimulated by a kinase complex, IκB kinase (IKK), which is composed of three core proteins: IKK1/ IKK-α, IKK2/IKK-β and NEMO/IKK-γ [2,3], being IKK-α and IKK-β two catalytic subunits which are structurally related kinases.

IKK-β, is an interesting target due to its role in the inflammation-induced tumour growth and progression, as well as an important modulator of tumour surveillance and rejection [4].

In turn, extracellular signal-regulated kinase 1 (ERK1) being one of the two isoforms of ERK described. It is present in the Ras/Raf/MEK/ERK signaling pathway, which has vital importance, since many essential cell processes are involved. In ~30 % of cancers and cognitive disorders, exists an abnormally activation [5].

ERK1 is a serine/threonine kinase of the GMGC group that plays a critical role in the regulation of cell growth and differentiation [6].

Benzylisoquinoline alkaloids (BIA) are metabolites which present a great diversity (~2,500 BIAs are known today) and several pharmacological activities such as antimicrobial agent, muscle relaxant, and potential anticancer drug, among others [7,8].

In this work, through of a random forest model using 2D molecular descriptors was performed to in order to predict the anti-inflammatory activity 179 BIAs structures.

2. Results and Discussion

Ligand-based virtual screening

(Q)SAR models was performed using a Random Forest algorithm. Models were evaluated through cross validation (leave-one-out), $Q^2_{LOO} = 0.69$ and 0.66 as well as external test, $Q^2_{ext} = 0.74$ and 0.58 for IKK-B and ERK1 respectively (see Figure 1 and 2)

After, pIC50 value of 179 benzylisoquinoline alkaloids were predicted in the (Q)SAR models. For IKK-B and ERK1, four structures presented the highest anti-inflammatory activity for each target (Table 2).

The two BIAs with highest anti-inflammatory activity for IKK-B present a structural similarity regarding to the presence of methoxyl and hydroxyl groups at the positions 9 and 10 respectively (Table 2a).

Meanwhile for ERK1, protoberberine skeleton presents the highest predicted value of pIC50, therefore, none structural modification regarding the core of this structure present in our database seem to increase the activity (Table 2b).

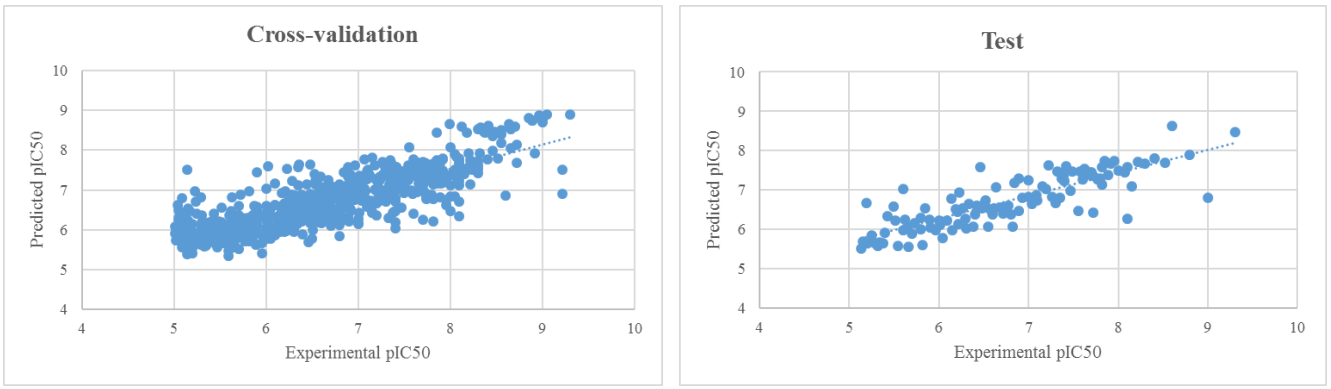


Figure 1. Plot of experimental vs predicted values of pIC50 for cross-validation and test set generated by IKK-B QSAR model

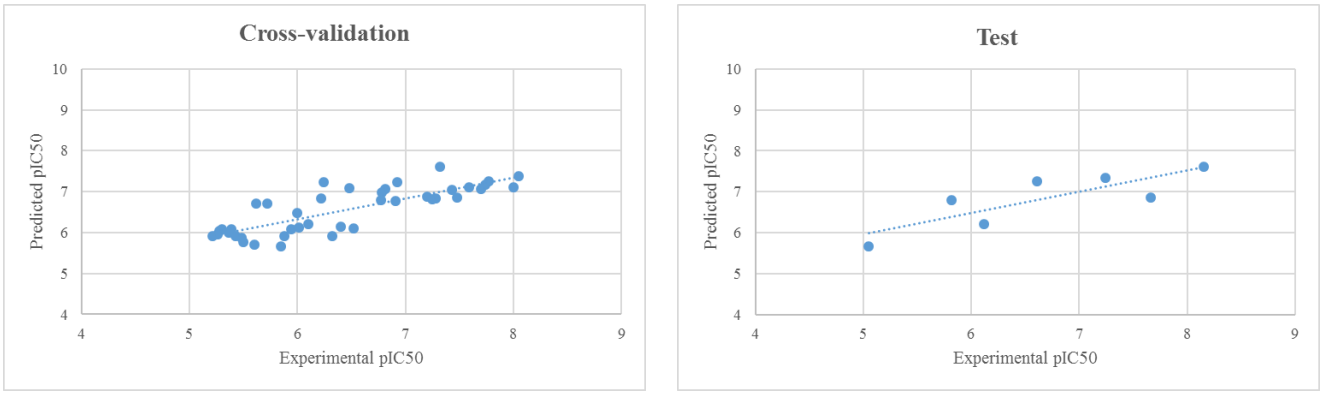
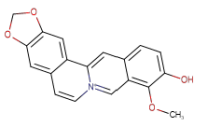
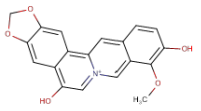
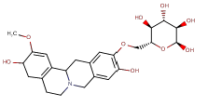
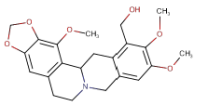
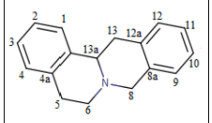
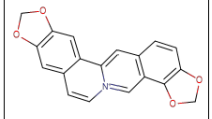
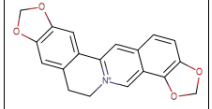
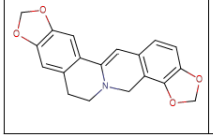


Figure 2. Plot of experimental vs predicted values of pIC50 for cross-validation and test set generated by ERK1 QSAR model

Table 1. benzylisoquinoline alkaloids with the highest anti-inflammatory activities for a) IKK-B and b) ERK1

Structure	Predicted pIC50
	6.39
	6.38
	6.36
	6.33

a)

Structure	Predicted pIC50
	6.17
	6.13
	6.12
	6.11

b)

3. Materials and Methods

From the ChemBL database were obtained 775 and 48 structures with activity against IKK-B (CHEMBL1991) and ERK1 (CHEMBL3385) respectively (https://www.ebi.ac.uk/chembl/). The compounds were classified using values of pIC_{50} (-log IC_{50}), presenting a range of 4.29 (from 5.01 to 9.30) for IKK-B and 3.10 (From 5.05 to 8.15) for ERK1. In this case, IC_{50} represents the concentration required for 50% inhibition of enzymatic activity. From SMILES codes, two-dimensional (2D) structures were generated in Standardizer software that canonized structures, added hydrogens, performed aromatic form conversions [JChem 14.9.1.0, 2014; ChemAxon (http://www.chemaxon.com)]. After were calculated 1064 two-dimensional molecular descriptors in Dragon 7 software. Obtained results were imported to Knime 3.1.0 software (www.knime.org). All variables were submitted to autoscaling and after were partitioned to generate two groups, a training group composed by the 80% of the whole molecules set and a test group composed by the remaining 20%. (Q)SAR models was performed using a Random Forest algorithm. Models were evaluated through cross validation (leave-one-out), $Q^2_{LOO,}$ as well as external test, Q^2_{ext}

4. Conclusions

The Ligand-based model using Random Forest and 2D molecular descriptors selected protoberberine skeleton with methoxyl and hydroxyl groups at the positions 9 and 10 as the most potential activity structures against IKK-B from an in-house database of benzylisoquinoline alkaloids. For ERK1 the RF model selected the same core without any substitution as the most potential active compound.

Acknowledgments

We would like to thank the Student Agreement Program of Graduate—PEC-PG of Brazilian National Research Council (CNPq)—Brazil

Author Contributions

LS, MFA built database; CHA performed all calculus; and CHA, MTS and MFFMD wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Maślińska, D.; Gajewski, M. Some aspects of the inflammatory process. *Folia neuropathologica/Association of Polish Neuropathologists and Medical Research Centre, Polish Academy of Sciences* **1997**, *36*, 199-204.
2. Senegas, A.; Gautheron, J.; Maurin, A.G.; Courtois, G. Ikk-related genetic diseases: Probing nf-kappab functions in humans and other matters. *Cell Mol Life Sci* **2015**, *72*, 1275-1287. DOI: 10.1007/s00018-014-1793-y.
3. Li, T.; Wong, V.K.; Jiang, Z.H.; Jiang, S.P.; Liu, Y.; Wang, T.Y.; Yao, X.J.; Su, X.H.; Yan, F.G.; Liu, J., *et al.* Mutation of cysteine 46 in ikk-beta increases inflammatory responses. *Oncotarget* **2015**, *6*, 31805-31819. DOI: 10.18632/oncotarget.5567.
4. Karin, M.; Greten, F.R. Nf-kappab: Linking inflammation and immunity to cancer development and progression. *Nat Rev Immunol* **2005**, *5*, 749-759. DOI: 10.1038/nri1703.

5. Busca, R.; Christen, R.; Lovern, M.; Clifford, A.M.; Yue, J.X.; Goss, G.G.; Pouyssegur, J.; Lenormand, P. Erk1 and erk2 present functional redundancy in tetrapods despite higher evolution rate of erk1. *BMC Evol Biol* **2015**, *15*, 179. DOI: 10.1186/s12862-015-0450-x.
6. PhosphoSitePlus. http://www.phosphosite.org/proteinAction?id=595&showAllSites=true (accessed on 20/12/2016).
7. Beaudoin, G.A.; Facchini, P.J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **2014**, *240*, 19-32.
8. Hawkins, K.M.; Smolke, C.D. Production of benzylisoquinoline alkaloids in saccharomyces cerevisiae. *Nature Chemical Biology* **2008**, *4*, 564-573.

Notes:
MDPI do not released one specific template for Sciforum conference. Consequentyl, this is not official template released by MDPI. In principle, the papers may be presented without specific format. However, the chairperson and the secretariat of the conference decided to create this template to give homogeneity to Mol2Net works. As is, this template should be used only for Mol2Net conference. Please, do not use this template for MDPI journals. Please, delete these notes before saving your final version. Type of the Paper (Proceeding, Letter, Expert Opinion, Communication, etc.). MDPI generates doi upon author request.
YouTube link: this option is only for those authors with welcome videos and/or oral presentations, plenary conferences uploaded to Mol2Net YouTube site.
Reference list: *We recommend the use of reference management software to prepare the references list (e.g., Endnote, http://endnote.com/styles/MDPI.ens).*

6TH BCNP
Brazilian Conference
On Natural Products
XXXII RESEM

Federal University of Espírito Santo
Vitória – ES/Brazil
November 05–08/2017

Certificate of poster presentation

This is to inform that the paper:

In silico Studies to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an Asteraceae in-house Database

authored by:

Marcus Scotti, Chonny Herrera-Acevedo, Luciana Scotti

was presented in the 6th BCNP Brazilian Conference on Natural Products/XXXII RESEM, Nov. 5th-8th, in Vitoria, ES.

Vitoria, 08th November 2017

Dr Warley de Souza Borges
Organizing Committee

Dr. Paulo Cezar Vieira
Scientific Organizing Committee

9130d0b902f67f6aa9495c9e8910f753

Galoá { Certificados

Este documento que você recebeu utiliza o sistema Galoá para emissão de comprovantes online, permitindo agilidade e segurança baseada em código de autenticidade e ainda ajuda a preservar o meio ambiente.*
Até o próximo evento!

Equipe **Galoá**

*As informações contidas neste são de inteira responsabilidade da organização do evento. Caso tenha algum equívoco em seu certificado, por favor verifique seus dados no sistema, ou entre em contato com o comitê organizador.

» Validação

Consulte a autenticidade e dados desse documento em nosso site: http://certificados.galoa.com.br/validar
Basta inserir o código
9130d0b902f67f6aa9495c9e8910f753
e será realizada a validação em alguns segundos.

» Próximos eventos

Fique sabendo de outros acontecimentos através do maior repositório científico e acadêmico do Brasil:
http://galoa.com.br/eventos-cientificos/agenda

» Saiba mais!

Esperamos que você tenha gostado de utilizar a plataforma Galoá e gostariamos de lhe convidar para acompanhar o nosso blog, tenha sua pesquisas em destaque!
http://galoa.com.br/blog

5° Encontro de Química da UFPB
Formação dos Profissionais de Química:
Novos Desafios para Novos Tempos
25 a 28 de Setembro de 2017

Triagem virtual baseada nas estruturas dos ligantes e receptores para selecionar compostos de um banco de dados de lactonas sesquiterpénicas de Asteraceae com potencial atividade leishmanicida.

Chonny Herrera-Acevedo (PG)[1], Luciana Scotti (PQ)[1], Marcus Tullius Scotti (PQ)[1,2*]

[1]*Universidade Federal da Paraíba, Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos - Campus I.*
[2] *Universidade Federal da Paraíba. Departamento de química -Campus I.*
** mtscotti@gmail.com*

Palavras Chave: *Leishmania donovani, Triagem virtual, Asteraceae, Lactonas sesquiterpénicas, Aprendizado de máquina*

Introdução

Entre os produtos naturais, que têm sido uma fonte de inspiração no desenvolvimento de agentes terapêuticos, destacam-se as lactonas sesquiterpénicas (LSS), as quais já têm presentado atividade contra doenças parasitárias[1-3].
Neste trabalho, foi desenvolvida uma aproximação combinada dos resultados de probabilidade obtidos de uma triagem virtual (TV) baseada na estrutura dos ligantes e uma triagem virtual baseada na estrutura dos receptores, o que permitiu a seleção de algumas LSS com potencial atividade leishmanicida.

Resultados e Discussão

Inicialmente foi feita uma TV baseada nas estruturas dos ligantes, desenvolvendo modelos Random Forest (RF) em Knime, para amastigota e promastigota de *L. donovani* usando um set de 1.306 lactonas sesquiterpénicas (LSS), cujas estruturas estão armazenadas no banco de dados SistematX (http://sistematx.ufpb.br) e moléculas com testes de atividade leishmanicida *in vitro* obtidas do ChEMBL, Sendo calculada a probabilidade de atividade leishmanicida das 1.306 LSS.
Estruturalmente, foi observado que no caso das promastigotas as 5 moléculas são dímeros de LSS entanto para amastigotas dois classes de esqueletos são observados guaianolideos e germacranolídeos.
Em seguida, no software Molegro, realizou-se uma TV baseada na estrutura dos receptores, (docking molecular) usando todo o conjunto de LSS e três proteínas alvo de *L. donovani*: N-miristoil transferase (NMT), ornitina decarboxilase (ODC) e MPK3.
Dois dímeros de LSS obtidos de *Smanthallus sonchifolius:* enidrina e uvedafolina apresentaram os menores valores de energia para as três enzimas alvo de *L. donovani* testadas
Finalmente, foram combinados os resultados de probabilidade obtidos nas metodologias anteriores, com a finalidade de verificar as moléculas potencialmente leishmanicidas bem como elucidar seu possível mecanismo de ação, permitindo a identificação de compostos *multitarget*. A probabilidade combinada foi calculada mediante a equação:

$$p_c = \frac{p_s + (1 + TN)\ x\ p}{2 + TN}$$

Onde *p*= probabilidade da TV baseada na estrutura dos ligantes, p_s = Probabilidade na TV baseada na estrutura dos receptores. p_c = Probabilidade combinada e TN a taxa de falsos negativos da validação cruzada.

Tabela 1. Estruturas melhor classificadas na aproximação combinada de triagens virtuais.

	Amastigota				Promastigota			
Proteín	LS	*p*	p_s	p_c	LS	*p*	p_s	p_c
NMT	Enidrofolina	0.66	0.99	0.78	Japonicona V	0.61	0.87	0.70
ODC	Enidrofolina	0.66	1	0.78	Bedforia Sym.	0.72	0.61	0.68
MPK3	Enidrofolina	0.66	1	0.78	Bedforia Sym.	0.72	0.61	0.68

Conclusões

Usando uma abordagem combinada de TV baseada em estrutura e em ligantes, foi possível a identificação de LSS leishmanicidas *multitarget,* sendo enhidrofolina um interessante dímero de SLL que age nas amastigotas de *L. donovani.*

Agradecimentos

Gostaríamos de agradecer ao CNPq e Capes pelo Apoio Financeiro e o Programa estudante convenio - Pós-Graduação (PEC-PG) do CNPq-Brasil.

[1] C. Herrera Acevedo, L. Scotti, M. Feitosa Alves, M. F. M. Diniz, M. Scotti, *Molecules* **2017**, *22*, 79
[2] T. J. Schmidt, A. M. Nour, S. A. Khalid, M. Kaiser, R. Brun *Molecules*. **2009**, *14*, 2062-2076
[3] D. J. Newman, G. M. Cragg, *J. Nat. Prod* **2016**, *79*, 629-661