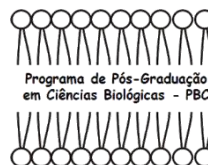**UNIVERSIDADE ESTADUAL DE MARINGÁ**

CENTRO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

ÁREA DE CONCENTRAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Paulo Sérgio Alves Bueno

Aplicações de técnicas de bioquímica e biofísica computacionais na descoberta de moléculas candidatas a medicamentos: Estudos *in silico*, *in vitro* e *in vivo*

Maringá

Agosto-2019

Paulo Sérgio Alves Bueno

Aplicações de técnicas de bioquímica e biofísica computacionais na descoberta de moléculas candidatas a medicamentos: Estudos *in silico*, *in vitro* e *in vivo*

Tese apresentada ao Programa de Pós-graduação em Ciências Biológicas (área de concentração – Biologia Celular e Molecular), da Universidade Estadual de Maringá para a obtenção do grau de Doutor em Ciências Biológicas.

Orientador: Dr. Flavio Augusto Vicente Seixas

Maringá

Agosto-2019

PAULO SÉRGIO ALVES BUENO

APLICAÇÕES DE TÉCNICAS DE BIOQUÍMICA E BIOFÍSICA
COMPUTACIONAIS NA DESCOBERTA DE MOLÉCULAS
CANDIDATAS A MEDICAMENTOS: ESTUDOS *IN SILICO*, *IN VITRO*
E *IN VIVO*.
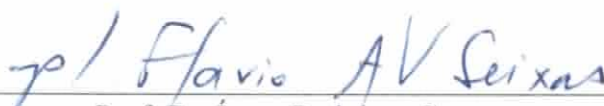
Tese apresentada como requisito parcial
para obtenção do grau de Doutor em
Ciências Biológicas, do Programa de Pós-
Graduação em Ciências Biológicas, da
Universidade Estadual de Maringá, sob a
apreciação da seguinte banca examinadora:

Aprovado em: 02 de Agosto de 2019.

BANCA EXAMINADORA

Prof. Dr. Flavio Augusto Vicente Seixas (Orientador)
Universidade Estadual de Maringá – UEM

Prof. Dr. Ícaro Putinhon Caruso
Universidade Estadual Paulista – IBILCE/UNESP
(Videoconferência)

Prof. Dr. João Carlos Palazzo de Mello
Universidade Estadual de Maringá – UEM

Prof. Dr. Celso Vataru Nakamura
Universidade Estadual de Maringá – UEM

Prof. Dr. Marco Aurélio Schüler de Oliveira
Universidade Estadual de Maringá – UEM

I

BIOGRAFIA

Paulo Sérgio Alves Bueno nasceu em Jandaia do Sul, estado do Paraná, no dia 6 de junho de 1990. Iniciou sua graduação em 2011 no curso Bacharelado em Bioquímica pela Universidade Estadual de Maringá (UEM). Durante esse período atuou em projetos de iniciação científica no Departamento de Bioquímica, Laboratório de Bioquímica e Fisiologia de Micro-organismos. Graduou-se em 2014 e no ano seguinte ingressou no mestrado do Programa d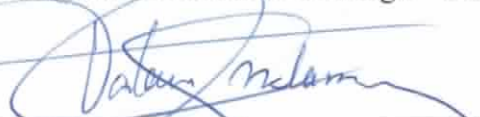e Pós-Graduação em Ciências Biológicas (Biologia Celular e Molecular) da UEM. Atualmente é aluno de doutorado no mesmo programa e professor temporário do Departamento de Bioquímica da UEM. Realizou suas pesquisas sob orientação do professor Dr. Flavio Augusto Vicente Seixas, no Laboratório de Bioquímica Estrutural, em Umuarama-PR, onde localiza-se o *campus* de Tecnologia da universidade. Sua dissertação de mestrado e atuais estudos no doutorado envolveram a utilização de métodos de simulação computacional, como modelagem, atracamento (*docking*) e dinâmica molecular para o estudo da interação proteína-ligante, com o intuito de descobrir novos fármacos. Além disso, está envolvido em outras pesquisas, como a descoberta de novos RNAs não-codificantes. A seguir, publicações científicas e um depósito de patente realizados no período em que foi desenvolvido seu doutorado:

Artigos:

Duarte Junior, Francisco Ferreira; Bueno, Paulo Sérgio Alves; Pedersen, Sofia L; Rando, Fabiana Dos Santos; Pattaro Júnior, José Renato; Caligari, Daniel; Ramos, Anelise Cardoso; Polizelli, Lorena Gomes; Lima, Ailson Francisco Dos Santos; De Lima Neto, Quirino Alves; Krude, Torsten; Seixas, Flavio Augusto Vicente; Fernandez, Maria Aparecida. Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*. RNA BIOLOGY, v. 1, p. 1, 2019. **(F.I. JCR/2017: 5,22)**

Bagatin, Mariane C; F Rozada, Andrew M; V Rodrigues, Franciele A; A Bueno, Paulo S; Santos, Jessyka L; Canduri, Fernanda; Kioshima, Érika S; V Seixas, Flavio A; Basso, Ernani A; Gauze, Gisele F. New 4-methoxy-naphthalene derivatives as promisor antifungal agents for paracoccidioidomycosis treatment. FUTURE MICROBIOLOGY, v. 1, p. 1, 2019. **(F.I. JCR/2018: 2,75)**

Bueno, Paulo Sérgio Alves; Biavatti, Débora Carina; Chiarello, Alex Sandro Duarte; Fassina, Verônica Aureliana; Fernandez, Maria Aparecida; Seixas, Flávio Augusto Vicente. The structure of viral cathepsin from *Bombyx mori Nuclear Polyhedrosis Virus* as a target against grasserie: Docking and molecular dynamics simulations. JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS, v. -, p. 1-24, 2018. **(F.I. JCR/2017: 3,11)**

De Paula Reis, Michelle; De Lima, Daniely Alves; Pauli, Karoline Bach; Andreotti, Carlos Eduardo Linhares; De Moraes, André Luiz Soares; Gonçalves, Daniela Dib; Navarro, Italmar Teodorico; Bueno, Paulo Sérgio Alves; Seixas, Flavio Augusto Vicente; Gasparotto Junior, Arquimedes; Lourenço, Emerson Luiz Botelho. Molecular docking to *Toxoplasma gondii* thymidylate synthase-dihydrofolate reductase and efficacy of raltitrexed in infected mice. PARASITOLOGY RESEARCH, v. 117, p. 1-7, 2018. **(F.I. JCR/2018: 2,08)**

Patente:

Bueno, P. S. A.; Pimentel, A. L.; Rodrigues, F. A. V.; Bandoch, G. F. G.; Bagatin, M. C.; Cotica, E. S. K. Seixas, F. A. V. Composição farmacêutica baseada na molécula *(S)-2- Ammonio-3-{(6-Bromo-2-Methylquinolin-4-Yl)Thio}Propanoate* e seu uso na preparação de medicamentos para o tratamento de infecções causadas por micro-organismos do gênero *paracoccidioides spp.*. 2018, Brasil. Número do Registro: Br10201801245, Instituição De Registro: INPI - Instituto Nacional Da Propriedade Industrial. Depósito: 18/06/2018.

AGRADECIMENTOS

Deus, obrigado por guiar-me pelos caminhos que fizeram eu chegar até aqui.

Aos meus pais, Osvaldo e Maria; e meus irmãos Anderson, Tatiane, Fabiano e Daiane, por demonstrar o verdadeiro amor.

Aos meus familiares e amigos, pela verdadeira amizade, mesmo quando não podia estar sempre presente.

Ao meu orientador e amigo prof. Dr. Flavio Augusto Vicente Seixas, por demonstrar o quanto a ciência é bonita, apesar de complexa. Por todos os ensinamentos, não só neste campo, mas para a vida. Com certeza, você é uma fonte de inspiração.

Aos colegas de laboratório: José Renato Pattaro, Quirino Lima Neto, Francisco Junior, Diego de Souza Lima, Elisangela Andrade Angelo, Jovana Chiapetti Tartari, Tainá Cruz, Arethusa Lobo Pimentel pelo companheirismo na pesquisa.

Aos amigos Ivan Lúcio Rosina, Rodrigo Prizon, Marcelo Aoki e João Henrique de Souza.

A todos os professores do Departamento de Bioquímica da Universidade Estadual de Maringá.

A todos os professores que fizeram parte da minha trajetória escolar e acadêmica.

Ao Programa de Pós-graduação em Ciências Biológicas da Universidade Estadual de Maringá.

À CAPES, CNPq, Fundação Araucária e outros órgãos de fomento pelo apoio financeiro durante toda a minha trajetória acadêmica.

E a todos aqueles não mencionados que também contribuíram com este trabalho.

Muito Obrigado!

APRESENTAÇÃO

Em consonância com as regras do Programa de Pós-Graduação em Ciências Biológicas, esta tese é composta por um resumo geral do trabalho limitado a 3 páginas em português e em inglês, e quatro artigos de pesquisa científica original contemplando os resultados obtidos ao longo do desenvolvimento do doutorado. Em todos os trabalhos foram utilizadas simulações computacionais de macromoléculas. O primeiro artigo foi redigido e submetido de acordo com as normas da revista científica *Journal of Molecular Modeling* (ISSN: 1610-2940/ *Impact Factor* JCR: 1,33). Neste trabalho é descrito a descoberta de moléculas com atividade antifúngica que atuam na enzima homoserina desidrogenase do fungo patogênico *Paracoccidioides brasiliensis*, causador da paracoccidioidomicose (PCM). O segundo artigo está sob revisão pela mesma revista citada anteriormente. Nele foi utilizada a enzima α-amilase para compreender o mecanismo de inibição de compostos com ação hipoglicemiante. Este artigo já foi avaliado pelos revisores, as correções incorporadas e aguardamos a resposta final do editor. O terceiro artigo foi aceito para publicação na revista *Future Microbiology* (ISSN: 1746-0921/*Impact Factor* JCR: 2,75) e refere-se a descoberta de um potente inibidor da enzima corismato sintase de *P. brasiliensis*, com ação antifúngica. Por fim, o último artigo, publicado na revista *Journal of Biomolecular Structure and Dynamics* (ISSN: 1538-0254/*Impact Factor* JCR: 3,11) apresenta a descoberta de uma molécula com ação antiviral contra *Bombyx mori Nuclear Polyhedrosis Virus*, hospedeiro do inseto bicho-da-seda (*Bombyx mori*). Em todos os estudos foram realizados testes experimentais para a validação dos resultados.

RESUMO

A compreensão das interações microscópicas entre um ligante e um determinado alvo celular é um grande desafio para pesquisa científica. O entendimento destes detalhes permite desenvolver medicamentos, por vezes, mais eficientes, ou ainda, podem trazer benefícios em aplicações biotecnológicas de interesse econômico. Neste campo, técnicas de simulação computacional, tais como modelagem, atracamento (*docking*) e dinâmica molecular, têm permitido entender vários aspectos deste mecanismo de interação. O conteúdo da presente tese descreve resultados de quatro artigos científicos que utilizaram as técnicas computacionais citadas. Nos artigos 1 e 3 é descrito a descoberta de moléculas com atividade antifúngica, que atuam em dois diferentes alvos proteicos (homoserina desidrogenase e corismato sintase), contra o fungo patogênico *Paracoccidioides brasiliensis*, causador da paracoccidioidomicose (PCM). No artigo 2 foi utilizada a enzima α-amilase para compreender o mecanismo de inibição de compostos com ação hipoglicemiante. Por fim, o artigo 4 descreve a descoberta de uma molécula com ação antiviral contra *Bombyx mori Nuclear Polyhedrosis Virus*, que afeta o inseto bicho-da-seda (*Bombyx mori*). Em todos os estudos foram realizados testes experimentais para a validação dos resultados.

Metodologia: Inicialmente as estruturas tridimensionais dos alvos proteicos foram determinadas por modelagem molecular por homologia, exceto a da α-amilase, que já possuía estrutura cristalográfica depositada no *protein data bank* (PDB). Todas as estruturas dos alvos foram submetidas a procedimentos de minimização de energia. Em seguida, protocolos de *docking* molecular foram validados pelo método de *redocking* para serem utilizados em simulações de varredura virtual em larga escala (VS) de pequenas moléculas com disponibilidade comercial. Estas moléculas foram obtidas do banco de dados Zinc e subdivididas por catálogo de fornecedores. Estudos de dinâmica molecular (MD) foram realizados para verificar o comportamento e calcular as interações envolvidas na estabilização dos complexos proteína-ligante. Em alguns casos (artigo 2 e 4) foram utilizadas metodologias mais sensíveis como *Molecular Mechanics Poisson-Boltzmann Surface Area* (MM-PBSA) para o cálculo de $\Delta G_{ligação}$. Para a comprovação das teorias foram utilizados testes *in vitro* e *in vivo*. Para os testes *in vitro*, diferentes metodologias foram empregadas, tais como: concentração inibitória mínima (MIC), concentração fungicida mínima (MFC), estudos cinéticos com a enzima alvo na presença do inibidor (IC_{50}) e, testes de citotoxicidade em células de mamíferos.

Artigo 1 - **Resultados e discussão:** Neste estudo, o alvo celular foi a enzima homoserina desidrogenase de *Paracoccidioides brasiliensis* (*Pb*HSD). A estrutura da enzima foi obtida por modelagem em estudos anteriores. A biblioteca virtual analisada por VS continha 112.572 moléculas. Procedimentos de *docking* foram realizados por três diferentes programas. As moléculas HS9 (Zinc2123137), HS7 (Zinc15967722), HS6 (Zinc20611644) foram selecionadas em comum por todos os programas e adquiridas para ensaios *in vitro*. Os ensaios microbiológicos demostraram que todas as moléculas apresentaram atividade antifúngica com valores de MICs/MCFs de 8, 32 e 128 μg.mL^{-1}, respectivamente. As duas moléculas mais promissoras apresentaram resultados satisfatórios com ampla faixa terapêutica nos ensaios de citotoxicidade. As simulações de MD da *Pb*HSD com as moléculas selecionadas indicaram que o ligante se manteve preso ao sítio ativo da proteína ao longo de toda a simulação e, sua presença não levou a proteína a se desenovelar. Além disso, as moléculas apresentaram um maior número de contatos com a proteína em relação ao substrato HSE, o que pode justificar o valor de CIM obtido. Os resultados apresentados sugerem que a molécula HS9 pode ser considerada como um *hit* no desenvolvimento de fármacos contra a PCM.

Artigo 2 - **Resultados e discussão**: Neste trabalho a estrutura cristalográfica da enzima α-amilase pancreática (pdbid: 3l2l) foi utilizada como alvo de VS. A biblioteca de ligantes continha as estruturas de taninos hidrolisáveis e condensados, mais os inibidores já conhecidos da enzima. Cálculos teóricos de *ΔG$_{ligação}$* pelo método MM-PBSA foram realizados e indicaram que o tanino hidrolizado (HTN), acarbose (ACA) e o tanino condensado (CTN) possuem alta afinidade em relação ao substrato amilose (AMY), com valores de −350,0, −346,2, −320,5 e −209,2 kJ.mol^{-1}, respectivamente. A afinidade relativa encontrada para o HTN e CTN estão de acordo com os valores obtidos experimentalmente. Os resultados deste estudo forneceram informações úteis para a caracterização da ligação de taninos em α-amilase, que podem ser aplicados em futuros estudos visando encontrar novas moléculas hipoglicemiantes.

Artigo 3 - **Resultados e discussão**: A enzima alvo deste estudo foi a corismato sintase de *Candida albicans* (CaCS). A estrutura da enzima foi modelada na forma tetramérica, ligada ao substrato 5-EPSP e ao cofator FMNH$_2$. O melhor modelo de CaCS foi submetido a minimização de energia e, em seguida, utilizada em procedimentos de VS por meio do protocolo validado por *redocking*. Foram selecionadas duas moléculas identificadas como CS8 (Zinc6445805) e CaCS02 (Zinc4217277), dentro de uma biblioteca virtual que continha 196.099 compostos. Simulações de MD da enzima CaCS

durante 50 ns ligadas ao 5-EPSP, CS8 ou ao CaCS02 confirmaram a estabilidade da unidade biológica como um homotetrâmero e, que os compostos CS8 e CaCS02 se ligam ao sítio ativo por um número maior de contatos em relação ao substrato 5-EPSP. Adicionalmente, estudos de MIC e MFC demonstraram atividade antifúngica contra *Paracoccidioides brasiliensis* com valores 512 e 32 $\mu g.mL^{-1}$ para o CS8 e CaCS02, respectivamente. Além disso, o CaCS02 demonstrou forte efeito sinérgico em combinação com amfotericina-B. Estudos *in vitro* utilizando a CS recombinante de *P. brasiliensis* mostraram um IC_{50} de 29 μM para o CaCS02 dando suporte a resultados *in silico*, os quais sugeriram uma forte ligação com a CS, podendo ser responsável pela atividade fungicida observada. Os resultados demonstraram que utilizando técnicas de simulação computacional foi possível obter uma molécula com grande potencial fungicida.

Artigo 4 - Resultados e discussão: A estrutura da enzima catepsina viral (V-Cath) de *Bombyx mori Nuclear Polyhedrosis Virus* (*Bm*NPV-Cath) foi modelada em complexo com o inibidor competitivo MYP. A estrutura minimizada da *Bm*NPV-Cath foi utilizada como alvo na VS de uma biblioteca com 111.500 moléculas de produtos naturais. Quatro moléculas foram selecionadas em comum por dois programas de *docking*, sendo denominadas de: Bm1 (Zinc70705449), Bm2 (Zinc70701476), Bm4 (Zinc 02496946) e Bm5 (Zinc1288007). Em simulações de MD, as análises dos complexos indicaram que o equilíbrio foi alcançado após 5 ns. Os cálculos teóricos de $\Delta G_{ligação}$ pelo método MM-PBSA, indicaram que a molécula Zinc12888007 (Bm5) tem alta afinidade pela enzima. A partir de testes *in vivo* com lagartas de *B. mori* infectadas com BmNPV e tratadas com os compostos selecionados, confirmou-se que o candidato Bm5 foi o mais promissor. O tratamento com dose única de 100 μg do composto Bm5 dissolvido em Pluronic-F127 0,02%, foi possível obter uma redução da mortalidade das lagartas em 22,6%. Estudos com diferentes doses e em diferentes concentrações do Bm5 ainda não foram realizados. Contudo, estes ensaios podem melhorar ainda mais os resultados desta pesquisa, que demonstrou uma aplicação biotecnológica direta contra uma das pragas mais comuns na sericicultura mundial.

Palavras-chave: simulação computacional, varredura virtual, *docking* molecular, dinâmica molecular, antifúngicos, antivirais, hipoglicemiantes, descoberta de fármacos.

ABSTRACT

The knowledge about the microscopic nature of the interactions between a ligand and a particular cellular target is a major challenge for scientific research. The understanding of these details allows to develop new medicines, which sometimes are more efficient, or can bring benefits in biotechnological applications of economic interest. In this field, computational simulation techniques, such as modeling, docking and molecular dynamics, have allowed to understand several aspects of these mechanisms of interaction. The content of the present thesis describes the results of four scientific papers that have make use of the cited computational techniques. In articles 1 and 3 the discovery of molecules with antifungal activity, acting on two different protein targets (homoserine dehydrogenase and chorismate synthase), against the pathogenic fungus *Paracoccidioides brasiliensis*, the causing agent of paracoccidioidomycosis disease (PCM) is described. In article 2, the α-amylase enzyme was used to understand the mechanism of inhibition of new hypoglycemic compounds. Finally, the article 4 describes the discovery of a molecule with antiviral activity against *Bombyx mori Nuclear Polyhedrosis Virus*, that affects the silkworm (*Bombyx mori*). In all studies, experimental assays were carried out to validate the results.

Methodology: Initially, the three-dimensional structures of the protein targets were modeled by homology, except for α-amylase, which had already a crystallographic structure deposited in the protein databank (PDB). All protein targets were submitted to energy minimization procedures. Then, molecular docking protocols were validated by the redocking method in order to be used in virtual screening simulations (VS) of purchasable small molecules library. These libraries were obtained from the Zinc database and subdivided by supplier catalog. Molecular dynamics (MD) studies were performed to verify the behavior of the complex and also to calculate the interactions involved in the stabilization of protein-ligands. In some cases, (article 2 and 4), more sensitive methodologies such as Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) were used to calculate $\Delta G_{binding}$. The *in vitro* and *in vivo* assays were used to validate the theoretical findings. For *in vitro* assays, different methodologies were used, such as: minimal inhibitory concentration (MIC), minimal fungicidal concentration (MFC), kinetic studies using the target enzyme in the presence of the inhibitor (IC_{50}), and cytotoxicity assays in mammalian cells.

Article 1 - Results and discussion: In this study, the cellular target was the enzyme homoserine dehydrogenase from *Paracoccidioides brasiliensis* (*Pb*HSD). The structure

of the enzyme was obtained by homology modeling in previous studies. The virtual library analyzed by VS contained 112,572 molecules. Docking procedures were performed by three different programs. The molecules HS9 (Zinc2123137), HS7 (Zinc15967722) and HS6 (Zinc20611644) were selected in common by all programs and purchased for *in vitro* assays. Microbiological assays showed that all the molecules presented antifungal activity with MICs/MCFs values of 8, 32 and 128 µg.mL^{-1}, respectively. The two most promising molecules presented satisfactory results with wide therapeutic range in the cytotoxicity assays. The MD simulations of *Pb*HSD with the selected molecules indicated that the ligand remained bonded to the active site of the protein throughout the simulation and its presence did not let the protein to unfold. In addition, the molecules showed a greater number of contacts with the protein regarding the HSE substrate, which may justify the MIC value obtained. The results presented suggest that the molecule HS9 can be considered as a hit in the development of antifungals against PCM.

Article 2 - Results and discussion: In this work the crystallographic structure of the enzyme pig pancreatic α-amylase (pdbid: 3l2l) was used as target for VS. The library of ligands contained the structures of hydrolysable and condensed tannins, plus the known inhibitors of the enzyme. Theoretical calculations of ΔG$_{binding}$ by the MM-PBSA method were performed and indicated that hydrolyzed tannin (HTN), acarbose (ACA) and condensed tannin (CTN) have highest affinity than the amylose substrate (AMY), with values of -350.0, -346.2, -320.5 and -209.2 kJ.mol^{-1}, respectively. The relative affinity found for HTN and CTN are in agreement with the values obtained experimentally. The results of this study provided useful information for the characterization of tannin binding in α-amylase, which can be applied in future *in vivo* studies aiming the discovery of new hypoglycemic molecules.

Article 3 - Results and discussion: The target enzyme of this study was the chorismate synthase from *Candida albicans* (*Ca*CS). The structure of the enzyme was modeled in tetrameric form, bounded to the 5-EPSP substrate and the FMNH2 cofactor. The best *Ca*CS model was submitted to energy minimization and then, used in VS procedures through a protocol validated by redocking. Two molecules identified as CS8 (Zinc6445805) and CaCS02 (Zinc4217277) were screened within a virtual library containing 196,099 compounds. The MD simulations of the *Ca*CS enzyme bounded to 5-EPSP, CS8 or CaCS02 by 50 ns confirmed the stability of the biological unit as a homotetramer, and that compounds CS8 and CaCS02 bind to the active site by a greater

number of contacts than substrate 5-EPSP. Additionally, MIC and MFC studies demonstrated antifungal activity against *Paracoccidioides brasiliensis* with values of 512 and 32 μg.mL^{-1} for CS8 and CaCS02, respectively. In addition, CaCS02 demonstrated a strong synergistic effect in combination with amphotericin-B. *In vitro* studies using the recombinant CS from *P. brasiliensis* showed an IC$_{50}$ of 29 μM for the CaCS02 supporting our *in silico* results, which suggested a strong binding of this ligand to CS and this fact could be responsible for the observed fungicidal activity. These results demonstrated that using computational simulation techniques, it was possible to identify a molecule with great fungicidal potential.

Article 4 - Results and discussion: The structure of the viral cathepsin (V-Cath) enzyme from *Bombyx mori Nuclear Polyhedrosis Virus* (*Bm*NPV-Cath) was modeled in complex with the competitive inhibitor MYP. The minimized structure of *Bm*NPV-Cath was used as target in the VS of a library with 111,500 molecules of natural products from Zinc database. Four molecules were selected in common by two docking programs, being named: Bm1 (Zinc70705449), Bm2 (Zinc70701476), Bm4 (Zinc 02496946) and Bm5 (Zinc1288007). In MD simulations, the analyzes of the complexes indicated that the equilibrium was reached after 5 ns. The theoretical calculations of ΔG$_{binding}$ by the MM-PBSA method, indicated that the molecule Zinc12888007 (Bm5) has high affinity for the enzyme. The *in vivo* assays with *B. mori* caterpillars infected with *Bm*NPV and treated with the selected compounds was used to confirm the Bm5 as the most promising candidate. The treatment with a single dose of 100 μg of Bm5 dissolved in Pluronic-F127 0.02% promotes a reduction in mortality of the caterpillars by 22.6%. Studies with different doses and at different concentrations of Bm5 have not yet been carried out. However, these assays may further improve the results of this research, which demonstrates a direct biotechnological application against one of the most common diseases in sericulture worldwide.

Key words: computational simulation, virtual screening, molecular docking, molecular dynamics, antifungal, antiviral, hypoglycemic, drug discovery.

Artigo 1: New inhibitors of homoserine dehydrogenase from *Paracoccidioides brasiliensis* presenting antifungal activity.

New inhibitors of homoserine dehydrogenase from *Paracoccidioides brasiliensis* presenting antifungal activity.

Paulo Sérgio Alves Bueno[1], Franciele Abigail Vilugron Rodrigues[2], Jessyka Lima Santos[4], Fernanda Canduri[4], Débora Carina Biavatti[1], Arethusa Lobo Pimentel[1], Mariane Cristóvão Bagatin[3], Érika Seki Kioshima[2], Gisele de Freitas Gauze[3], Flavio Augusto Vicente Seixas[1]*.

[1]*Department of Technology, Universidade Estadual de Maringá, Umuarama, PR, Brazil.*
[2]*Department of Clinical Analysis and Biomedicine, Universidade Estadual de Maringá, Maringá, PR, Brazil.*
[3]*Department of Chemistry, Universidade Estadual de Maringá, Maringá, PR, Brazil.*
[4]*São Carlos Institute of Chemistry, Universidade de São Paulo, São Carlos, SP, Brazil.*

* Corresponding author: Departamento de Tecnologia, Universidade Estadual de Maringá, Campus Umuarama. Av. Ângelo Moreira da Fonseca, 1800, 87506-370, Umuarama, PR, Brazil. E-mail: favseixas@uem.br, Phone: +55 44 36219337

Abstract

   Paracoccidioidomycosis (PCM) is a systemic mycosis caused by fungi of the genus *Paracoccidioides* spp., which mainly affects workers in rural regions of Latin America. Although the antifungal agents currently available for the treatment of PCM are effective in controlling the disease, many months are needed for healing, making the side effects and drug interactions relevant. In addition, conventional treatments are not able to control the sequelae left by PCM, even after the cure, justifying the search for new therapeutic options against PCM. In this context, the enzyme homoserine dehydrogenase of *P. brasiliensis* (*Pb*HSD) was used to screen a library of natural products from the Zinc database using three different docking programs, i.e. Autodock, Molegro and CLC Drugdiscovery Workbench. Three molecules (Zinc codes: 2123137, 15967722 and 20611644) were better ranked than the homoserine substrate (HSE), and were used for *in vitro* trials of the minimum inhibitory concentration (MIC) and minimal fungicidal concentration (MCF). All three molecules presented a fungicidal profile with MICs/MCFs of 8, 32 and 128 μg.mL^{-1}, respectively. The two most promising molecules presented satisfactory results with wide therapeutic ranges in the cytotoxicity assays. Molecular dynamics simulations of *Pb*HSD indicated that the ligands remained bound to the protein by a common mechanism throughout the simulation. The molecule with the lowest MIC value presented the highest number of contacts with the protein. The results presented in this work suggest that the molecule Zinc2123137 may be considered as a hit in the development of new therapeutic options for PCM.

Keywords: *Paracoccidioides brasiliensis*; virtual screening; homoserine dehydrogenase; drug discovery.

Introduction

Paracoccidioidomycosis (PCM) is a systemic mycosis caused by fungi of the genus *Paracoccidioides* spp., where *P. brasiliensis* is the best known species [1]. This disease is restricted to countries in Latin America, where it affects mainly low-income rural workers and is considered a neglected disease [2]. In Brazil, PCM is considered a major public health problem, classified as one of the major causes of death among the predominantly chronic or recurrent infectious diseases, and its mortality rate is the highest among the systemic mycoses [3].

PCM manifests itself in two well-established clinical forms. The acute/subacute (juvenile) form occurs predominantly in children and young adults. In these cases, the disease develops in a few weeks or months after exposure to *Paracoccidioides* spp. Among the most common clinical manifestations in the affected patient are fever, exacerbated weight loss, hepatosplenomegaly and lymph node infarction. In contrast, the chronic (adult) form of PCM, which affects about 80% to 95% of cases, usually appears after age 30 and is limited to lesions in the lungs, upper airways, oral mucosa and skin adjacent to the mouth and nose; in rare cases, the fungus spreads to the adrenal glands and brain [4].

There are no specific drugs for PCM, and the therapeutic arsenal is limited to two classes of antifungal agents. The first group are derivatives of polyenes, represented mainly by amphotericin-B (AMB) [5]. Its mechanism of action occurs in the cell membrane, which binds to the sterol ergosterol by altering the permeability of the membrane [6]. This medication is usually effective in the treatment of mycoses and used as the initial therapy for PCM, although its use is restricted to the hospital environment [7]. It has several adverse effects, with nephrotoxicity as the most common and most relevant [8].

In less severe cases or after clinical improvement in severe cases, the patient undergoes continued treatment, mainly involving the use of azole derivatives [2]. These drugs act specifically on the enzyme lanosterol demethylase, inhibiting ergosterol biosynthesis [9]. Representatives of this group are itraconazole, voriconazole and posaconazole. These drugs are less toxic than AMB, but are fungistatics and therefore require a longer treatment time; this depends on the severity of the disease, and can reach 24 months in severe cases.[2] This long-term treatment incurs high costs and adverse effects [10,11], which increase the possibility of treatment abandonment and thus promote the appearance of resistant strains [12]. Reports of *P. brasiliensis* strains

resistant to azole derivatives are common in the literature [13,14]; however, the pharmacological arsenal against fungal infections is quite limited when compared to antibacterial drugs. In addition, the fungus presents virulence attributes that may contribute to the infection process in patients [15]. All these facts contribute to the limitations in the treatment of PCM. These demonstrate the need for discovery of new metabolic targets and new drugs with different and more effective mechanisms of action.

In this context, homoserine dehydrogenase (HSD) is an enzyme (EC 1.1.1.3) of the aspartate metabolic pathway, which reduces L-aspartate 4-semialdehyde to L-homoserine in a NAD(P)H-dependent reaction. This enzyme is essentially required for the biosynthesis of methionine, threonine and isoleucine in fungi. Humans do not synthesise these amino acids, as they are obtained from the diet [16]. Therefore, HSD has recently been explored as an interesting metabolic target in the search for drugs with selective antifungal action and low toxicity to the host [17-19].

In this context, the objective of this work was to perform virtual screening simulations for the discovery of new inhibitor candidates of the enzyme homoserine dehydrogenase from *Paracoccidioides brasiliensis* strain Pb18, the most virulent of this species, and validate the activity of the hits by means of *in vitro* assays.

Material and Methods

Target preparation, library selection and filtering.

The biological unity of homoserine dehydrogenase from *Paracoccidioides brasiliensis* (*Pb*HSD) strain Pb18, in the presence of the cofactor NAD$^+$ plus the ligand L-homoserine (HSE), has already been modelled and validated by previous results [17]. In this work, the structure of *Pb*HSD was used in a new screening of a library of purchasable natural products from the Zinc database [20]. The redocking protocol was validated by three different programs using three different algorithms, i.e. Autodock 4.2.3 [21] implemented in the Pyrx-0.9 graphical interface [22], the program Molegro Virtual Docker v6.0 [23] and the program CLC Drug Discovery Workbench v1.0.2 [23]. At the end of the screening, the compounds selected in common by the three programs were filtered by the ADMTox criteria through the DataWarrior program (openmolecules.org) and then acquired from InterBioScreen (Russia) for *in vitro* trials.

Molecular dynamic simulations

*Pb*HSD in a complex with ligands and the cofactor was subjected to energy minimisation and equilibration procedures by the molecular dynamics (MD) method through the NAMD2 program [24]. The Charmm c35b2-c36a2 force field (FF) [25] was used for the atoms of the protein and the cofactor NAD^+. In order to use accurate FF for the L-homoserine and other ligands found by virtual screening, we initially downloaded from Zinc database, the coordinate files in the *.mol2 format containing all hydrogens at the expected protonation state for pH 7.0. Then, these files were submitted to Swissparam server [26] for generation of the force fields in the Charmm format. Simultaneously, the same files were also subjected to quantum optimization calculations using the B3LYP/6-311G level of theory with the salvation model COSMO, by means of the ORCA program [27]. Then, the topology files of the ligands provided by the Swissparam were edited so that the partial charges of each atom were replaced by the respective Mülliken's charge provided by the ORCA program output.

Each protein-ligand complex was immersed in a periodic box with TIP3P water, with limits at least 10 Å away from the outermost surface of the molecule, under NPT conditions (constants: number of atoms (N), pressure (1 atm) and temperature (300 K)). A sufficient amount of Na^+ counterions was added to neutralise the system charges.

The simulations were performed in steps. Initially, the coordinates of the ligand atoms were fixed in space, while all other atoms of the system were subjected to 30,000 steps of minimisation by the conjugate gradient (CG) method. In the second step, all the atoms of the system were minimised by 10,000 steps of CG. The resulting minimised coordinates were used as input parameters for redocking and virtual screening. In the third step, the atoms of the protein, cofactor and ligand were fixed in space, while the water molecules and ions were equilibrated by 60 ps of MD. In the fourth and last step, all atoms were minimised again by another 20,000 steps of GC. For prolonged MD, all atoms were equilibrated for 50 ns, a sufficient amount of time to verify the convergence of the system to the conditions of thermodynamic equilibrium.

The analyses were performed in terms of root mean square deviation (rmsd); the radius of gyration of the protein-ligand complexes and by the contact frequency of the protein residues with the ligands within a range of up to 4.0 Å, and also by the root mean square fluctuation (rmsf) of the alpha carbons for each residue. All this information was extracted from the trajectory file. The prolonged simulations were performed on 20 nodes of an Intel Xeon E5-2695v2 Ivy Bridge, 2.4 GHz (480 cores)

from the SDumont cluster at Laboratório Nacional de Computação Científica (LNCC), Brazil.

In vitro assays

Compounds selected by virtual screening and the control drug amphotericin B (Sigma-Aldrich) were tested. Initially, a stock solution of the selected compounds was prepared in dimethylsulphoxide (DMSO) at 50 mg.mL^{-1}. The compounds were diluted in RPMI [RPMI-1640 medium without bicarbonate, with L-glutamine, supplemented with 2% glucose and buffered (pH 7.0) with 0.165 M morpholinopropanesulphonic acid (MOPS)] (Sigma-Aldrich), DMSO (1%) and Pluronic F-127 at 0.1% (Sigma-Aldrich) for trials. The fungi were incubated in the absence and presence of the diluent, in order to confirm that exposition to it did not influence the growth of the tested isolates.

To evaluate the susceptibility profile of the pathogenic fungi to the compounds, the broth microdilution test was performed using a method proposed by CLSI (*Clinical and Laboratory Standards Institute*, 2008), according to protocol M27-A3 (old NCCLS - *National Committee for Clinical Laboratory Standards*), with modifications [28]. The fungal isolates of *P. brasiliensis* (Pb18) were subcultured in Fava-Netto agar medium [29] at 37°C for five days. The viability of Pb18 was determined by counting viable cells in a Neubauer chamber by the trypan blue method [30]. For these experiments, ≥80% was considered a viable count; cells were then used in the *in vitro* susceptibility assays.

The inoculum were suspended in sterile saline buffer (0.85%) and added to RPMI media at a final concentration of 1-2×10^3 cells.mL^{-1}. The tests were performed on sterile 96-well microplates (TPP Zellkultur Test Plate 96F, Switzerland) and the final tested concentrations of the evaluated molecules ranged from 1 to 512 µg.mL^{-1}. The incubation time was seven days at 37°C. The data are presented as MIC, which corresponds to the concentration of each antifungal capable of inhibiting 80% of the isolate's growth. As the negative control was used only the RPMI medium plus DMSO (2%), in order to verify possible culture medium contamination. For the positive control, the yeast inoculum was added to the same medium (RPMI plus DMSO). The optical density found in this control is considered to be 100% growth. After reading the MIC, aliquots of 5 µL of all wells from the 96-well microplate were transferred and subcultured into brain-heart infusion agar at 37°C for seven days to determine the

minimum fungicidal concentration (MFC). The fungicide profile considered the lowest concentration capable of eliminating any fungal growth. All MIC (minimum inhibitory concentration) and MFC (minimum fungicidal concentration) determinations were done in duplicate.

Cytotoxicity evaluation

Cytotoxicity was assessed in two cell lineages, i.e. HeLa and Vero, using the CellTiter 96 assay (Promega, Madison, WI, USA), based on the reduction of MTS (*3-[4,5-dimethylthiazol-2-yl]-5-[3-carboxymethoxyphenyl]-2-[4-sulphophenyl]-2H-tetrazolium*). Briefly, cells were incubated in 96-well culture plates and maintained in Dulbecco's modified Eagle's medium (DMEM) with foetal bovine serum (5%) and treated with compounds in concentrations ranging between 2 and 1024 $\mu g.mL^{-1}$. After 24 h, the MTS solution was added to each well and the plate was incubated for 180 min in a CO_2 (5%) incubator at 37°C. Cell viability was determined by measurement of the optical density at 490 nm using a spectrophotometer. The cytotoxicity results are presented as the average from three independent experiments with three replicates for each compound. Percent cell viability was calculated in relation to the viability of the controls, which were treated with medium only.

Results and discussion

Virtual screening

The library of natural products obtained from the Zinc database contained 112,572 molecules. They were divided into 56 folders with approximately 2,000 molecules each, plus the structure of the substrate L-homoserine. The library was initially screened with the Autodock program and 388 molecules that were better ranked than the substrate (scores smaller than −5.61) were selected and stored in a folder called "Best Results". The best results were screened four times with the Autodock program, another four times with the Molegro program and another four times with the CLC program. The purpose of these replicates was to eliminate false positive results, since if a molecule is selected by three programs using distinct search and ranking methods, it is more likely that the selected ligand is indeed a promising hit. In order to perform the screening, the Autodock and CLC programs used their standard search and rank algorithms centred in the modelled ligand HSE. The Molegro program used the search algorithm Moldock optimiser and the ranking function Moldock score and search radius

of 10 Å, also centred in the ligand HSE. At the end, the molecules 2061164, 15967722 and 2123137 (Zinc codes) were selected in common by all the programs in all twelve screens; these compounds were subsequently given the abbreviations HS6, HS7 and HS9, respectively (Fig. 1), and acquired for *in vitro* assays.
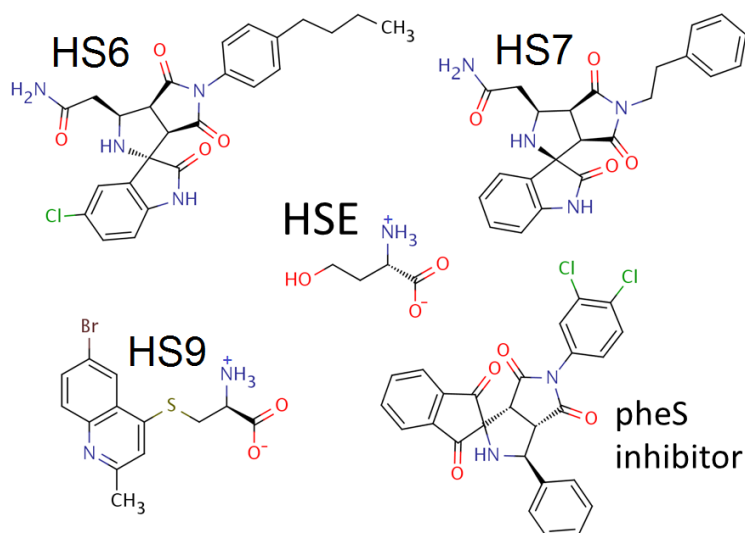

Fig. 1. Chemical structures of the molecules selected by the virtual screening and the structure of the substrate L-homoserine (HSE) and a known inhibitor of phenylalanine-tRNA ligase (pheS).

The values of the docking scores provided by the three programs used in the virtual screening for the identified ligands plus the reference ligand HSE are shown in Table 1.

Table 1 Docking scores provided by the programs used to identify the three ligands evaluated in this work. Values represent the mean of four simulations and the standard deviation.

Ligand	Autodock score	Molegro score	CLC score
HS6	-9.8 ± 0.19	-117.3 ± 0.06	-43.5 ± 1.07
HS7	-8.6 ± 0.24	-120.0 ± 0.15	-52.1 ± 0.57
HS9	-7.5 ± 0.15	-110.6 ± 1.73	-50.7 ± 0.94
HSE	-5.6 ± 0.07	-67.4 ± 0.98	-33.8 ± 1.02

Biological trials

The MIC assays showed that the three molecules evaluated were able to inhibit the growth of the *P. brasiliensis* isolate Pb18, where the molecule HS9 (MIC 8 µg.mL^{-1}) presented the best result, followed by the molecule HS7 (MIC 32 µg.mL^{-1}). The molecule HS6 also presented antifungal activity *in vitro*; however, with an MIC of

128 µg.mL^{-1}, this was not considered a sufficiently promising result by our group. The assay for the determination of the MFC showed that the values found were the same as those observed in the MIC assays (Table 2). Thus, the ratio of the MIC and MFC values equals 1, suggesting a fungicidal profile for these three molecules [31].

Table 2. Minimum inhibitory and fungicidal concentration of the molecules selected by virtual screening against *Paracoccidioides brasiliensis*.

Ligand	MIC µg.mL^{-1}	MFC µg.mL^{-1}
HS6	128	128
HS7	32	32
HS9	8	8

The fact that all three identified molecules presented antifungal activity is quite interesting. We attributed this result to the rigor of the adopted selection criterion, which eliminated false positive results and highlighted the most promising ones. Among the tested molecules, HS6 showed the lowest solubility in the assays. This characteristic may have contributed to the high MIC value, since molecules with low solubility may not be absorbed by the cell [32]. The water-octanol partition coefficients (*x*logP) calculated at pH 7.0 for the selected molecules were 0.48 (HS9), 1.89 (HS7) and 4.30 (HS6). The higher the *x*logP value, the lower the solubility in water; this may have influenced the results by different ways [33,34].

Molecule HS6

The molecule HS6 (Zinc20611644; InterBioScreen STOCK1N-58514, 92% purity by NMR) has a neutral charge at pH 7.0, which, in theory, would facilitate its permeability through the membrane [35]. However, its poor solubility in aqueous media may have reduced its activity in the medium. So far, it has no patent or known applicability. The analysis by the SEA predictions server, which compares the structure of the query molecule with a database of molecules that are shown to bind to a given protein/enzyme, suggested that HS6 bears some resemblance to the ligands of the alpha subunit of phenylalanine-tRNA ligase (pheS), which is also present in fungi of the genus *Paracoccidioides* (Uniprot id: c1gjz6), with a score of 0.29 and *p*-value of 3.367×10^{-25}, which is considered a low probability [36]. In fact, the molecule HS6 bears some resemblance to known inhibitors of pheS (Fig. 1), with a Tanimoto index ranging from 0.28 to 0.30. Thus, although the molecule HS6 was discovered by docking with

*Pb*HSD as a target, we evaluated the docking of HS6 and HS7 on the active site of the pheS alpha subunit as well (PDBid: 3pco). The results did not present favourable interactions and generated prohibitive scores.

Molecule HS7

The molecule HS7 (Zinc15967722; InterBioScreen STOCK1N-61006, 92% purity by NMR) also has a neutral charge at pH 7.0 and, to date, has no patent or known applicability. The SEA predictions server did not indicate a similarity between this molecule and any other known ligand that could act on another enzyme.

The molecules used in the library were not selected by the resemblance of their structure with a known ligand of *Pb*HSD, but because of their origin as natural products. The virtual screening simulation selected only three molecules out of 112,572 in the library. Because of this small number, all three molecules were acquired for *in vitro* assays regardless of their structural similarity.

The molecules HS6 and HS7 have in common the core *2-[(3S,3'S,3'aS,6'aR)-2,4',6'-trioxo-1,2,3',3'a,4',5',6',6'a-octahydro-2'H-spiro[indole-3,1'-pyrrolo[3,4-c]pyrrole]-3'-yl]acetamide*, also known as *spiro[pyrrolidine-3, 3´-oxindole]*. The Tanimoto index between them is 0.88, indicating high identity. Of the 388 molecules selected in the first screening, at least 19 molecules were identified with this core. This means that a new virtual screening from a library of molecules with the common core *spiro[pyrrolidine-3, 3´-oxindole]*, that are similar to HS6 and HS7, may identify other more promising molecules. Neither molecule HS6 nor HS7 have patents deposited, but other similar molecules that have the *spiro[pyrrolidine-3, 3´-oxindole]* core have recently been described as potent drugs against breast cancer [37].

Molecule HS9

The molecule HS9 (Zinc2123137; InterBioScreen STOCK5S-99718, 90% purity by mass) has net charge 0 at pH 7.0 and, to date, has no patent or known applicability. The analysis by the SEA predictions server indicated similarity (p-value: 1.315×10^{-53}, score 0.42) with inhibitors of the enzyme cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A from mouse (Uniprot: q8ca95) and similarity (p-value: 1.247×10^{-52}, score 0.38) with inhibitors of proline-tRNA ligase, a cytoplasmic protein in *Candida albicans* (Uniprot: p78600). However, molecule HS9 was not active against

Candida albicans and had no cytotoxic effect on mammalian cells, which discards these enzymes as alternative targets, i.e. acting as pan-assay interference compound (PAINS).

Cytotoxicity assays were performed with the most promising molecules HS7 and HS9 in the mammalian HeLa and Vero cell lineages (Fig. 2). The HS7 molecule was not toxic to the HeLa cells up to the tested concentration of 1,024 $\mu g.mL^{-1}$, indicating a therapeutic range of at least 992 $\mu g.mL^{-1}$, more than 32 times the MIC value. In the Vero cell line, the concentration of 1,024 $\mu g.mL^{-1}$ was shown to be significantly toxic (p<0.05), indicating an interval of at least 480 $\mu g.mL^{-1}$, which is equivalent to 16 times the MIC value. Since the therapeutic window is calculated according to the LD_{50} not by the MIC [38], this window may be slightly larger than the range shown above. However, we must consider that, because of the solubility and quantity of compounds tested, it was not possible to determine the value of the LD_{50}, since concentrations higher than 1,024 $\mu g.mL^{-1}$ were not possible to prepare.
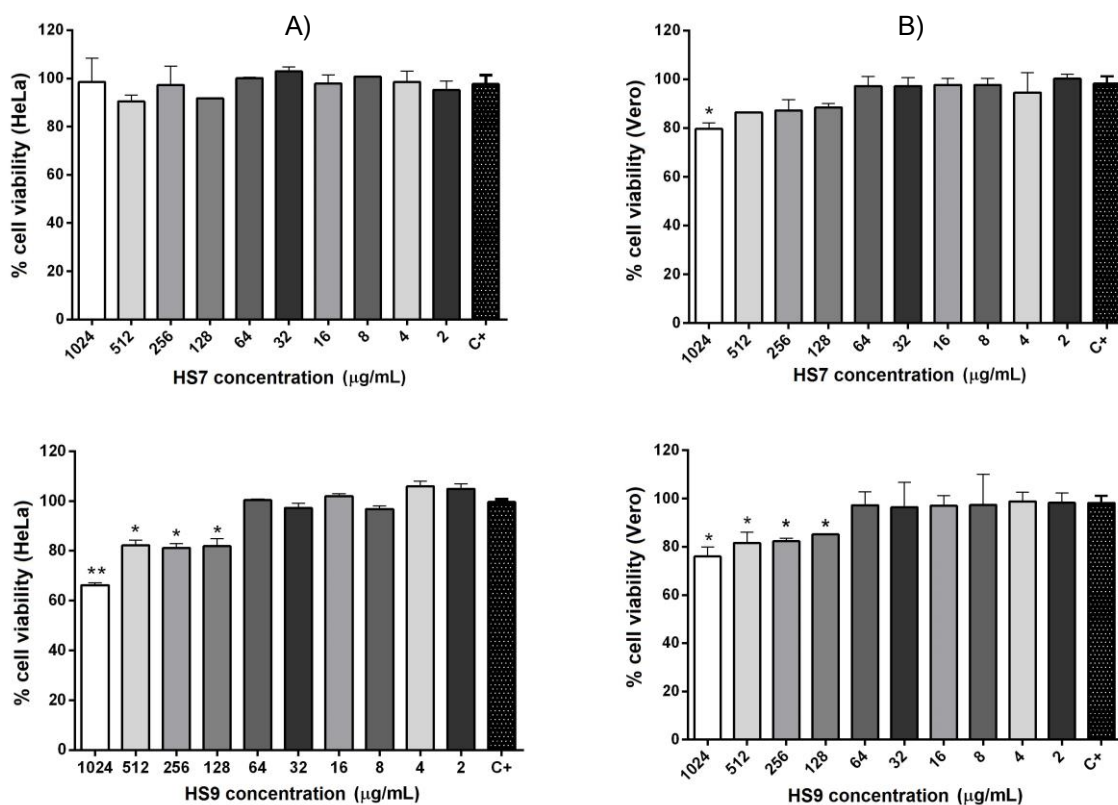
Fig. 2. *In vitro* cytotoxicity in HeLa (**a**) and Vero cells (**b**) in the presence of the molecules HS7 and HS9. The tested concentrations ranged from 2 to 1,024 $\mu g.mL^{-1}$, and cell viability was assessed by determining mitochondrial activity using a CellTiter 96 AQueous One Solution cell proliferation assay (which contains MTS) after 48 h of incubation. The percentage of viable cells after exposure to the compounds at different concentrations was evaluated by considering the viability of untreated cells to be 100%. The experiments were performed in triplicate. (*) indicates a significant reduction in cell viability (p<0.05).

The HS9 molecule presented the same toxicity profile for the two cell lines after 128 μg.mL^{-1}. The interval between the MIC value and the minimum cytotoxic dose was at least 120 μg.mL^{-1}, demonstrating a therapeutic range smaller than HS7.

To better understand the nature of the interaction of the selected molecules with the target enzyme, MD simulations were performed in comparison to the HSE substrate. Fig. 3 depicts the behaviour of the biological unit (homodimer) of *Pb*HSD in the presence of the three ligands mentioned, plus the cofactor NAD$^+$, over time. The rmsd value in the last 10 ns of the simulation remained stable, forming a plateau, in all evaluated complexes (Fig. 3a), which indicates that the system entered into thermodynamic equilibrium, allowing this interval to be used to analyse the interactions that each ligand makes with the enzyme. Fig. 3b shows that the radius of gyration of the complexes oscillated very little, i.e. within one Ångstrom, throughout the simulation. This indicates that the protein maintained its folded state, and the presence of the ligands did not allow the protein to unfold. Fig. 3c shows that the greatest oscillations in the protein occurred in the N-terminal region, which was expected, since the first residues do not form any ordered element of secondary structure. When comparing the rmsf of the three complexes, it was possible to observe that none of the ligands caused greater oscillations in a given region of the protein. By means of visual inspection of the trajectory, it was possible to verify that the ligands remained within the catalytic site of the enzyme throughout the simulation (Fig. S1, supplementary data).

The ligands have distinct structures, which promote distinct interactions with the protein (Figure S2, supplementary data). Because of this, the most important residues in the protein for docking are those with a higher frequency of contact. The frequency of contact that the protein residues made with each ligand is shown in Table S1 (Supplementary Material). In this analysis, it is considered the contact that each heavy atom of a given ligand makes with a given protein residue, within a defined distance of up to 4.0 Å. Next, a statistics was made regarding the number of times that this contact appears in the frames, being total number of frames = 1.00 (100%). Table S1 indicates that Asp229 was essential for the binding of all evaluated ligands. Tyr216 made contact with all inhibitors, suggesting a specific anchoring mechanism for these ligands. The high frequency of contact with the cofactor NAD$^+$ indicates that the ligands bind to the active form of the enzyme, which is essential for antifungal activity since there would be no point in inhibiting an inactive enzyme. Table S1 also demonstrates that molecule HS9 makes 14 contacts with a frequency greater than 70% with protein residues,

including the cofactor NAD^+. This suggests a common mechanism of interaction of these ligands with the protein, that starts with ligand entry into the catalytic site and interaction with residues in common. In this case, the Asp224 and Asp229 by means of charge-charge or charge-dipole interactions (depending on the ligand) and with Tyr216 and NAD^+ by Van der Waals contacts.

There are many aspects involving the biological activity of a compound such as solubility, cell permeability, target affinity, etc. Molecules HS6 and HS7 make nine contacts each, while the HSE substrate makes only five contacts. This is certainly reflected in the different affinities of the enzyme for each ligand and may explain in part, the *in vitro* inhibition results (MIC and MFC). Coincidence or not, the ligand with the highest number of contacts (HS9) had the lowest MIC, whereas the ligand with the lowest number of contacts and lowest solubility in water (HS6) had the highest MIC. However, the theoretical affinity of a protein for a ligand can only be estimated by means of calculations, for example using MM-PBSA [39] or LIE [40] methods.

Details of the electrostatic contacts involving the protein residues with each ligand identified by virtual screening are shown in Table 3. The full list of contacts which includes van der Waals interactions are presented in Table S1 (Supplementary Material).

Table 3 Details of electrostatic contacts between protein residues and the respective ligand atoms.

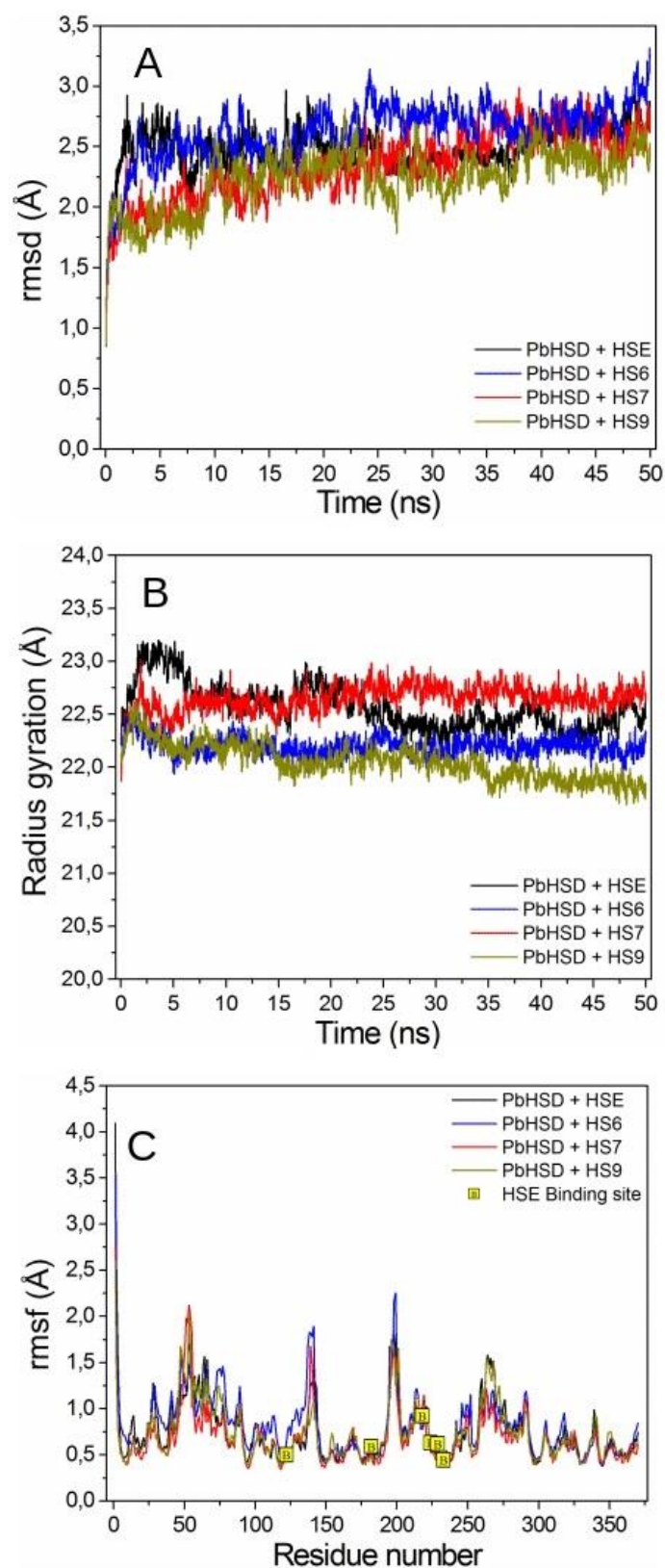| Protein residue	HS6		HS7		HS9	
	Atom	Distance (Å)	Atom	Distance (Å)	Atom	Distance (Å)
Lys122 Nζ	O1	3.01	-	-	O1	2.80
Gly215 O	N3	3.99	-	-	-	-
Glu218 Oε2	N2	3.34	-	-	N2	2.70
Asp224 Oδ2	N4	2.73	N4	2.84	N2	2.65
Asp229 Oδ2	N4	2.77	N4	2.83	N2	4.00
Lys233 Nζ	-	-	O4	2.87	O2	2.07

Fig. 3. Behaviour of the main chain atoms of *Pb*HSD amino acids throughout the simulation in terms of rmsd (**a**) and radius of gyration (**b**). The rmsf of the Cα of each residue in the last 10 ns of simulation (**c**). The yellow boxes highlight the position of the most important residues for the docking of the HSE substrate.

Molecules HS6, HS7 and HS9 were screened into a library of 112,572 compounds. The chances of these three molecules having been discovered at random and demonstrating fungicidal activity while having another enzyme as the target would be $(1/112,572)^3$. Therefore, it is most likely that these molecules are in fact ligands of *Pb*HSD. However, definitive confirmation should be obtained with spectroscopic binding studies of these molecules with recombinant *Pb*HSD. For this goal, heterologous expression experiments are already underway to obtain the purified enzyme to perform kinetic assays.

The literature reports other HSD inhibitors that are amino acids or derivatives in their chemical nature. The amino acid cysteine acts as competitive inhibitor of HSD from the hyperthermophilic archaeon *Sulfolobus tokodaii* with a K_i of 11 μM and as uncompetitive inhibitor of NAD^+ and NADP with K_i values of 0.55 and 1.2 mM, respectively [41]. The amino acid derivatives HS1 and HS2 were detected by virtual screening and showed *in vitro* fungiostatic activity against *Paracoccidioides brasiliensis*, with an MIC of 64 and 32 μg·mL^{-1}, respectively [17]. The amino acid derivative *5-hydroxy-4-oxo-l-norvaline* (HONV) incorporated into dipeptides was able to inhibit the growth of yeasts of the *Candida* genus *in vitro*, with MIC values ranging between 32 and 64 μg.mL^{-1} [42].

Since HSD is an enzyme that participates in the process of essential amino acid biosynthesis, it is to be expected that certain amino acids or derivatives may inhibit it in some way. In this sense, the molecule HS9 is an amino acid derivative and had the lowest reported MIC value for the HSD enzyme of *Paracoccidioides*. The identification of non-amino acid inhibitors is a major challenge. In this sense, a series of bis-sulphanylalkylphenol molecules with IC$_{50}$ values ranging between 2 to 40 μM with activity against strains of *Candida albicans* and *Candida parapsilosis* were the first inhibitors of this category described [18].

The MIC values found in this work for molecules HS6 and HS7 (non-amino acid inhibitors) are similar to the values found in the studies cited above. The differences are related to their fungicidal profile and the relatively low cytotoxicity of the molecule HS7 to human cell lines, which provides a greater therapeutic range than amphotericin B [43], the reference fungicide for the treatment of PCM.

Conclusions

This work identified three molecules with activity against the most virulent strain of *Paracoccidioides brasiliensis* (Pb18), using virtual screening techniques. The compounds HS6 (Zinc2061164), HS7 (Zinc15967722) and HS9 (Zinc2123137) presented MIC/MCF values of 128, 32 and 8 µg.mL^{-1} respectively, presenting a fungicidal profile. The HS9 molecule was the most promising, since the MIC/MCF value found for it is lower than values recently reported in the literature for other inhibitor candidates of *Pb*HSD.

Molecular dynamics simulations of protein-ligand-cofactor complexes showed that ligands remained within the catalytic site throughout the simulation, and Tyr216 residue proved to be very important in stabilizing HS6, HS7 and HS9 on the site. The residue Asp219 was the most important in the stabilization of all ligands, including the substrate. The stability of them within the site depends on the presence of NAD cofactor configuring the active form of the enzyme.

By having fungicidal activity against *P. brasiliensis* and low cytotoxicity in human cell lines, the molecule HS9 (Zinc2123137) may be explored as a hit against paracoccidioidomycosis.

References

1. Martinez R (2015) Epidemiology of Paracoccidioidomycosis. Rev Inst Med Trop Sao Paulo 57 (Suppl 19):11-20. doi:10.1590/S0036-46652015000700004

2. Shikanai-Yasuda MA, Mendes RP, Colombo AL et al (2017) Brazilian guidelines for the clinical management of paracoccidioidomycosis. Rev Soc Bras de Med Trop 50:715-40. doi:10.1590/0037-8682-0230-2017

3. Mendes RP, Cavalcante RS, Marques SA et al (2017) Paracoccidioidomycosis: Current perspectives from Brazil. The open microb j 11:224-82. doi:10.2174/1874285801711010224

4. Martinez R (2017) New Trends in Paracoccidioidomycosis Epidemiology. J Fungi 3. doi:10.3390/jof3010001

5. Pecanha PM, de Souza S, Falqueto A et al (2016) Amphotericin B lipid complex in the treatment of severe paracoccidioidomycosis: a case series. Int J Antimicrob Agents 48:428-30. doi:10.1016/j.ijantimicag.2016.06.011

6. Ghannoum MA, Rice LB (1999) Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. Clin Microbiol Rev 12:501-17.

7. Blau IW, Fauser AA (2000) Review of comparative studies between conventional and liposomal amphotericin B (Ambisome) in neutropenic patients with fever of unknown origin and patients with systemic mycosis. Mycoses 43:325-32. doi:10.1046/j.1439-0507.2000.00577.x

8. Abadio AK, Kioshima ES, Teixeira MM et al (2011) Comparative genomics allowed the identification of drug targets against human fungal pathogens. BMC genomics 12:75. doi:10.1186/1471-2164-12-75

9. Sagatova AA, Keniya MV, Wilson RK et al (2015) Structural insights into binding of the antifungal drug fluconazole to *Saccharomyces cerevisiae* lanosterol 14alpha-demethylase. Antimicrob Agents Chemother 59:4982-9. doi:10.1128/AAC.00925-15

10. Cunha BA (2001) Antibiotic side effects. Emerg Med Clin North Am 85:149-85. doi:10.1016/s0025-7125(05)70309-6

11. Verweij PE, Snelders E, Kema GH et al (2009) Azole resistance in *Aspergillus fumigatus*: a side-effect of environmental fungicide use? Lancet Infect Dis 9:789-95. doi:10.1016/S1473-3099(09)70265-8

12. Hahn RC, Morato Conceicao YT, Santos NL et al (2003) Disseminated paracoccidioidomycosis: correlation between clinical and *in vitro* resistance to ketoconazole and trimethoprim sulphamethoxazole. Mycoses 46:342-7. doi:10.1046/j.1439-0507.2003.00901.x

13. Donnici CL, Nogueira LJ, Araujo MH et al (2014) *In vitro* studies of the activity of dithiocarbamate organoruthenium complexes against clinically relevant fungal pathogens. Molecules 19:5402-20. doi:10.3390/molecules19045402

14. Pfaller MA (2012) Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. Am J Med 125:S3-13. doi:10.1016/j.amjmed.2011.11.001

15. Calich VL, Kashino SS (1998) Cytokines produced by susceptible and resistant mice in the course of *Paracoccidioides brasiliensis* infection. Braz J Med Biol Res 31:615-23. doi:10.1590/S0100-879X1998000500003

16. Jacques SL, Nieman C, Bareich D et al (2001) Characterization of yeast homoserine dehydrogenase, an antifungal target: the invariant histidine 309 is important for enzyme integrity. Biochim Biophys Acta 1544:28-41. doi:10.1016/S0167-4838(00)00203-X

17. Bagatin MC, Pimentel AL, Biavatti DC et al (2017) Targeting the homoserine dehydrogenase of Paracoccidioides species for treatment of systemic fungal infections. Antimicrob Agents Chemother 61. doi:10.1128/AAC.00165-17

18. Ejim L, Mirza IA, Capone C et al (2004) New phenolic inhibitors of yeast homoserine dehydrogenase. Bioorg Med Chem 12:3825-30. doi:10.1016/j.bmc.2004.05.009

19. Jastrzebowska K, Gabriel I (2015) Inhibitors of amino acids biosynthesis as antifungal agents. Amino acids 47:227-49. doi:10.1007/s00726-014-1873-1

20. Irwin JJ, Sterling T, Mysinger MM et al (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757-68. doi:10.1021/ci3001277

21. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785-91. doi:10.1002/jcc.21256

22. Wolf LK (2009) New software and websites for the chemical enterprise. Chem & Eng News 87:32. doi:10.1021/cen-v087n005.p032

23. Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 49:3315-21. doi:10.1021/jm051197e

24. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-802. doi:10.1002/jcc.20289

25. Mackerell AD, Jr., Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400-15. doi:10.1002/jcc.20065

26. Zoete V, Cuendet MA, Grosdidier A (2011) SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 32:2359-68. doi:10.1002/jcc.21816

27. Neese, F (2012) The ORCA program system. Comp Mol Sci 2(1):73-78. doi:10.1002/wcms.81

28. Rodrigues-Vendramini FAV, Marschalk C, Toplak M, et al. (2018) Promising new antifungal treatment targeting chorismate synthase from Paracoccidioides brasiliensis. Antimicrob Agents Chemother 63:e01097-18. doi:10.1128/AAC.01097-18

29. Netto CF, Vegas VS, Sciannamea IM et al (1969) The polysaccharidic antigen from *Paracoccidioides brasiliensis*. Study of the time of cultivation necessary for the preparation of the antigen. Rev Inst Med Trop Sao Paulo 11:177-81. PMID:5824780

30. Strober W. (2015) Trypan blue exclusion test of cell viability. Curr Protoc Immunol 111:A3.B.1-3. doi:10.1002/0471142735.ima03bs111

31. Sun N, Li D, Zhang Y et al (2017) Repurposing an inhibitor of ribosomal biogenesis with broad anti-fungal activity. Sci Rep 7:17014. doi:10.1038/s41598-017-17147-x

32. Ishikawa M, Hashimoto Y (2011) Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. J Med Chem 54:1539-54. doi:10.1021/jm101356p

33. Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods 44:235-249. PMID:11274893

34. Fahr A, Liu X (2007) Drug delivery strategies for poorly water-soluble drugs. Expert Opin Drug Deliv 4:403-416. doi: 10.1517/17425247.4.4.403

35. Yang NJ, Hinner MJ (2015) Getting across the cell membrane: an overview for small molecules, peptides, and proteins. Methods Mol Biol 1266:29-53. doi:10.1007/978-1-4939-2272-7_3

36. Keiser MJ, Roth BL et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197-206. doi:10.1038/nbt1284

37. Hati S, Tripathy S, Dutta PK et al (2016) Spiro[pyrrolidine-3, 3 -oxindole] as potent anti-breast cancer compounds: Their design, synthesis, biological evaluation and cellular target identification. Sci Rep 6:32213. doi:10.1038/srep32213

38. Di, L Kerns, E (2008) Drug-like properties: concepts, structure design and methods from ADME to toxicity optimization. 1st ed. Academic press.

39. Kumari R, Kumar R, Lynn A, Consort OSDD (2014) g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54 (7):1951−1962. doi:10.1021/Ci500020m

40. Gutierrez-de-Teran H, Aqvist J (2012) Linear interaction energy: method and applications in drug design. Methods Mol Biol 819:305−323. doi:10.1007/978-1-61779-465-0_20

41. Ogata K, Yajima Y, Nakamura S et al (2018) Inhibition of homoserine dehydrogenase by formation of a cysteine-NAD covalent complex. Sci Rep 8:5749. doi:10.1038/s41598-018-24063-1

42. Skwarecki AS, Schielmann M, Martynow D et al (2018) Antifungal dipeptides incorporating an inhibitor of homoserine dehydrogenase. J Pept Sci 24. doi:10.1002/psc.3060

43. Andes D, Stamsted T, Conklin R (2001) Pharmacodynamics of amphotericin B in a neutropenic-mouse disseminated-candidiasis model. Antimicrob Agents Chemother 45:922-6. doi:10.1128/AAC.45.3.922-926.2001

Supplementary Materials

New inhibitors of homoserine dehydrogenase from *Paracoccidioides brasiliensis* presenting antifungal activity

Paulo Sérgio Alves Bueno[1], Franciele Abigail Vilugron Rodrigues[2], Débora Carina Biavatti[1], Arethusa Lobo Pimentel[1], Mariane Cristóvão Bagatin[3], Érika Seki Kioshima[2], Gisele de Freitas Gauze[3], Flavio Augusto Vicente Seixas[1]*.

[1]Department of Technology, Universidade Estadual de Maringá, Umuarama, PR, Brazil.
[2]Department of Clinical Analysis and Biomedicine, Universidade Estadual de Maringá, Maringá, PR, Brazil.
[3]Department of Chemistry, Universidade Estadual de Maringá, Maringá, PR, Brazil.

*Correspondence author
favseixas@uem.br
Department of Technology.
Universidade Estadual de Maringá
Av. Ângelo Moreira da Fonseca, 1800
87506-370, Umuarama, PR, Brazil.
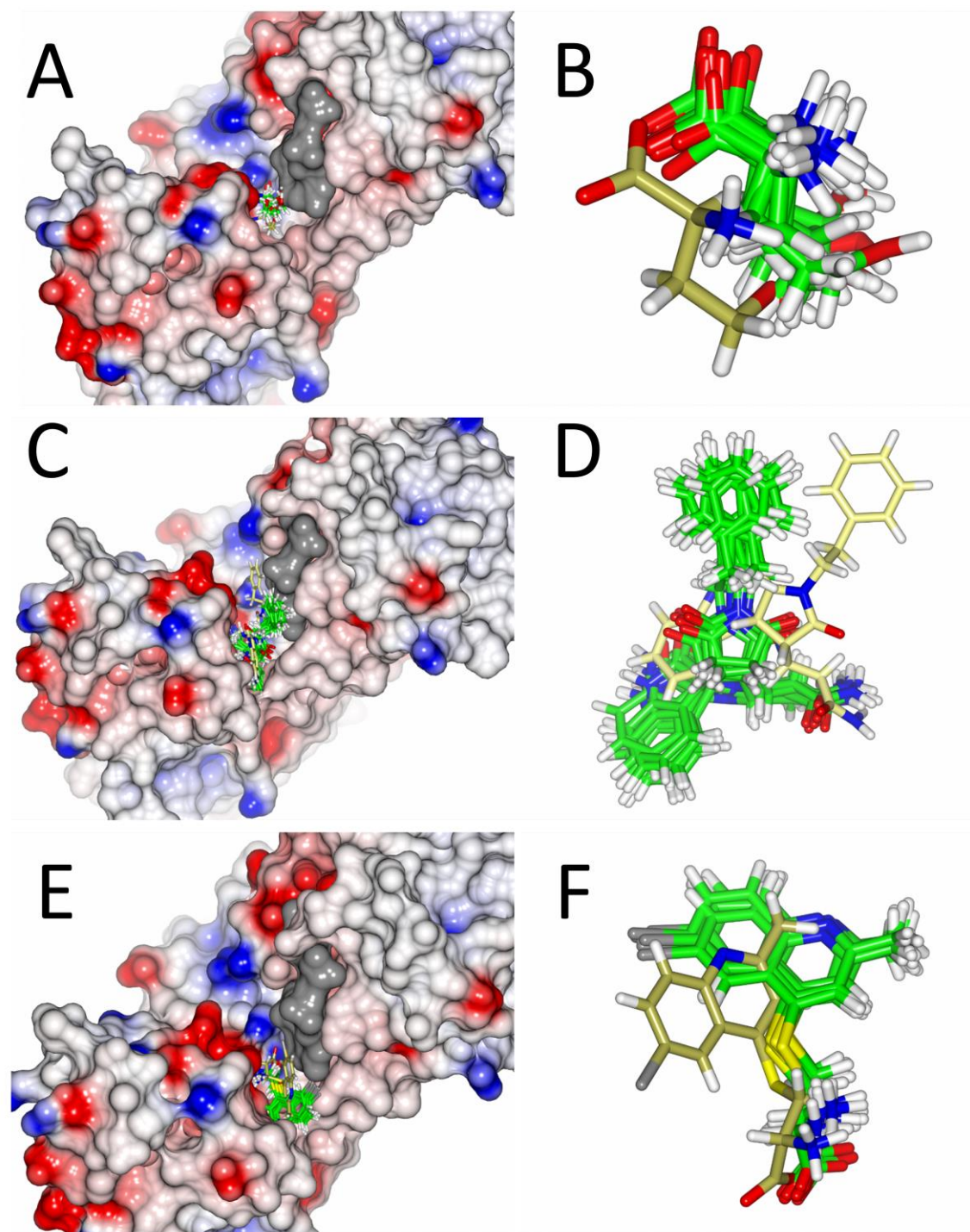
Fig. S1. The representations in (**A, C, E**) shows that the main ligands evaluated in this study remained within the catalytic cavity of *Pb*HSD along simulations. The gray surface in these figures represents the NAD$^+$ cofactor. In **B, D, F**, the protein surface was removed for better view. The gold-colored structures represent the initial position of the ligands HSE (**B**), HS7 (**D**) and HS9 (**E**) obtained by docking at the beginning of the simulation. The green structures represent the positions of the ligands in the last 10 frames (100 ps) of the simulations. See that all ligands moved regarding to their initial position, however, they have found a place to better interact with the protein and remain in this position until the end. This is clear when we see a cluster of close conformations at the last stages of the simulation (equilibrium reached). Figures generated using CCP4mg program.

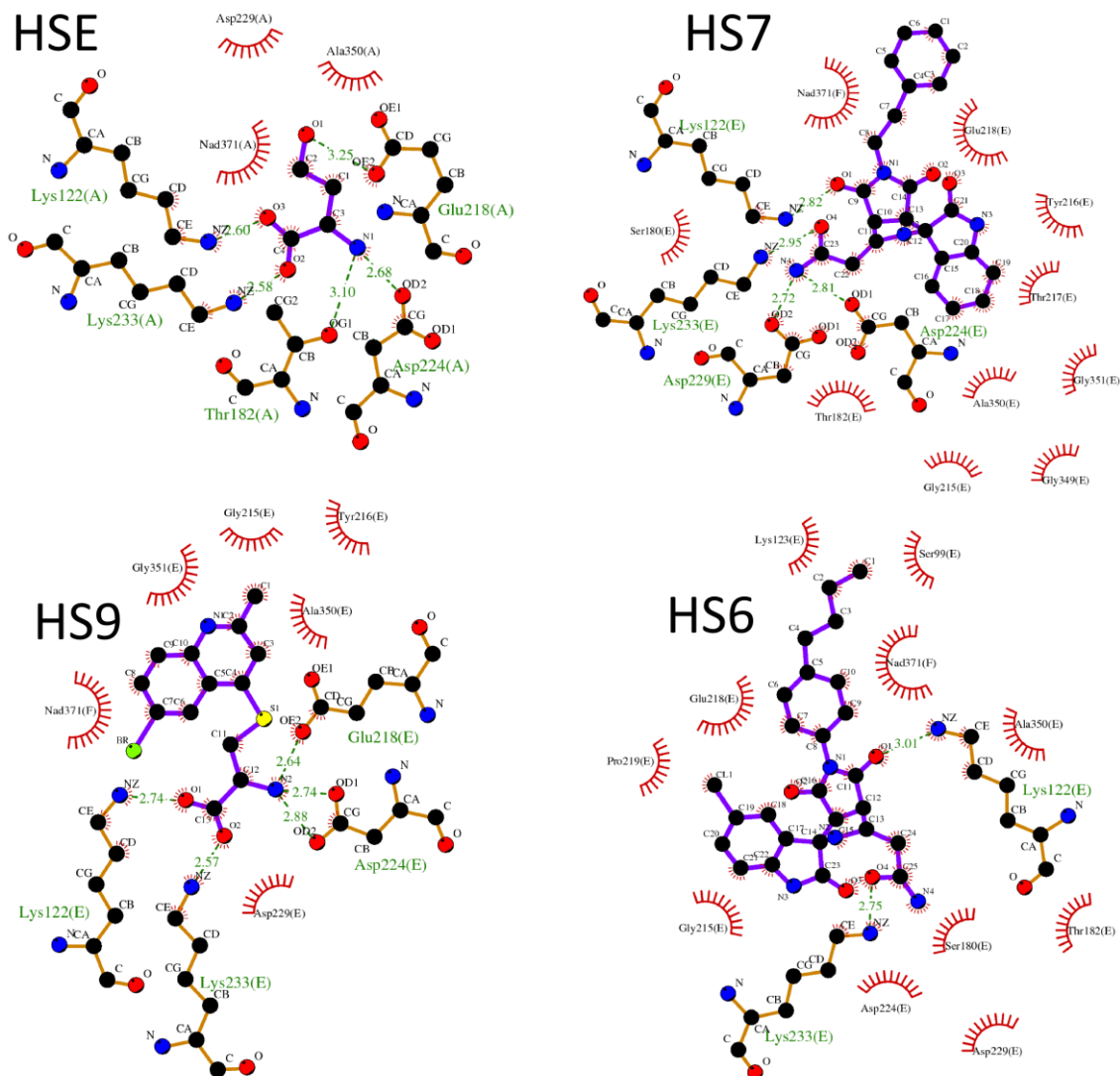**Fig. S2.** Schematic representation of the interaction of ligands with *Pb*HSD. Figure generated using LigPlot[+] program.

Table S1. Contact frequency between *Pb*HSD residues and ligands. The common residues with a higher frequency of contact are in bold. The contacts in bold letters had frequency greater than 0.7 with all the ligands, indicating a common docking mechanism. Blue contacts had a frequency greater than 0.9, indicating a docking mechanism specific to inhibitors

Residue	HSE (substrate)	HS6 58	HS7 61	HS9 99
VAL-12	-	0.18	-	-
SER-99	-	0.65	-	-
LYS-122	0.89	0.78	-	0.97
ALA-154	-	-	-	0.17
PHE-179	0.14	-	0.31	0.62
SER-180	0.21	-	0.93	0.96
GLY-181	0.14	-	0.87	0.84
THR-182	0.16	0.89	0.75	0.97
PHE-185	-	-	-	0.12
GLY-215	-	0.99	0.46	0.72
TYR-216	-	**0.99**	**0.92**	**0.99**
THR-217	-	1.00	-	0.71
GLU-218	0.75	1.00	0.23	0.99
PRO-219	-	0.60	-	0.21
ASP-223	0.35	-	-	-
ASP-224	0.48	0.97	0.52	0.94
ASP-229	**0.98**	**0.99**	**0.99**	**0.95**
LYS-233	0.98	0.22	0.89	1.00
SER-327	-	-	0.85	0.33
ASN-329	-	-	0.11	-
GLU-336	-	-	0.23	-
ARG-337	-	-	0.19	-
TYR-338	-	-	0.10	-
GLY-339	-	-	0.43	-
LYS-340	-	-	0.24	-
ASN-341	-	-	0.17	-
GLY-349	0.45	-	0.61	0.89
ALA-350	0.17	0.10	0.96	0.96
NAD-371	**0.70**	**1.00**	**0.87**	**1.00**

Artigo 2: *In silico* evaluation of condensed and hydrolysable tannins as inhibitors of pancreatic α-amylase.

In silico evaluation of condensed and hydrolysable tannins as inhibitors of pancreatic α-amylase

Paulo Sérgio Alves Bueno[1], Camila Gabriel Kato-Schwartz[1,2], Diego de Souza Lima[3], Adelar Bracht[1,2], Rosane Marina Peralta[1,2], Flavio Augusto Vicente Seixas[3].

[1]Department of Biochemistry. Universidade Estadual de Maringá, Maringá, PR, Brazil.
[2]Post-graduate program of Food Science. Universidade Estadual de Maringá, PR, Brazil
[3]Department of Technology, Universidade Estadual de Maringá, Umuarama, PR, Brazil

* Corresponding author
Universidade Estadual de Maringá, UEM
Department of Technology
Av. Angelo Moreira da Fonseca, 1800, Umuarama, PR, Brazil, 87506-370.
Phone: +55 44 3621 9300
e-mail: favseixas@uem.br

Abstract

Amylases are interesting targets for antidiabetic drugs because their inhibition is able to lower glycemia without the need of hormonal control, as promoted by insulin or glibenclamide. In this context, the comparison between the binding features of α-amylases with their substrate and known inhibitors, may provide insights aiming at the discovery of new antidiabetic drugs. In this work, the structure of the porcine pancreatic α-amylase was modelled with the acarbose pentasaccharide inhibitor, and used in structure-based virtual screening simulations based on a library containing the structures of amylose (AMY), acarbose (ACA) and the more representative structures of condensed tannin (CTN) and hydrolysable tannin (HTN). After validation of the methodology by redocking (mean rmsd ~0.8 Å), the scores provided by programs AutoDock/Molegro were contradictory ($-1.5/-23.3$; $-3.5/-24.6$; $-4.3/-14.6$; $-/-19.5$ for AMY, ACA, CTN and HTN respectively), indicating that a more sensitive methodology was necessary. The $\Delta G_{binding}$ was calculated by the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method, which indicated that the HTN, ACA and CTN had higher affinities for the enzyme regarding the AMY substrate, with values of -350.0, -346.2, -320.5 and -209.2 kJ·mol^{-1}, respectively. The predicted relative affinities of HTN and CTN are in agreement with those obtained experimentally. The results provided useful information for the characterization of tannin binding to α-amylase, which can be applied in future studies aiming at finding new hypoglycemic molecules among natural products.

Key words: enzyme, diabetes, natural products, mm-pbsa, docking.

1. Introduction

Alpha-amylase (E.C. 3.2.1.1) belongs to the alpha-glucosidase family that hydrolyses α-1,4-glycosidic bonds present in starch, glycogen and other oligo- and polysaccharides. The human body produces two α-amylases, salivary and pancreatic, which are involved in the digestion process, with consequent production of glucose, maltose, and oligosaccharides [1]. The presence of these fermentable sugars in the oral cavity is related to the development of dental caries [2], whereas the absorption of monosaccharides into the bloodstream leads to an increase in glycemia, which is related to metabolic disorders especially obesity and type 2 diabetes [3]. The prevalence of type 2 diabetes is increasing worldwide. A recent report from the International Diabetes Federation (IDF; 2017) reveals that diabetes affects approximately 425 million people worldwide [4,5] This represents approximately 8.8% of the world population and there are projections indicating that by 2045, 628 million people will have diabetes. Type 2 diabetes presents the highest incidence with 85% of the total prevalence of diabetes mellitus [4,5]. An appropriate response to these projections would be an intensified search for new agents and approaches. In this respect, α-amylases have been extensively investigated as potential targets for antidiabetic drugs [6]. The advantage of this target is that its inhibition does not involve mechanisms of hormonal regulation, as occurs with insulin and glibenclamide, whose continuous use may lead to the development of tolerance to these drugs [7,8].

Drugs such as acarbose, voglibose, and miglitol have been described as inhibitors of alpha-glucosidases, acting in the lumen of the intestine. Compared with acarbose, however, miglitol is systemically absorbed [9]. In the intestine, the normal flora acts on these drugs, especially on acarbose, generating new molecules that cause discomforts such as flatulence and moderate diarrhoea [10]. To avoid these collateral effects, natural products identified as inhibitors of alpha-glucosidases may be a good option for the reduction of glycemia without propitiating such undesirable adverse effects [11]. Among these products are the flavonoids, polyphenolics, cinnamic acid derivatives, terpenes and tannins, all found in plant extracts [12].

The tannins are polyphenols with molecular weights varying between 500 and 3,000 Da. They are relatively large molecules, so that they are not absorbed into the bloodstream, and may induce a local inhibition of alpha-amylase, resulting in a hypoglycemic effect. The tannins comprise two groups, hydrolysable and condensed. Among the hydrolysable tannins, one of the most studied is tannic acid (gallotannin) with up to 12 esterified galloyl groups and the core glucose which is found in the Chinese natural gallnuts [13]. Among the condensed tannins, the best known is that one extracted from the bark of the black wattle tree (*Acacia mearnsii*), rich in the catechin-like flavan-3-ol monomers robinetinidol and fisetinidol [14]. In a preceding

work we have characterized the kinetics of the α-amylase inhibitory activity of both the hydrolysable tannin from Chinese natural gall and the condensed tannin from *Acacia nearnsii* [15]. We also found that both compounds are able to reduce the post-prandial hyperglycemia *in vivo* after starch administration [15]. Kinetic studies, however, do not reveal mechanistic details about the inhibitory activity at the molecular level. Taking this into consideration, the objective of this work was to compare, through computational simulations, the binding characteristics of condensed and hydrolysable tannins to porcine pancreatic α-amylase in order to get mechanistic insights into the experimentally detected differences in the behaviour of both classes of compounds.

2. Material and methods

2.1. Protein ligand modelling and model minimization

The structure of porcine pancreatic alpha-amylase (PDBid 3l2l) at 2.11 Ångström resolution [16] was used as the target in the simulations. Its glycidyl ligands were removed, whereas the Ca^{2+} and Cl^- cofactors were maintained. Next, the structure of the human pancreatic alpha-amylase (PDBid: 1xd0) bound to acarbose pentasaccharide (ACA) [17] was overlapped with the porcine pancreatic alpha-amylase and the ACA ligand and the water molecules 511, 513 and 766 were copied in the resulting pose to the structure of the porcine pancreatic alpha-amylase (geometric docking).

The modeled complex of porcine alpha-amylase bonded to ligand ACA and cofactors Ca^{2+} and Cl^- was minimized by conjugate gradient (CG) using the NAMD2 program [18], due to the possibility of stereochemical collisions resulting from the geometric docking. The minimization of the system was done stepwise. In the first step, the protein-ligand complex was placed in a periodic box filled with Tip3 water and a sufficient amount of sodium counterions to neutralize the system charges, and subsequently it was minimized by 20,000 steps of conjugate gradient with the atoms of ligands and cofactors fixed in space. In the second step, the atoms of the protein and ligand were fixed in space and waters and ions submitted to 60 ps of molecular dynamics under NPT conditions. In the third and last step, the whole system was minimized again by new 20,000 steps of CG. The CHARMM C35b2-C36a2 force field [19] was used for simulation of protein, water and ions. The force field for ACA was generated by the Swissparam server [20] in the same format. The partial charges of the atoms of all ligands at pH 7.0 were calculated by the B3LYP/6-311G method by means of the ORCA program (https://orcaforum.cec.mpg.de/), in order to increase the accuracy of the ligand force field. The

minimized structure that arose from this third step was used as input parameter in the docking and long time molecular dynamics simulations.

2.2. Docking simulations

The docking simulations were carried out by means of two programs selected from five available in our laboratory. The choice of the programs as well as the search and ranking algorithms (when needed) were done by redocking the ACA ligand on the modeled structure. The protocol was chosen and considered validated when the best score pose provided an rmsd below 2.0 Å in all repetitions. The first program was Molegro-6.0 virtual docker [21] using 12 Å as search radius centralized at the ligand molecule. The Moldock Simplex Evolution was used as search algorithm adjusted to 50 runs, energy threshold as 1000, and 400 maximum steps. The Moldock Score [grid] algorithm was used for ranking. The second program was AutoDock-4.2.3 [22] implemented in the Pyrx-0.9 graphical interface [23]. The search box was centralized at the ligand, with grid dimensions of 50 at the x, y and z axes. The number of runs was increased to 50 due to the large number of rotatable bonds of the ligands and the number of energy evaluations was set to 2500000 (medium). Since the ligands have no acid or basic character (proton donor or acceptor), neutral charges at pH 7.0 were assumed, as suggested by Zinc database.

Pancreatic α-amylase has a pH optimum at 6.9 (neutral), so the degree of protonation chosen for all the protonatable residues was that for pH 7.0: 1 negative charge for Asp and Glu (charge -0.5 for both oxygen atoms of carboxyl groups); 1 positive charge for Lys and Arg and neutral charge for His (residue Hsd, proton on Nδ1).

2.3. ΔG$_{binding}$ calculations

The pose of the ligand found by docking was incorporated into the spatial coordinates and the entire system was minimized by 2,000 steps of steepest descent through the Gromacs-4.5.7 program [24]. The long time molecular dynamics simulations (20 ns) were performed using the same program at NPT conditions (300 K, 1 atm, pH 7.0, 150 mM NaCl) using the force field Gromos96 v43a1 [25]. The force fields of the ligands were generated in the same format by Prodrg2 server [26]. The Gibbs free energy of binding (ΔG$_{binding}$) was calculated by Molecular Mechanics Poisson-Boltzmann Surface Analysis method implemented in the g_mmpbsa program [27,28]. All analyses of the simulations were carried out from the trajectory file. Root mean square deviation (rmsd) and radius of gyration (r$_{gyr}$) were calculated from the main chain atoms of the protein after the alignment of the frames on the reference structure by

overlapping the protein backbone by means of the application "RMSD trajectory tool" available in the VMD program. The simulations took place in four nodes of an SGI Altix ICE 8400 LX running in parallel on 2 processors Intel Six Core 5680 of 3.33 GHz (48 cores) and 36 GB RAM, at CENAPAD/Unicamp, Brazil.

2.4. Compound library.

The library of compounds evaluated in this study includes known α-amylase ligands: an amylose tetrasaccharide (AMY), which is the enzyme's natural substrate; acarbose derived pentasaccharide (ACA), as a crystallographic template, and inhibitors: condensed tannin (CTN), as the most probable structure present in the extract of *A. mearnsii* [29] and the hydrolysable tannin, tannic acid (HTN), MW 1701.2 g/mol acquired from Sigma-Aldrich. The choice of the latter was due to the fact that this compound was isolated, what would allow greater control in the experimental trials.

3. Results and discussion.

In order to maintain the crystallographic structure of the pancreatic α-amylase (pdbid: 3l2l) in a docking-friendly conformation, especially at the catalytic site region, it was necessary to copy the structure of the water molecules 511, 513 and 766 due to the presence of hydrogen bonds between the acarbose pentasaccharide inhibitor (ACA) and some amino acid residues of the 1xd0 template structure. Calcium and chloride cofactors, original from structure 3l2l, were maintained for the same reason.

The structure of the inhibitor ACA was used in the redocking simulations on the porcine α-amylase-ACA complex in order to validate the docking protocol. The ligand overlap gave an average r.m.s.d. of 0.8 Å (Fig. 1), suggesting the protocol is valid and may be used in the docking of unknown molecules. The modeled structure of ACA has generally served as a template for the proper guidance in the selection of the better pose of unknown inhibitors.
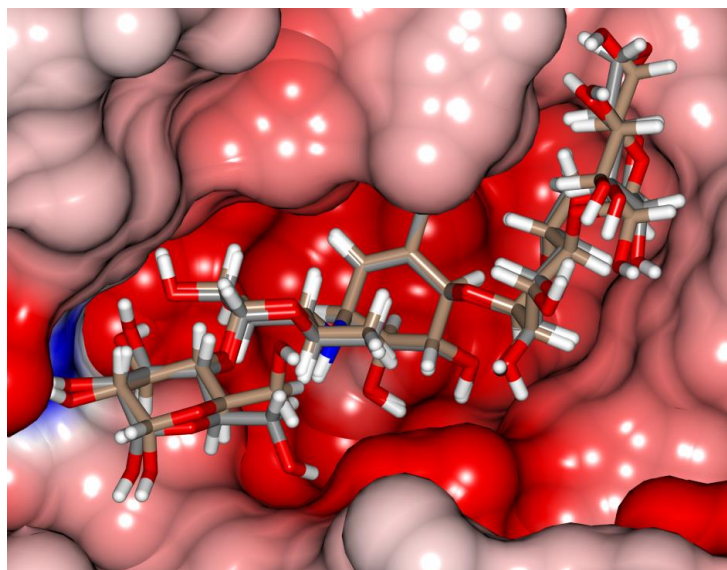
Fig. 1 Redocking of the ligand acarbose pentasaccharide (ACA) (brown) on the modeled complex α-amylase-ACA (gray), showing a very good overlay with an rmsd value of 0.8 Å. Figure generated using the CCP4mg program.

The redocking method consists in docking a ligand that is already present in a crystallographic structure (or modeled) in this same complex. If the program can reproduce the crystallographic pose observed in the complex and rank it as the best pose, it means that the program was able to "understand" the forces involved in the interaction of the protein with the ligand and will most likely be able to apply them for identification of a new molecule (unknown ligand) from a library. Redocking is a way of looking for the best program (or docking algorithm) to be applied to a particular target. The application of this method allowed to choose the programs AutoDock-4.2.3 and Molegro Virtual Docker 6.0, among the five that were evaluated, as those ones able to reproduce the pose of the ACA ligand in the form that it assumes in the model complex porcine α-amylase-ACA, with reproducibility in all attempts. The scores obtained with the redocking of the ACA ligand in the pig alpha-amylase using both programs are shown in Table 01. This procedure generates reliability in the identification of the best unknown ligand pose because if two different programs using distinct search and rank algorithms identify the same pose of the same ligand from a library of compounds, it is more likely that the selected ligand is, in fact, a true inhibitor.

A library with four compounds was used in the analysis of the porcine amylase inhibitors (Fig. 2) and the scores obtained in their docking to this target by means of the two validated programs are shown in Table 1.
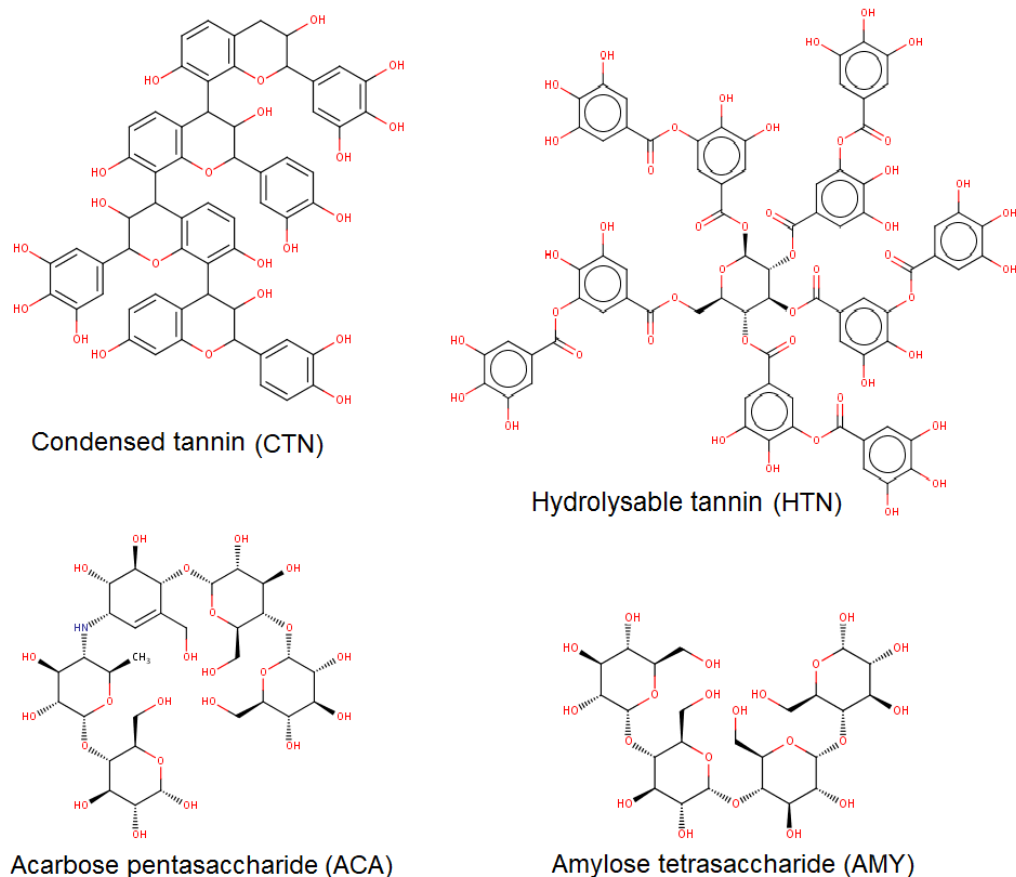
Fig. 2 Structures of the ligands used in this study.

The molecules HTN and CTN are the units that are repeated in the structures of the tannins, therefore, they were used in the simulations. Due to the relatively large number of rotational bonds present in the structures of the compounds, we increased the number of runs from 10 (default) to 50, the energy evaluations and other sensitive parameters of the Molegro's simplex evolution search algorithm, so that reproducibility was obtained in all simulations. Even so, the AutoDock program failed to anchor the hydrolysable tannin (HTN) ligand because its number of rotatable bonds exceeds the limit of 32 set by the pertinent search algorithm. The conformational search was considered sufficient, since each simulation produced a number of similar poses (cluster), as shown in Fig. S1 (supplementary material). In addition, these clusters were also found in different simulations. In this way, we chose the pose that we believe best represents the cluster, so that it could be used in molecular dynamics simulations

Table 1. Scores of the different ligands obtained by docking to the porcine α-amylase and free energy of binding (ΔG$_{binding}$) calculated from the best pose found for each ligand.

Ligand	Docking Score		ΔG$_{binding}$ (mm-pbsa) kJ·mol^{-1}
	AutoDock	Molegro	
Hydrolysable tannin (HTN)	–	−19.5	−350.0 ± 34.05
Acarbose pentasaccharide (ACA)*	−3.5	−24.6	−346.2 ± 35.77
Condensed tannin (CTN)	−4.3	−14.6	−320.5 ± 27.55
Amylose tetrasaccharide (AMY)	−1.5	−23.3	−209.2 ± 43.82

* The scores are the same for docking and redocking of this ligand.

The AutoDock and Molegro program scores shown in Table 1 are the mean values of four simulations for each ligand. Because they are relatively low and close values, it is difficult to establish an affinity ranking for each ligand. These results indicate, thus, that the use of more precise methods is necessary for the calculations.

The programs AutoDock and Vina give an energy dimension, kcal·mol^{-1}, to the score of the poses. For this reason, it is a common practice in the literature to regard this score as being the calculated free energy of binding (ΔG$_{binding}$) [30,31]. However, a more realistic numerical value for ΔG$_{binding}$ can be obtained by means of calculations that make use of much more sensitive and specific methods that analyse the trajectory of the molecular dynamics. Examples of these methods are the Linear Interaction Energy (LIE) [32] or the Molecular Mechanics Poisson–Boltzmann surface area (MM-PBSA) [28].

In order to improve our calculations of the ΔG$_{binding}$, thus, we performed molecular dynamics simulations with the complexes of porcine α-amylase bonded to the ligands listed in Fig. 2 during 20 nano seconds. This time was appropriate because all systems attained equilibrium at approximately 15 ns. The system was assumed to be in equilibrium when the rmsd of the atoms in the main chain of the protein began to oscillate at a constant frequency taking into consideration the last five nanoseconds. (Fig. 3A-E).

The regions of low rmsd represented by the light regions in all-to-all plots (Fig. 3F-H), indicate that the ligands moved little during a defined time interval of the simulation. When the rmsd of the last frames on the *x*-axis is larger regarding the first frames of the *y*-axis, it means that the position of the ligand at the end of the simulation is different from the initial position. For the AMY ligand, which represents the enzymatic substrate, we can see after the frame ~ 300, that it moved around 7 Å of rmsd regarding its initial position (Fig. 3C and G) and remained in this position along the rest of the simulation ranging within a cluster of similar conformations. This is evident since the rmsd in the *x* and *y* axes are small after the frame ~300 (Fig. 3G). A plausible explanation for this behaviour is that the AMY ligand, which is composed

of four glucoside unities $(Gluc)_4$ is relatively small compared to the natural starch substrate $(Gluc)_n$, where n can range in the order of thousands, until it be hydrolysed into smaller portions $(Gluc)_2$ and released from the alpha-amylase. Since the protein binding site is relatively large, the smaller size of AMY allowed it to move inside this cavity. This movement makes the ligand to stabilize at a nearby location, but still within the binding site after 8 ns of simulation. This did not affect the calculation of $\Delta G_{binding}$ by the MM-PBSA method because AMY was still inside the cavity (Fig S2, supplementary material).

The ligands ACA, HTN and specially AMY moved significantly with respect to their initial positions (Fig. 3E, F and G). For CTN ligand, the distribution of the light colour for a longer time interval indicates it moved little regarding its initial position (Fig. 3H) [33]. In addition, the ligands evaluated here are relatively large, and part of them situates outside the protein binding site and is exposed to the solvent. Without the constraints imposed by the protein residues, these structure groups have a greater degree of movement freedom, what justifies the relatively higher values of rmsd even when the ligand is bonded to its site. In spite of this, no ligand completely left the binding site.

Although we have chosen the pose that best represents the cluster of lower energy to be used in the simulations, the docking programs identified other clusters containing poses of close scores. Because of the large number of rotatable bonds of the compounds, it is statistically possible that the ligands may bind to the protein with slightly different but energetically close poses. This could justify the results observed in the *in vitro* experiments, which revealed a mixed (competitive-noncompetitive) parabolic inhibition type [15], meaning that more than one inhibitor molecule (or the same inhibitor with different poses) can eventually bind simultaneously to a given binding site. This could justify the variation in rmsd of some ligands observed in the simulations, since these ligands would be alternating between these poses of close energy.

The radius of gyration calculated for the alpha carbons of the protein oscillated around 1.0 Å in all systems (Fig. 3I), a relatively low value for a protein with 496 residues and approximate 75×55 Å dimensions. The low oscillation of the radius of gyration also indicates that the protein maintained its fold design and that the presence of ligands did not let the protein to unfold. Remarkable in Fig. 3 is that all ligands had an rmsd smaller that the substrate amylose (AMY), suggesting that the putative inhibitors may be interacting more firmly with the enzyme.

Each ligand interacts with the amino acid residues in a distinct manner in consequence of the different intermolecular forces that are active in their vicinity. For a best evaluation of the ligand interactions with the porcine α-amylase the contact frequencies of each ligand with

the various amino acid residues within a distance of 4 Å was calculated (Table S1, Supplementary Material). This provides a much more precise vision of the residues involved in the anchoring of the ligands than that one provided by a sole docking frame. Table S1 shows that the residues Trp58, Trp59, Gln63, His101, Tyr151, Leu162, Leu165, Lys200, His201, Ile235, Leu237, Asp300, Gly306 and Ala307 are involved in the anchoring of all the four ligands that were evaluated. They present a contact frequency higher than 30% with all ligands (Fig. 4A-D), what makes them the most important ones in this process. A similar analysis carried out with the human salivar α-amylase [34] concluded that the three residues most strongly interacting with acarbose are Asp300 > Trp59 > Gln63, in this order. The latter observation would, thus, in general terms at least, be in agreement with the present results. Most of these contacts are of the dipole-dipole type (hydrogen bonds) or charge-dipole with the OH groups of the ligands. Due to the relatively large size of the catalytic site of the porcine α-amylase and its pronounced exposition to the solvent, water molecules also intermediate some of the contacts between the residues and the ligands. These contacts, however, are not listed in Table S1. Figure 4 illustrates how the ligands are bound to the active site of the enzyme.
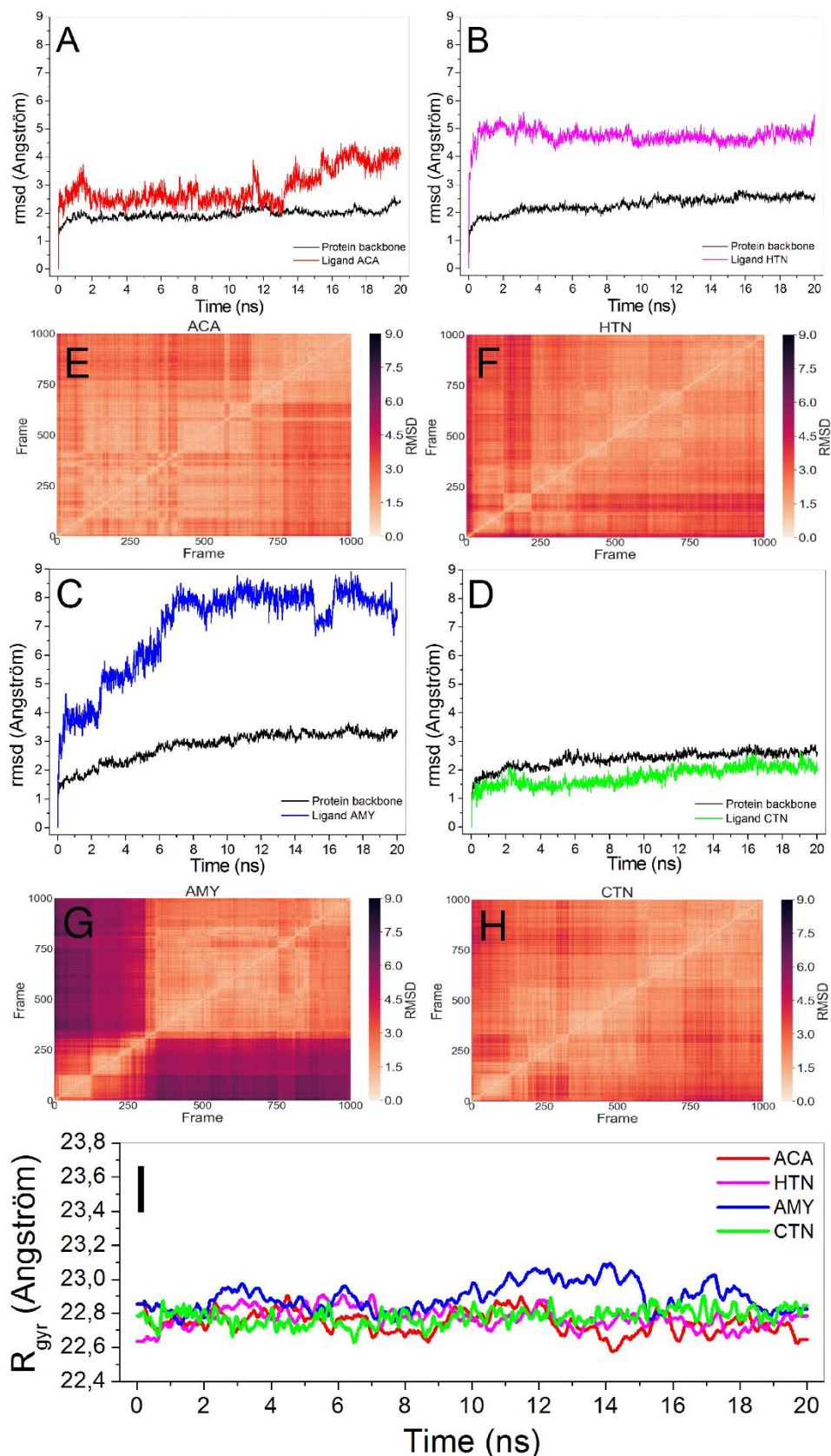
Fig. 3 Root mean square deviation (RMSD-based) convergence plots for all complexes evaluated along simulations. Color lines on white background plots **(A-D)** represent the rmsd for the corresponding ligand atoms. Below them are the all-to-all RMSD plot **(E-H)** for each ligand. Initially the frames were aligned by the superimposition of the protein backbone atoms on the reference structure. Then, the rmsd was calculated for the protein backbone (black line on white background plot) and for the ligands (colored lines on white background plot). **I)** The color lines represent the radius of gyration for the Cα atoms of porcine α-amylase bonded to each corresponding ligand identified by the color pattern.
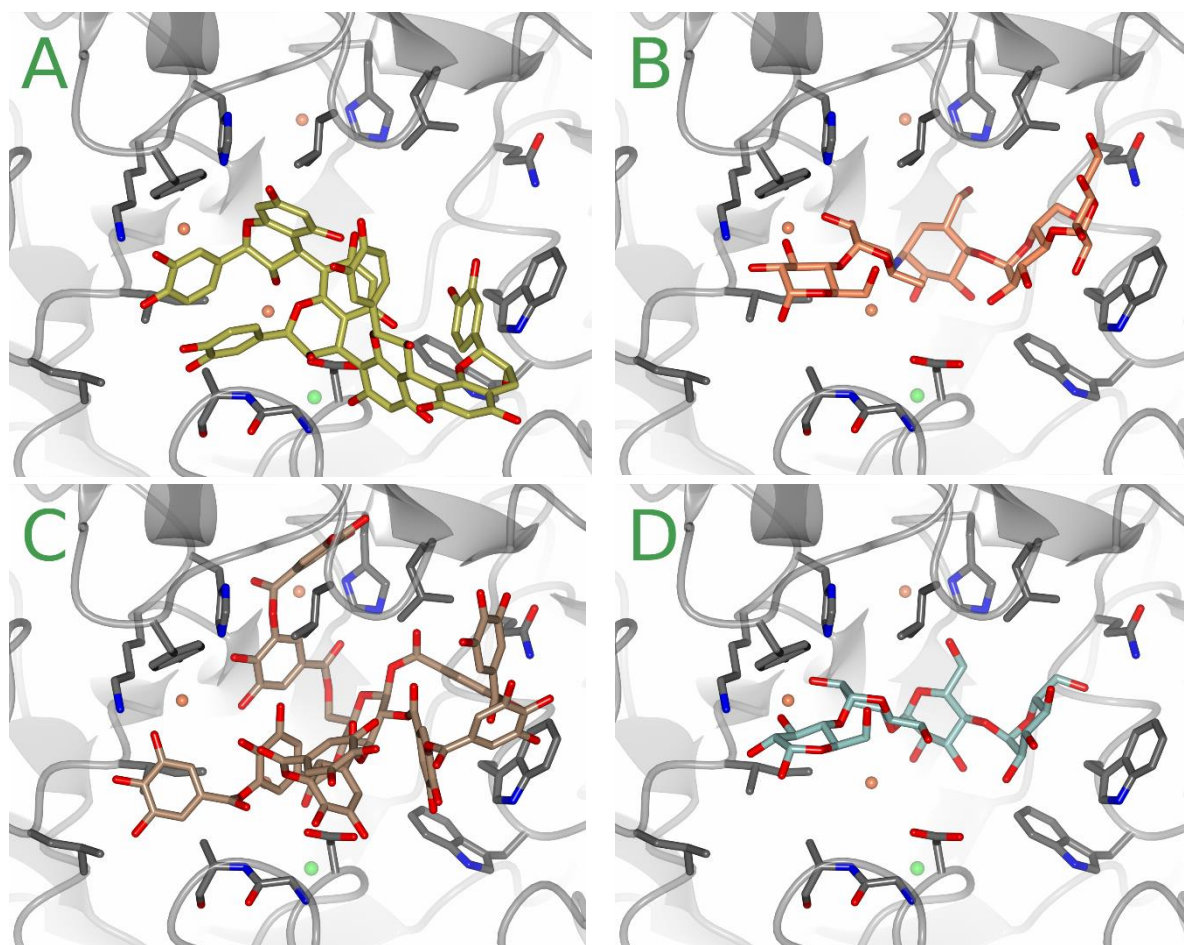
Fig. 4. Representation of the active site of the pancreatic α-amylase porcine showing the residues with higher frequency of contact with all the ligands that were evaluated, according to Table S1. The orange beads represent the water molecules that intermediate the contact of the substrate amylose tetrasaccharide (**D**) with the enzyme. The green sphere represents the cofactor Cl⁻ and the ligands in **A**, **B** and **C** are, respectively, the condensed tannin, acarbose pentasaccharide and hydrolysable tannin. The picture was generated using the CCP4mg program.

From the molecular dynamics trajectory, it was possible to calculate $\Delta G_{binding}$ values by means of the mm-pbsa method for each ligand, whose results are shown in Table 1. They show that the three methods used to identify the ligand with greater affinity with the enzyme diverged greatly. In addition, the AutoDock program was not able to evaluate the hydrolisable tannin ligand (HTN). In this way, only the Molegro program was used to dock this ligand. The mm-pbsa method, apparently the most sensible one, identified the hydrolysable tannin (HTN) as the ligand for which the pancreatic porcine α-amylase has the highest affinity, above the condensed tannin (CTN), when expressed in terms of $\Delta G_{binding}$. It looks appropriate at this point to compare these relative affinities with those computed from kinetic experiments [15]. Kinetic experiments allow to compute inhibition constants (K_i) which reflect or are, at least,

proportional to dissociation constants. For the formation of the enzyme-inhibitor complex (EI) at the active site Kato et al. [15] found K_i values of 37.7 and 147 μM for the hydrolysable and condensed tannins respectively. Kinetics reveals, thus that the pancreatic porcine amylase, binds more strongly the hydrolysable tannin than the condensed tannin at its active site. When the $\Delta G_{binding}$ (Table 1) values are converted into K_i by means of the relation $\Delta G = -RT ln K_i$, it is quite obvious that they differ by several orders of magnitude from the values obtained in kinetic studies. Among the reasons for this discrepancy is the fact that docking simulates a mechanism that corresponds to competitive inhibition with only one binding event whereas the kinetic assays actually revealed a mixed (competitive-noncompetitive) inhibition type. Additionally, inhibition was parabolic [15], meaning that more than one inhibitor molecule can eventually bind simultaneously to a given binding site. Besides, one has to consider other methodological factors in the computation of the $\Delta G_{binding}$ that are likely to influence the final outcome. The MM-PBSA model makes several approximations as for example the assumptions that the conformational phase space of the biomolecules is similar in vacuum and solvent and that there is no contribution of entropy. Several physical parameters are assumed, as for example the dielectric constant, the polar-solvation energy, the nonpolar-solvation energy the partial atomic charge and the force-field [28,35,36]. For all these reasons there is ample consent that values calculated by applying the mm-pbsa method cannot be compared to binding parameters computed from kinetic experiments but are perfectly comparable between themselves when the same simulation conditions are maintained. In the latter aspect, precisely, the docking simulations are not in disagreement with the kinetic measurements in that both predict that the hydrolysable tannin will be bound more strongly than the condensed tannin at the active site of the enzyme.

The results provided useful information for characterizing tannin binding to α-amylase. These and other similar future results can be applied in studies aiming at finding new hypoglycemic molecules among natural products. There are numerous studies in which new natural products, especially from plants, have been identified and characterized by their structural features [37-39]. Biological testing can be quite laborious, especially when it comes to *in vivo* experiments. The latter should be preceded by *in vitro* experiments in which enzyme activities linked to starch hydrolysis are assessed. Previous knowledge of the binding strengths of a given set of compounds to hydrolytic enzymes will allow to establish priorities for the experimental work, saving time, financial resources and, thus, rationalizing the approach to potentially useful new compounds. This applies to future studies as well as to the compounds

for which the binding features to the pancreatic α-amylase has been obtained in the present work.

5. Conclusion

The docking assays in addition to molecular dynamics simulation techniques of protein-ligand complexes provided a valuable contribution to the characterization of both hydrolysable and condensed tannins as inhibitors of porcine pancreatic α-amylase. The binding characteristics can be used as a way of predicting if a given compound, especially tannin derivatives, is or not able to inhibit the porcine pancreatic α-amylase and, in consequence, to offer perspectives for being used as a hypoglycemic drug. This work also suggests that the hydrolysable tannin would be a better inhibitor of the pig pancreatic α-amylase than the condensed tannin.

6. References

1. Boehlke C, Zierau O, Hannig C (2015) Salivary amylase - The enzyme of unspecialized euryphagous animals. Arch Oral Biol 60 (8):1162−1176. doi:10.1016/j.archoralbio.2015.05.008

2. Lynge Pedersen AM, Belstrom D (2019) The role of natural salivary defences in maintaining a healthy oral microbiota. J Dent 80 Suppl 1:S3−S12. doi:10.1016/j.jdent.2018.08.010

3. Saeedi M, Hadjiakhondi A, Nabavi SM, Manayi A (2017) Heterocyclic compounds: effective alpha-amylase and alpha-glucosidase inhibitors. Curr Top Med Chem 17 (4):428−440. doi:10.2174/1568026616666160824104655

4. IDF Diabetes Atlas (2017). 8th ed. Brussels, Belgium: International Diabetes Federation. http://www.diabetesatlas.org. Accessed 04/23/2019.

5. Forouhi NG, Wareham NJ (2019) Epidemiology of diabetes. Medicine, 41 (1): 22−27.

6. Teng H, Chen L (2017) alpha-Glucosidase and alpha-amylase inhibitors from seed oil: A review of liposoluble substance to treat diabetes. Crit Rev Food Sci Nutr 57 (16):3438−3448. doi:10.1080/10408398.2015.1129309

7. Taylor R (2012) Insulin resistance and type 2 diabetes. Diabetes 61 (4):778−779. doi:10.2337/db12-0073

8. Fischer S, Patzak A, Rietzsch H, Schwanebeck U, Kohler C, Wildbrett J, Fuecker K, Temelkova-Kurktschiev T, Hanefeld M (2003) Influence of treatment with acarbose or glibenclamide on insulin sensitivity in type 2 diabetic patients. Diabetes Obes Metab 5 (1):38−44. doi: 10.1046/j.1463-1326.2003.00239.x

9. Ueno H, Tsuchimochi W, Wang HW, Yamashita E, Tsubouchi C, Nagamine K, Sakoda H, Nakazato M (2015) Effects of miglitol, acarbose, and sitagliptin on plasma insulin and gut peptides in type 2 diabetes mellitus: A Crossover Study. Diabetes Ther 6 (2):187−196. doi:10.1007/s13300-015-0113-3

10. Hanefeld M, Cagatay M, Petrowitsch T, Neuser D, Petzinna D, Rupp M (2004) Acarbose reduces the risk for myocardial infarction in type 2 diabetic patients: meta-analysis of seven long-term studies. Eur Heart J 25 (1):10−16. doi:10.1016/S0195-668X(03)00468-8

11. Tundis R, Loizzo MR, Menichini F (2010) Natural products as alpha-amylase and alpha-glucosidase inhibitors and their hypoglycaemic potential in the treatment of diabetes: an update. Mini Rev Med Chem 10 (4):315−331. doi:10.2174/138955710791331007

12. Funke I, Melzig M (2005) Effect of different phenolic compounds on alpha-amylase activity: screening by microplate-reader based kinetic assay. Die Pharmazie 60 (10):796−797. ingentaconnect.com/contentone/govi/pharmaz/2005/00000060/00000010/art00017

13. Zhang B, Wang L, Luo L, King MW (2014) Natural dye extracted from Chinese gall–the application of color and antibacterial activity to wool fabric. J Clean Prod 80:204−210. doi:10.1016/j.jclepro.2014.05.100

14. Kusano R, Ogawa S, Matsuo Y, Tanaka T, Yazaki Y, Kouno I (2010) α-Amylase and lipase inhibitory activity and structural characterization of acacia bark proanthocyanidins. J Nat Prod 74 (2):119−128. doi:10.1021/np100372t

15. Kato CG, Goncalves GA, Peralta RA, Seixas FAV, de Sa-Nakanishi AB, Bracht L, Comar JF, Bracht A, Peralta RM (2017) Inhibition of alpha-amylases by condensed and hydrolysable tannins: focus on kinetics and hypoglycemic actions. Enzyme Res 2017:5724902. doi:10.1155/2017/5724902

16. Larson SB, Day JS, McPherson A (2010) X-ray crystallographic analyses of pig pancreatic alpha-amylase with limit dextrin, oligosaccharide, and alpha-cyclodextrin. Biochemistry 49 (14):3101−3115. doi:10.1021/bi902183w

17. Li C, Begum A, Numao S, Park KH, Withers SG, Brayer GD (2005) Acarbose rearrangement mechanism implied by the kinetic and structural analysis of human pancreatic alpha-amylase in complex with analogues and their elongated counterparts. Biochemistry 44 (9):3347−3357. doi:10.1021/bi048334e

18. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26 (16):1781−1802. doi:10.1002/jcc.20289

19. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25 (11):1400−1415. doi:10.1002/jcc.20065

20. Zoete V, Cuendet MA, Grosdidier A, Michielin O (2011) SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 32 (11):2359−2368. doi:10.1002/jcc.21816

21. Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 49 (11):3315−3321. doi:10.1021/jm051197e

22. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30 (16):2785−2791. doi:10.1002/jcc.21256

23. Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. In: Chemical Biology. Humana Press, New York, NY, p.243−250. doi:10.1007/978-1-4939-2269-7_19

24. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4 (3):435−447. doi:10.1021/Ct700301q

25. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem A 103 (19):3596−3607. doi:10.1021/Jp984217f

26. Schuttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D 60:1355−1363. doi: 10.1107/S0907444904011679

27. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. P Natl Acad Sci USA 98 (18):10037−10041. doi:10.1073/pnas.181342398

28. Kumari R, Kumar R, Lynn A, Consort OSDD (2014) g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54 (7):1951−1962. doi: 10.1021/Ci500020m

29. da Silva SM, Koehnlein EA, Bracht A, Castoldi R, de Morais GR, Baesso ML, Peralta RA, de Souza CGM, de Sá-Nakanishi AB, Peralta RM (2014) Inhibition of salivary and pancreatic α-amylases by a pinhão coat (*Araucaria angustifolia*) extract rich in condensed tannin. Food Res Int 56:1−8. doi:10.1016/j.foodres.2013.12.004

30. Ambrose GO, Afees OJ, Nwamaka NC, Simon N, Oluwaseun AA, Soyinka T, Oluwaseun AS, Bankole S (2018) Selection of luteolin as a potential antagonist from molecular docking analysis of EGFR mutant. Bioinformation 14 (5):241−247. doi:10.6026/97320630014241

31. Shakeel E, Akhtar S, Khan MKA, Lohani M, Arif JM, Siddiqui MH (2017) Molecular docking analysis of aplysin analogs targeting survivin protein. Bioinformation 13 (9):293−300. doi:10.6026/97320630013293

32. Gutierrez-de-Teran H, Aqvist J (2012) Linear interaction energy: method and applications in drug design. Methods Mol Biol 819:305−323. doi:10.1007/978-1-61779-465-0_20

33 . Grossfield A, Patrone PN, Roe DR, Schultz AJ, Siderius DW, Zuckerman DM (2018) Best practices for quantification of uncertainty and sampling quality in molecular

simulations [Article v1. 0]. Living J Comput Mol Sci 1 (1):5067. doi: 10.33011/livecoms.1.1.5067

34. Kato-Schwartz CG, Bracht F, Gonçalves GA, Soares AA, Vieira TF, Brugnari T, Bracht A, Peralta RM (2018) Inhibition of α-amylases by pentagalloil glucose: kinetics, molecular dynamics and consequences for starch absorption. J Funct Foods 44: 265–273. doi:10.1016/j.jff.2018.03.025

35. g_mmpbsa Forum. (2014). https://groups.google.com/forum/#!topic/g_mmpbsa/YR2fuPMifjM. Accessed 07/10/2018.

36. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 10 (5):449–461. doi:10.1517/17460441.2015.1032936

37. Souza AH, Correa RC, Barros L, Calhelha RC, Santos-Buelga C, Peralta RM, Bracht Adelar, Matsushita M, Ferreira I C (2015) Phytochemicals and bioactive properties of Ilex paraguariensis: An *in-vitro* comparative study between the whole plant, leaves and stems. Food Res Int 78:286286-294. doi:10.1016/j.foodres.2015.09.032

38. Gonçalves GA, de Sá-Nakanishi AB, Comar JF, Bracht L, Dias MI, Barros L, Peralta RM, Ferreira ICFR, Bracht A (2018) Water soluble compounds of rosmarinus officinalis l. improve the oxidative and inflammatory states of rats with adjuvant-induced arthritis. Food Funct 9(4):2328-2340. doi:10.1039/c7fo01928a.

39. Correa RC, Peralta RM, Haminiuk CWI, Maciel GM, Bracht A, Ferreira ICF (2018) New phytochemicals as potential human anti-aging compounds: Reality, promise, and challenges. Crit Rev Food Sci Nutr 58(6):942-957. doi:10.1080/10408398.2016.1233860

Supplementary Materials

In silico evaluation of condensed and hydrolysable tannins as inhibitors of pancreatic α-amylase

Paulo Sérgio Alves Bueno[1], Camila Gabriel Kato Schwartz[1,2], Diego de Souza Lima[3], Adelar Bracht[1,2], Rosane Marina Peralta[1,2], Flavio Augusto Vicente Seixas[3].

[1]Department of Biochemistry. Universidade Estadual de Maringá, Maringá, PR, Brazil.

[2]Post-graduate program of Food Science. Universidade Estadual de Maringá, PR, Brazil

[3]Department of Technology, Universidade Estadual de Maringá, Umuarama, PR, Brazil

* Corresponding author

Universidade Estadual de Maringá, UEM

Department of Technology

Av. Angelo Moreira da Fonseca, 1800, 87506-370, Umuarama, PR, Brazil.

Phone: +55 44 3621 9300

e-mail: favseixas@uem.br
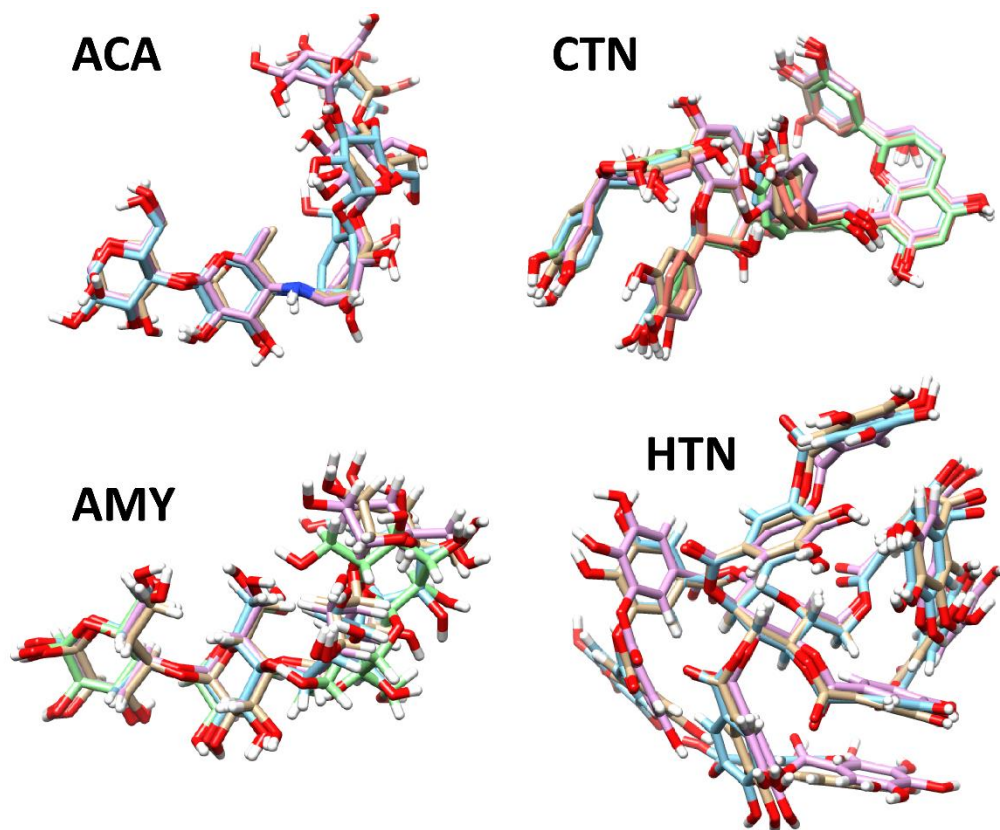
Fig. S1. Representation of the pose clusters of the evaluated ligands. The pose with best score from each cluster was chosen to be used in molecular dynamics simulations. The regions of the ligand with poor overlap is located outside the protein binding site and exposed to solvent. With fewer contacts to restrict their movements, these regions of the ligands become more flexible throughout the simulation, which may increase the rmsd value.

Table S1. Contact frequency of the pig pancreatic α-amylase residues with the evaluated ligands as computed from equilibrium molecular dynamics. Bold numbers refer to the residues with the highest contact frequency with all ligands, being thus more important for their anchoring. HTN (hydrolysable tannin), ACA (acarbose pentasaccharide), CTN (condensed tannin), AMY (amylase tetrasaccharide).

Residue	HTN	ACA	CTN	AMY	Residue	HTN	ACA	CTN	AMY
VAL-51	0.01	0.48	-	-	LEU-166	0.29	-	-	0.02
ASN-53	0.07	0.08	-	-	ASP-167	0.98	-	-	0.01
PRO-54	0.86	0.11	0.15	-	ARG-195	-	0.87	-	0.26
PRO-57	0.01	0.11	0.01	-	ASP-197	0.20	1.00	-	0.28
TRP-58	**0.95**	**0.94**	**0.94**	**0.89**	ALA-198	0.94	0.99	0.02	0.87
TRP-59	**1.00**	**1.00**	**1.00**	**0.99**	SER-199	0.12	0.01	-	-
GLU-60	-	0.10	-	-	**LYS-200**	**0.88**	**0.98**	**0.79**	**0.98**
TYR-62	0.53	0.90	0.01	0.99	**HIS-201**	**1.00**	**0.44**	**0.96**	**0.99**
GLN-63	**0.92**	**0.99**	**0.87**	**0.69**	GLU-233	0.63	1.00	-	0.42
VAL-98	-	-	-	0.83	VAL-234	0.18	-	-	-
ASN-100	0.97	-	-	-	**ILE-235**	**0.80**	**0.93**	**1.00**	**0.96**
HIS-101	**0.99**	**0.85**	**0.68**	**0.99**	ASP-236	0.98	0.46	0.75	-
CYS-103	-	-	0.08	-	**LEU-237**	**0.98**	**0.53**	**0.92**	**0.72**
GLY-104	-	0.49	0.29	-	GLY-238	0.28	-	0.17	0.70
SER-105	0.95	0.73	0.20	-	GLY-239	-	0.04	-	0.12
GLY-106	0.70	0.16	-	-	GLU-240	0.01	0.83	0.05	0.01
ALA-107	0.01	0.04	-	-	ALA-241	-	0.24	-	-
SER-145	0.02	-	-	-	ILE-242	-	0.44	0.57	-
GLY-146	0.07	-	-	-	GLU-255	0.02	-	-	-
GLY-147	0.01	-	-	0.04	PHE-256	0.95	-	-	-
ILE-148	0.02	-	0.03	0.83	LYS-257	0.33	-	-	-
GLU-149	-	0.01	0.08	-	ALA-260	0.03	-	-	-
SER-150	-	-	-	0.03	HIS-299	0.65	0.90	0.07	0.78
TYR-151	**0.17**	**0.99**	**1.00**	**0.79**	**ASP-300**	**1.00**	**0.93**	**0.43**	**0.95**
ASN-152	-	-	0.53	-	ASN-301	-	0.03	-	-
ASP-153	-	-	0.26	-	GLY-304	0.90	-	-	-
PRO-154	-	-	0.48	-	HIS-305	1.00	-	0.97	0.17
GLN-156	0.01	-	-	0.25	**GLY-306**	**0.92**	**0.30**	**0.52**	**0.82**
VAL-157	0.02	0.25	0.63	0.83	**ALA-307**	**1.00**	**0.37**	**0.36**	**0.65**
ARG-158	0.01	-	0.15	0.01	GLY-308	0.95	-	-	-
ASP-159	0.01	-	-	-	GLY-309	0.02	-	-	-
CYS-160	0.84	-	-	0.57	SER-311	0.01	-	-	-
GLN-161	0.27	-	-	0.25	ASN-350	0.01	-	-	-
LEU-162	**1.00**	**0.94**	**0.86**	**0.98**	GLN-352	0.06	-	-	-
VAL-163	0.82	0.99	0.45	0.09	VAL-354	0.10	-	-	-
GLY-164	0.10	0.33	0.24	0.01	ASP-356	0.99	0.01	0.85	-
LEU-165	**1.00**	**0.99**	**0.99**	**0.97**	TRP-357	0.95	-	0.87	0.03

Artigo 3: New inhibitors of chorismate synthase present antifungal activity against *Paracoccidioides brasiliensis.*

Future
MICROBIOLOGY

New inhibitors of chorismate synthase present antifungal activity against *Paracoccidioides brasiliensis*

Paulo SA Bueno[1], Franciele AV Rodrigues-Vendramini[2], Marina Toplak[3], Peter Macheroux[3], Érika S Kioshima[2] & Flavio AV Seixas*,[1] (ID)

[1]Department of Technology, Universidade Estadual de Maringá, Av. Ângelo Moreira da Fonseca, 1800, 87506–370 Umuarama, PR, Brazil

[2]Department of Clinical Analysis & Biomedicine, Universidade Estadual de Maringá, Av. Colombo 5790, 87020–900 Maringá, PR, Brazil

[3]Institute of Biochemistry, Graz University of Technology, Petersgasse 12/2, 8010, Graz, Austria

*Author for correspondence: Tel.: + 55 443 621 9337; Fax: +55 443 621 9326; favseixas@uem.br

Aim: A structural model of chorismate synthase (CS) from the pathogenic fungus *Candida albicans* was used for virtual screening simulations. **Methods:** Docking, molecular dynamics, cell growth inhibition and protein binding assays were used for search and validation. **Results:** Two molecules termed CS8 and CaCS02 were identified. Further studies of the minimal inhibitory concentration demonstrated fungicidal activity against *Paracoccidioides brasiliensis* with a minimal inhibitory concentration and minimal fungicidal concentration of 512 and 32 μg·ml^{-1} for CS8 and CaCS02, respectively. In addition, CaCS02 showed a strong synergistic effect in combination with amphotericin B without cytotoxic effects. *In vitro* studies using recombinant CS from *P. brasiliensis* showed IC$_{50}$ of 29 μM for CaCS02 supporting our interpretation that inhibition of CS causes the observed fungicidal activity.

Invasive fungal infections pose serious threats to human health, resulting in at least 1.5 million deaths worldwide each year [1]. Almost half of this number is caused by the *Candida* species [2]. It lives commensally in the intestine, urogenital tract and other mucosal surfaces with approximately 50% of the population being affected [3]. The immunosuppressed patients such as transplant recipients, patients with cancer and people affected by HIV are the most affected by these nosocomial infections [4,5].

The paracoccidioidomycosis (PCM) is another invasive fungal disease, endemic in Latin American countries, which presents a high prevalence in South America and stands out as the major cause of death among systemic mycoses in Brazil [6,7]. The genus *Paracoccidioides* includes the thermodymorphic fungi *Paracoccidioides brasiliensis* and *Paracoccidioides lutzii*, which are the species most commonly identified as agents of the PCM [8]. In Brazil, it is estimated that approximately 51.2% of the deaths by systemic mycosis are caused by PCM [9]. Among the risk factors for acquiring this infection are the external activities related to the management of soil contaminated with the fungus and geographical exposure in affected regions. In addition, there are other associated factors, such as sex, age, smoking, alcohol consumption and socioeconomic conditions [10].

PCM presents distinct symptoms, and may manifest in the acute/subacute and chronic form. The chronic form of PCM affects 90% of the patients. In these cases, a pulmonary architectural distortion is observed in 63.8 to 100% of the patients [11–13]. The pulmonary damage caused by fungal stimulation remains the same after the treatment and causes pulmonary fibrosis with the loss of respiratory function in 50% of the cases [14]. The arsenal of drugs against systemic mycoses is limited to antifungal agents such as polyenes (amphotericin B [AmB]), azoles (itraconazole and fluconazole) and echinocadins (caspofungin, micafungin). However, there are reports in the literature describing strains resistant to each of these drugs [15], which further limits their use [16]. Therefore, there

Future
Medicine

is an urgent need for the discovery of new antifungal agents, with preference for those that can act on different metabolic targets.

Nowadays, the development of an antifungal agent is a challenge, due to the few potential metabolic targets that are not shared by the fungus and the human species (host). Enzymes of the shikimate pathway are promising targets for the development of new antifungal agents [17], besides antimicrobial agents [18], antiparasites [19] and herbicides [20], because this pathway is absent in mammals [17]. It converts two compounds present in carbohydrate metabolism, phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P), into chorismate in a seven-step enzyme-catalyzed pathway. Metabolites such as the aromatic amino acids, folic acid, vitamin K and ubiquinone are produced from chorismate, which are essential for the survival of fungi, bacteria, algae, plants and parasites of the apicomplex phylum [21]. Thus, the enzymes that make up this pathway are excellent targets for drug discovery, despite still being unexplored in *Candida* spp and *Paracoccidioides* spp.

The enzyme chorismate synthase (CS) (EC 4.2.3.5) acts in the last step of the shikimate pathway, catalyzing the conversion of the substrate 5-enolpyruvylshikimate-3-phosphate (EPSP) to chorismate, in the presence of the reduced form of flavin mononucleotide ($FMNH_2$) [21,22]. Although the 3D structure of *C. albicans* CS (*Ca*CS) has not yet been solved by any experimental method, its amino acid sequence has already a confirmed status in the Uniprot database. In addition, there are other crystallographic structures of CS deposited in the Protein Data Bank (PDB) as for example from *Saccharomyces cerevisiae* [23], which shares 75% identity with the homologous enzyme from *C. albicans*.

In this context, the objective of this work was to use bioinformatics tools, such as homology modeling and large-scale virtual screening simulations to discover new molecules with potential antifungal activity. In order to validate the results obtained *in silico*, a set of *in vitro* assays of minimal inhibitory concentration (MIC) and minimal fungicidal concentration (MCF) in human pathogenic fungi were performed, as well as kinetic and binding studies using CS from *Paracoccidioides brasiliensis* in the presence of the selected ligands.

Material & methods

Sequence identification & modeling of the 3D structure

The amino acid sequence of *Ca*CS strain WO-1 was obtained from the Uniprot database [24] (uniprot id: C4YMI5), and was used in the search for structural templates by means of BLASTp [25] in the PDB [26]. The model and template sequences were aligned, and the result was used as the input parameter for the Modeller-9v14 program [27]. The structure was modeled in its biological unit (homotetramer) where each subunit was bound to the EPSP substrate and the $FMNH_2$ cofactor. A symmetry constraint was imposed between the Cα of the four chains, and then, 2000 models were generated. The top ten were selected based on Modeller DOPE score. From these, the best model was selected based on the stereochemical quality evaluated by the program Procheck [28]. The residues in forbidden regions of the Ramachandran plot were corrected by means of point modeling.

Energy minimization & molecular dynamics simulations

Energy minimization was performed using the software package NAMD2/VMD [29,30]. The structure of the homotetramer obtained from modeling was solvated with TIP3P water in a periodic box with limits of 10 Å away from the outermost surface of the complex. Na^+ counterions were added in sufficient amount to neutralize the system charges. The Charmm force field (FF) c35b2-c36a2 [31] was used for the protein atoms, while the FFs for the substrate, $FMNH_2$ cofactor and molecules obtained by the virtual screening were generated in the same format by the Swissparam server [32]. To increase the accuracy of the ligand FFs, the Mulliken's charges were calculated at the B3LYP/6-311 G level of theory using the Orca program [33], at the protonation state at pH 7.0.

The molecular dynamics (MD) process occurred in five steps. In the first step, the atoms of the ligand and cofactor were fixed in the space, while those of the protein and other atoms of the system were free to move. At this point, the system was subjected to 20,000 minimization steps by conjugate gradient (CG). In the second step, all the atoms of the system were again minimized by other 10,000 steps of CG. The structure resulting from this step was used in the redocking and docking.

In the third step, the atoms of protein and ligands were kept fixed, while the waters and salts were subjected to 60 ps of equilibrium MD. In the fourth step, all the atoms were subjected to a new CG cycle for 20,000 steps. In the fifth and the last step, the entire system was subjected to 50 ns of equilibrium MD under NPT conditions (constant: number of atoms, pressure 1 atm and temperature 300 K).

The analyses were performed in terms of root mean square deviation (RMSD) from the reference structure (second step of the MD), as well as the radius of gyration (R_{gyr}) of the protein–ligand complexes, the residues in contact with the inhibitors at a distance of up to 4.0 Å, as well as the root mean square fluctuation (RMSF) of the α carbons for each residue were extracted from the trajectory file.

The long time MD simulations required a higher processing power, being performed on 20 nodes of an Intel Xeon E5-2695v2 Ivy Bridge processors, 2.4 GHZ (480 cores) from the SDumont supercomputer at National Laboratory of Scientific Computation (LNCC), Campinas, SP, Brazil.

Docking & virtual screening simulations

The docking protocol was established by redocking simulations of the EPSP substrate in the protein–cofactor complex using the AutoDock-4.2.3 program [34] implemented in the Pyrx-0.9 graphical interface [35]. The parameters used were the standard search algorithm (Lamarckian GA), grid of 0.375 Å, search center on ligand EPSP and box size of 50, 50 and 50 on *x*, *y* and *z*. The torsion angles of the enzyme and cofactor were restricted, but flexibility of the ligands was allowed. All water molecules and ions were removed from the structure. The protocol was considered validated when the EPSP pose overlapped the pose of the modeled substrate with RMSD less than 0.8 Å.

For the large-scale screening simulations, three libraries were constructed. The first was assembled by searching for purchasable molecules, similar to the known inhibitors of CS, using 70% Tanimoto index at the databases PubChem [36], Zinc [37], DrugBank [38] and Brenda enzymes [39]. The second library consisted of the natural products from Zinc database. The third library was the catalog of SIGMA (Sigma-Aldrich, MO, USA). The EPSP structure was included in the libraries, so that, at the end of the simulation, only the best ranked molecules relative to the substrate EPSP were considered. The selected molecules were downloaded in *.sdf format. Then, the molecules selected in the first screening were subjected to three new screenings (repetitions) to select the ligands with reproducibility, in order to eliminate false-positive results.

Microorganisms

The *in vitro* susceptibility assays were performed on the standard strains *C. albicans* ATCC 90028 (American Type Culture Collection) and *P. brasiliensis* (Pb18), both strains belonging to the Library of the State University of Maringa. Prior to testing, the *C. albicans* strain was subcultured on agar sabouraud dextrose (SDA, Difco™, MI, USA) and incubated at 35°C for 24 h. The Pb18 isolate was subcultured on Fava–Netto agar medium [40] at 36°C for 5 days.

Ligand preparation

The compound CS8 (AG-690/37099007) was purchased from Molport (Riga, Latvia) and the compound CaCS02, as well as the control drug, AmB were purchased from Sigma-Aldrich. Initially a stock solution of the molecules to be tested was prepared in dimethylsulfoxide (DMSO) at a concentration of 50 mg·ml^{-1} and stored at -80°C until further use. For the assays, the molecules were diluted in RPMI [RPMI-1640 without bicarbonate, L-glutamine, supplemented with 2% glucose] and buffered (pH 7.0) with 0.165 M morpholinopropanesulfonic acid (MOPS) (Sigma-Aldrich) 0.2% plus Pluronic F-127 (Sigma-Aldrich). The fungi were incubated in the absence and presence of the diluent to confirm that the exposure did not influence the growth of the assayed isolates.

Antifungal activity assays

In order to evaluate the susceptibility profile of the pathogenic fungi to the compounds selected by virtual screening, a test based on the broth microdilution method proposed by CLSI (Clinical and Laboratory Standards Institute, 2008) was performed according to protocol M27-A3 for the determination of MIC and minimum fungicidal concentration (MFC) for *C. albicans*. The method was adapted to the conditions of *P. brasiliensis*, as previous described [41]. The tests were performed on sterile flat bottom 96-well microtiter plates (Techno Plastic Products, Trasadingen, Switzerland) and the final tested concentrations of the evaluated molecules ranged from 1 to 512 μg·ml^{-1}. The inoculums were suspended in sterile saline buffer (0.85%), to a final concentration of $1–2 \times 10^3$ cells·ml^{-1} for *C. albicans* and $1–2 \times 10^4$ cells·ml^{-1} for *P. brasiliensis*. The determination of the end points of the MIC and MFC readings were performed visually and correspond to the concentration of each antifungal capable of inhibiting the 100% growth of the isolates. The assays were performed in duplicate.

Checkerboard assay

CaCS02 was selected for further trials due to its better antifungal potential. In order to evaluate the synergy between AmB and CaCS02, microdilution checkerboard was performed. Briefly, concentrations below or above the MIC of the drugs against *P. brasiliensis* were combined in a 96-well plate, starting at 128 μg·ml^{-1} and vertically distributed, while 32 μg·ml^{-1} AmB were added horizontally [42]. After 6 days of incubation at 35°C, the fungal growth was visualized and 20 μl of Resazurin (Sigma-Aldrich) 0.01% in sterile water were added to each well and incubated for a further 24 h. A total of 7 days later, the inhibition was confirmed by the abrupt change in color when there was fungal growth. Fractional inhibitory concentration (FIC) was determined by $\Sigma FIC = FICA + FICB = (Comb_{AmB}/MIC_{AmB}) + (Comb_{CaCS02}/MIC_{CaCS02})$, revealing a strongly synergistic effect, FIC <0.5; synergistic effect, FIC < 1; additive effect, FIC = 1; without effect, 1 < FIC < 2 and antagonistic effect, FIC > 2 [43].

Recombinant production & binding of CaCS02 to CS from *Paracoccidioides brasiliensis*

Recombinant production of CS from *Paracoccidioides brasiliensis* (*Pb*CS) was performed as described previously [41]. For the binding studies, CaCS02 was dissolved in 50 mM MOPS, pH 7.5, diluted to a final concentration of approximately 10 μM (A_{590}, ~0.2; 800 μl) and transferred to a quartz cuvette (with 800 μl of buffer in the reference cuvette). After recording an initial UV-*visible* absorption spectrum between 300 and 800 nm, 10 μl aliquots of *Pb*CS (~393 μM) were added and spectral changes were monitored after each addition. In order to quantify the binding affinity of CaCS02 to *Pb*CS, the relative absorption changes at 530 nm were plotted as a function of the protein concentration in the cuvette, with the standard deviations displayed as error bars (from three titrations).

Inhibitory effect of CaCS02 on recombinant PbCS

To test the inhibitory effects of CaCS02 on *Pb*CS, a coupled assay involving EPSP, CS and anthranilate synthase was used [44,45]. This special assay setup allowed forward coupling of the CS reaction, leading to the formation of anthranilate, a stable product that exhibits a fluorescence emission maximum at 390 nm, and thus, can be detected spectrofluorometrically. Reaction mixtures containing 3 mM $MgSO_4$, 7.5 mM L-Gln, 22.5 mM $(NH_4)_2SO_4$, 1 mM dithiothreitol (DTT), 10 μM FMN, 80 μM EPSP, 4 μM *Pb*CS, ~20 μM anthranilate synthase and various concentrations of CaCS02 (0.1–500 μM) were prepared in 50 mM MOPS (pH 7.5) and pipetted into a 96-well plate (triplicate determinations). After the addition of 500 μM Nicotinamide adenine dinucleotide phosphate (NADPH), the plate was incubated at 37°C for 60 s before recording the fluorescence emission changes at 390 nm (λ_{ex} = 340 nm) using a Spectramax Gemini XS Microplate Spectrofluorometer (Molecular Devices, CA, USA) for 5 min (data points every 12 s). Initial velocities were extracted from the first 15 data points, and relative rates were plotted as a function of the log (CaCS02 micromolar concentration), with the standard deviations shown as error bars (triplicate determinations). Using the program GraphPad Prism 5 (GraphPad Software, Inc., CA, USA), the data were fitted applying the nonlinear curve fit 'log (inhibitor) versus normalized response − variable slope', which allowed for the determination of the IC_{50}.

Cytotoxicity assay

The CaCS02 cytotoxicity evaluation was performed in HeLa and Vero cell lineages, using the CellTiter 96 assay (Promega, WI, USA) based on the reduction of 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS). The cells were incubated in 96-well culture plates maintained in Dulbecco's Modified Eagle's Medium (DMEM) with fetal bovine serum (5%) and treated with different concentrations of CaCS02 (0.0 to 512 μg·ml^{-1}). After 24 h, the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) solution was added to each well culture and incubated in a CO_2 oven (5%) at 37°C for 180 min. Cell viability was determined by optical density in a spectrophotometer at 490 nm. The cytotoxicity was presented as the average of three independent experiments. The percentages of cell viability were calculated relative to controls-containing medium only [46].

Results

The amino acid sequence of *Ca*CS comprises 378 residues giving rise to a molecular mass of 40,895 Da and a pI of 6.14 [47]. In fungi, the first and seventh enzyme, in other words, phospho-2-dehydro-3-deoxyheptonate aldolase tyrosine-inhibited and CS encoded by ARO4 and ARO2, respectively, are monofunctional; whereas, the
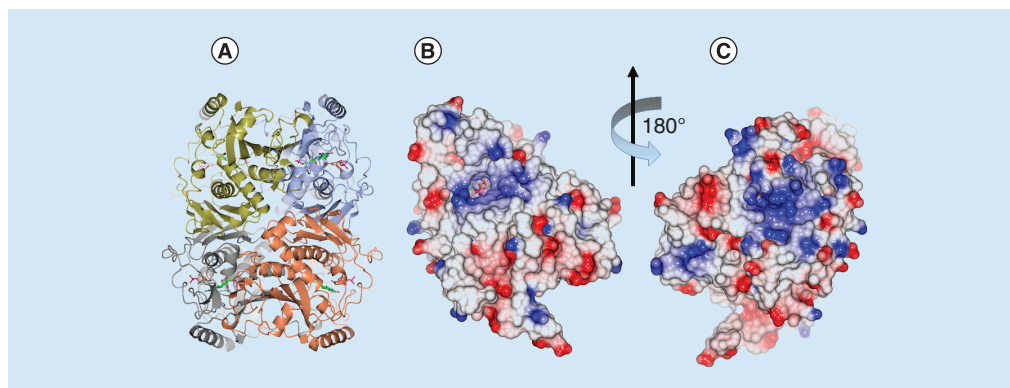
Figure 1. **Representations of the enzyme chorismate synthase from _Candida albicans_. (A)** Ribbon model of the quaternary structure colored by chain. The molecules bonded are 5-Enolpyruvylshikimate-3-phosphate (EPSP) (brown) and Flavin mononucleotide (FMNH$_2$) (green). **(B)** Representation of the surface of chain A colored by electrostatic potential seen from the neighboring subunit. The molecules in the cavity are EPSP and FMNH$_2$. **(C)** The same chain after 180° rotation seen from outside. Since the protein was modeled in the 'closed conformation' EPSP and FMNH$_2$ are not visible in this view.

remaining enzymatic activities are found on a single pentafunctional polypeptide (Supplementary Figure 5A & B, Supplementary Data), which forms the so-called 'arom complex' encoded by ARO1 [24].

Further analysis of the sequence of _Ca_CS by Blastp provided two structural templates. One of them, the CS from _Saccharomyces cerevisiae_ (PDB ID: 1r53) [23] shares 75% identity with _Pb_CS, and was used as a template for the polypeptide chain. Because this structure does not provide the spatial coordinates of the substrate EPSP and the cofactor, we modeled these two ligands using the structure of CS from _Streptococcus pneumoniae_ (PDB ID: 1qxo) [48], which shares 34% identity. The template 1r53 represents the biological unit of the homotetrameric enzyme; however, many residues around the active site are ill-defined due to the lack of electron density in this region. Therefore, the B chain of the 1qxo template was used to model these missing regions because they were well defined in this chain. CS exhibits two distinct conformations of the active site: in the unliganded apo-form several loops near the active site adopt an open conformation; whereas, in the presence of the substrate and FMN, these loops are in a closed conformation basically forming a lid blocking the access to the active site. The use of the 1qxo template allowed the _Ca_CS to be modeled in closed form and in the presence of the FMN cofactor plus the substrate, which presumably represents a catalytically active conformation.

The high degree of freedom of movement of the protein in the monomeric form could compromise the configuration of the active site, essentially structured by random segments (loops and turns). Previous MD simulations with the CS of _Mycobacterium tuberculosis_ [49], have demonstrated that the homotetrameric form is more stable than the monomeric and dimeric forms. Therefore, we have modeled _Ca_CS as a homotetramer, the most likely biological unit. The stereochemical constraints of this conformation diminished possible distortions in the loops contributing to the active site, and thus, renders this region more amenable to MD simulations. After modeling and minimization, only one of the subunits was effectively used in the docking studies.

The final model of _Ca_CS (Figure 1) showed 99.7% of the residues in allowed regions of the Ramachandran plot (Supplementary Figure 6, Supplementary Data) indicating excellent stereochemical quality.

Finally, we have performed a two-step minimization by using the FF of the FMN cofactor in the reduced form to account for the structural changes that occur upon reduction of the oxidized cofactor as present in the template 1qxo. At the end of the minimization, the isoalloxazine ring showed bending along the N(5)–N(10) axis indicating that the reduced FMN adopts the so-called 'butterfly conformation' [50].

Validation of docking protocol & large-scale virtual screening

The feasibility of using the _Ca_CS in large-scale screening studies as well as the validation of the docking protocol was performed by the redocking method using the Autodock program, which provided an RMSD of 0.762 Å (average of four simulations). Once the results were reproducible (Supplementary Figure 1, Supplementary Data), the _Ca_CS structure and the docking protocol were validated and were applied in large-scale virtual screening.
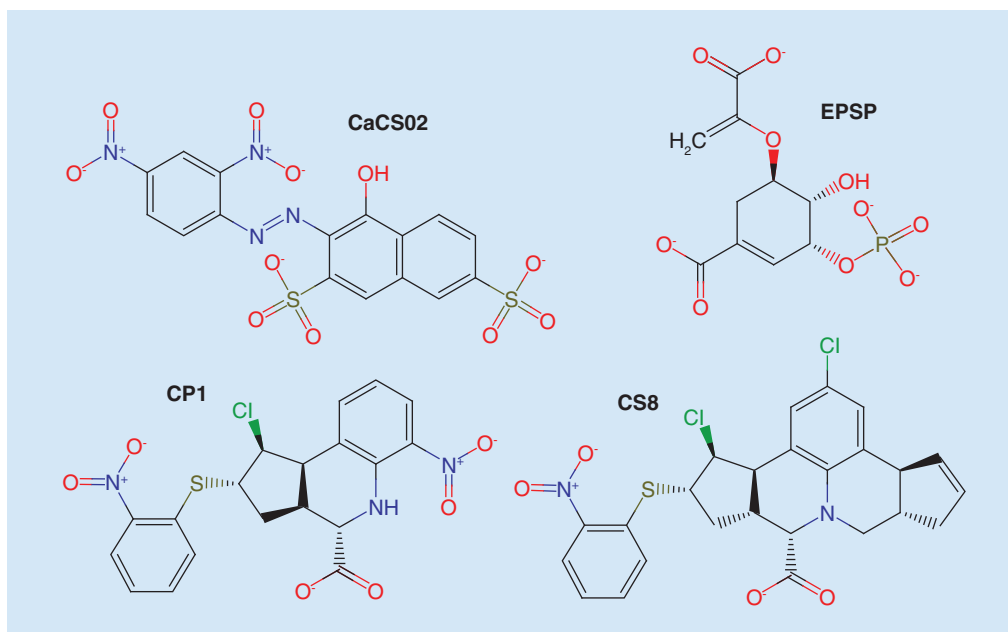
Figure 2. The chemical structure of the molecules selected by the virtual screening (CaCS02 and CS8), plus the structure of the 5-enolpyruvylshikimate-3-phosphate substrate and the CP1 molecule. CP1 was described as a Chorismate synthase from *Paracoccidioides brasiliensis* inhibitor.
EPSP: 5-Enolpyruvylshikimate-3-phosphate; *Pb*CS: Chorismate synthase from *Paracoccidioides brasiliensis*.

The library of compounds similar to the known inhibitors of CS contained 2501 molecules. The library of commercially available natural products from the Zinc database contained 120,435 molecules. The third library was the Sigma-Aldrich catalog containing 73,163 compounds (total 196,099 molecules). The candidates were subdivided into folders with 2500 molecules each, plus the EPSP structure. The final protocol validated in the redocking was then applied in the virtual screening of the libraries. In the first screening, 941 molecules were better ranked than EPSP (scores better than -11.0 ± 1.45). Of these molecules, 539 presented reproducibility in three new screens, which were used to eliminate false-positive results. In all replicates, the molecules Zinc6445805 (CS8) and Zinc4217277 (CaCS02) always appeared among the five best ranked, showing scores of -11.5 ± 0.48 and -11.2 ± 0.51, respectively. These two compounds were purchased for *in vitro* assays. Figure 2 shows the structures of CS8, CaCS02 and EPSP. In fact, CS8 is very similar to compound CP1, which was recently demonstrated to exhibit antifungal activity against *P. brasiliensis* and be a potent inhibitor of *Pb*CS [41].

MD simulations

In order to investigate the structural stability of *Ca*CS, as well as to obtain information about the residues involved in the stabilization of the ligands, MD simulations were performed for 50 ns. In these simulations, the final model with the cofactor $FMNH_2$ plus EPSP or the best pose of CaCS02 or CS8 were minimized and then equilibrated as described in 'Materials & methods'. The system reaches thermodynamic equilibrium when the RMSD calculated for the protein main chain atoms oscillates at a constant plateau for at least 10 ns. As shown in Supplementary Figure 2A (Supplementary Data), the structures converged to an equilibrium at 30 ns. For the *Ca*CS–EPSP complex, the RMSD was 3.0 Å ± 0.4. This value is close to those obtained in studies using CS from *Mycobacterium tuberculosis* [49] and *Shigella flexneri* [18], as well as for the parasite *Plasmodium falciparum* [19]. This result also showed that the enzyme in complex with either CaCS02 or CS8 is more stable as compared with the enzyme–substrate complex. This is probably due to the fact that these ligands are larger, and therefore, interact with a greater number of residues as compared with EPSP (Supplementary Tables 1 & 2, Supplementary Material).

The stability of the structures was also evaluated in terms of the radius of gyration (R_{gyr}), defined as the average quadratic distance of a set of atoms from their common center of gravity. As shown in Supplementary Figure 2B (Supplementary Material), the complexes stabilized after 10 ns and protein unfolding was not observed.

Table 1. Minimum inhibitory concentration effect of CS8 and synergic effect of *CaCS02* and amphotericin B for *Paracoccidioides brasiliensis*.

Compound	MIC/MFC (μg·ml^{-1})			FIC model	
	Alone	Combination	FIC	$\sum$ FIC	Interpretation
CaCS02	32	4	0.125	0.375	Strong synergistic
AmB	0.5	0.125	0.25		
CS8	512	–	–	–	–

AmB: Amphotericin B; FIC: Fractional inhibitory concentration; MIC: Minimum inhibitory concentration; MFC: Minimum fungicidal concentration.

From the MD simulations, it was possible to evaluate regions of the protein with greater flexibility by means of the RMSF of the Cα of each residue in the presence of the ligands. The RMSF was calculated in the last 5 ns of the simulation, when the protein was in its equilibrium state (Supplementary Figure 3). The residues with greater flexibility were those of the C-terminal regions in all chains, which is common in most proteins, and in the case of *Ca*CS, they refer to a segment without a defined secondary structure. Furthermore, we observed three regions with greater flexibility in the presence of any of the three ligands. These regions 1, 2 and 3 were also verified in a previous work [49], indicating that this pattern is a general property of CS in complex with a ligand. However, these regions comprise long loops, which by nature, have a high degree of freedom of movement.

From the trajectory of MD simulations, it was also possible to evaluate the frequency of contacts between amino acid residues of *Ca*CS with the ligands at a distance of up to 4.0 Å. A frequency greater than 60% was delineated to identify the most important residues in the stabilization of the ligands. Each ligand has specific geometry and composition, which makes them interact differently with the protein, and therefore, structuring different sites. Thus, a complete list of the residues composing the active site for each ligand is shown in Supplementary Table 1 (Supplementary Material), as a probability of binding. It clearly shows that the residues Ser-16, Thr-55, Ser-125, Ala-126, Arg-127, Arg-337 and Arg-345 are the most important in the anchoring of the substrate and ligands, which means that amino acid replacements at these positions may significantly affect the enzymatic activity and/or cause resistance to the presence of the inhibitors. Since our theoretical results strongly indicated that CS8 and CaCS02 bind to CS, we were also interested to evaluate their effect on the growth of fungi, e.g., *C. albicans* and *P. brasiliensis*. Toward this end, we determined the MIC using fungal strains from *Candida* and *Paracoccidioides*.

Antifungal activity assays

CS8 and CaCS02 were tested with five *Candida* species and *P. brasiliensis* (Pb18). Antifungal activity against *Candida* species was not observed. However, *in vitro* antifungal activity against *P. brasiliensis* was observed for both compounds with MICs of 512 and 32 μg·ml^{-1} for CS8 and CaCS02, respectively (Table 1). The MFC assays showed that the values found for CaCS02 were the same as the MIC, indicating a fungicidal profile (Table 1). However, the most promising result for this compound was the strong synergistic effect observed when combined with AmB. This combination was able to reduce the MIC of AmB fourfold (from 0.5 to 0.125 μg·ml^{-1}) and the MIC of CaCS02 eightfold (from 32 to 4 μg·ml^{-1}) (Supplementary Figure 4).

Binding assays with PbCS

Although ligands CS8 and CaCS02 were identified by virtual screening using modeled *Ca*CS, the binding studies were conducted with the enzyme from *P. brasiliensis* (*Pb*CS) mainly because the ligands were found to be active against *P. brasiliensis* but not *C. albicans*. Because CS8 showed very poor solubility in aqueous buffers, a detailed analysis was only conducted with CaCS02. As shown in Figure 3A, binding to *Pb*CS can be monitored by following the spectral changes occurring upon binding, yielding a dissociation constant of approximately 20 μM. Similarly, CaCS02 inhibited the activity of *Pb*CS exhibiting an IC$_{50}$ = 29 ± 3 μM (Figure 3B). These results nicely fit to our virtual screening and MD simulations.

Cytotoxicity assay

Next, we investigated the toxicity of CaCS02, which was predicted to be low (Supplementary Table 3, Supplementary Data). The cytotoxic effects *in vitro* were also evaluated. For this, two mammalian cell lines were incubated with different concentrations of CaCS02 (Figure 4). The results showed that CaCS02 does not present any cytotoxic effects even at the limit of the evaluated concentrations, in other words, up to 16-times the MIC value.
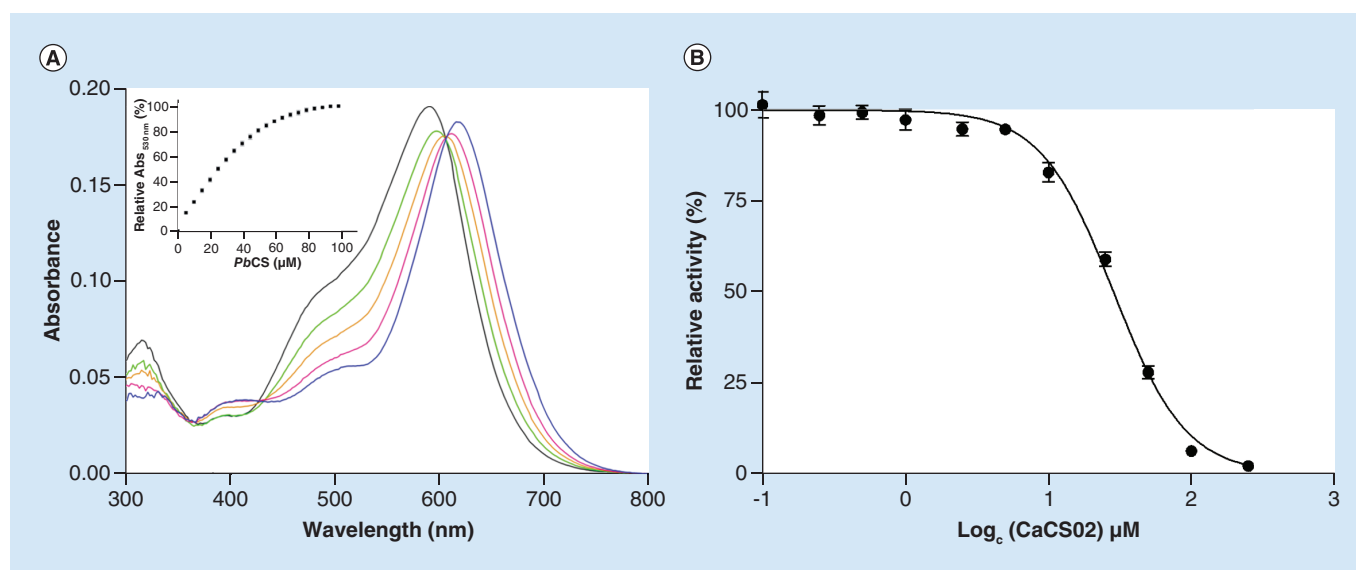
Figure 3.   Binding of *Ca*CS02 to chorismate synthase from *Paracoccidioides brasiliensis* and its inhibitory effect on the catalytic activity of chorismate synthase from *P. brasiliensis*. (A) CaCS02 was dissolved in 50 mM morpholinopropanesulfonic acid, pH 7.5 ($\sim$10 μM, black spectrum) and titrated with *Pb*CS ($\sim$393 μM) using 10 μl aliquots (see inset). Spectral changes were monitored between 300 and 800 nm. The green, orange and pink traces represent intermediate spectra at $\sim$15, $\sim$29 and $\sim$49 μM *Pb*CS, respectively, and the blue trace corresponds to the final spectrum recorded at $\sim$98 μM *Pb*CS. The inset shows the mean absorption changes at 530 nm (from three determinations) as a function of the protein concentration indicating the binding of CaCS02 to *Pb*CS (the standard deviation at each *Pb*CS concentration is shown as error bars). **(B)** The influence of CaCS02 on the activity of *Pb*CS was tested in an assay involving 80 μM 5-Enolpyruvylshikimate-3-phosphate, 4 μM *Pb*CS and 20 μM anthranilate synthase at 37°C. By forward coupling the CS with the anthranilate synthase reaction, *Pb*CS activity could be studied spectroflourometrically by following the increase in fluorescence intensity at 390 nm (λ_{ex} = 340 nm) for 5 min, which correlates with the amount of anthranilate formed in the course of the reaction. The slopes of the initial velocities were determined and plotted as a function of the log of the corresponding CaCS02 concentration. Using a nonlinear curve fit, based on the Hill function, an IC_{50} of 29 $\pm$ 3 μM was determined (rates were determined in triplicate with the standard deviations shown as error bars).
CS: Chorismate synthase; *Pb*CS: Chorismate synthase from *Paracoccidioides brasiliensis*.
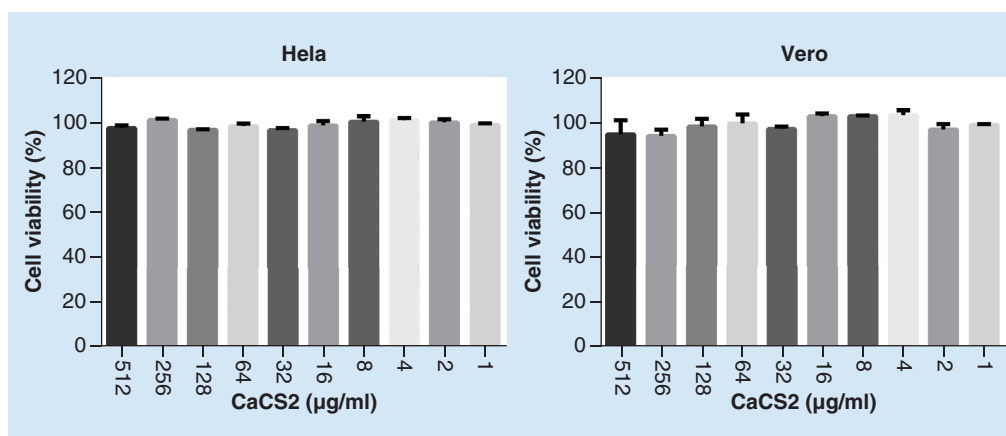

Figure 4.   Evaluation of the *in vitro* cytotoxicity of CaCS02 in two cell lines by means of the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium assay. Several concentrations of CaCS02 were tested (0.0 to 512 μg·ml^{-1}) without showing significant toxicity. The two cell lines evaluated (HeLa and Vero cells) showed high viability after 24 h in contact as CaCS02, even in concentrations 16 $\times$ minimum inhibitory concentration (32 μg·ml^{-1}). The results represent the mean of three independent trials. Error bars indicate their respective standard deviation.

Discussion

Although this study initially focused on Candida CS, the compounds did not cause any effect on this species. This may be linked to some kind of intrinsic resistance of the fungus *C. albicans* to the compounds [51]. In addition, studies have shown that the presence of efflux pumps may further hamper the action of inhibitors [52]. However, the most significant difficulty may be related to the reported differences in the cell wall structure and composition of the two species [53].

The Tanimoto index between the CS8 and the recently reported inhibitor CP1 (Figure 2) is 0.75. Although CS8 did not lend itself to a more detailed analysis, its similarity to CP1 [41] and its antifungal activity, strongly suggests that it may also bind to *Pb*CS, since CP1 was shown to be a potent inhibitor of *Pb*CS, which had *in vivo* antifungal activity similar to the reference drug itraconazole. The lower inhibitory activity found in the MIC trials with CS8 is possibly due to the loss of one of the negatively charged groups compared with EPSP, CP1 and CaCS02, indicating that negative charges are very important for the stabilization of the ligand in the active site.

The strongly synergistic effect observed between CaCS02 and AmB (Table 1) is a significant indication that the two ligands act in different metabolic pathways. Indeed, AmB has been shown to act on ergosterol, in order to alter the permeability of the cytoplasmic membrane [54]; whereas, CaCS02 interferes with the biosynthesis of chorismate, confirming that the two compounds exert their influence on different aspects of cellular processes. AmB is the most common antifungal agent used in severe forms of the disease, but despite its excellent activity in reducing fungal load, it produces highly toxic adverse effects, such as nephrotoxicity and hepatotoxicity [55]. Thus, a molecule, such as CaCS02, capable of enhancing the antifungal activity of AmB may reduce the damage since it could be administered in smaller doses. In order to distinguish that CaCS02 acts as a specific inhibitor of *Pb*CS and not as pan-assay interference compound [56], we also performed binding and activity assays with recombinant *Pb*CS (Figure 3).

The CaCS02 ligand has a negative net charge (-2) and partition coefficient (*x*logP) of -1.25 at pH 7.0. The presence of these negative charges appears to be crucial for the stabilization of this ligand in the active site, since the simulations predicted binding at the same site as the substrate EPSP, which has several negative charges that interact with the enzyme by means of charge–charge and charge–dipole interactions (Figure 2 & Supplementary Table 2, Supplementary Data).

Although other antifungal candidates against *P. brasiliensis* have already been described with lower MIC values [41,42,57], the low cytotoxicity presented by CaCS02 may compensate for the greater dosages required. CaCS02 is a water-soluble azo dye, and to date, this type of molecule has never been described as an inhibitor of CS. Since many azo dyes are commercially available, and the fact that isosteric mimetics may be accessible by chemical synthesis, our research group will continue this line of research in the quest to obtain highly specific CS inhibitors. Since CS is also present in bacteria, plants and apicomplexan parasites, this research may also contribute to the development of new antimicrobials, herbicides and/or antiparasitic (e.g., antimalaria) drugs.

There are currently eight specific CS inhibitors described in the Brenda database [58], all derived from modifications in the structure of the EPSP substrate. The discovery of specific inhibitors for CS whose structures differ from the substrate, makes this work even more interesting.

Conclusion & future perspective

This work presented the modeling, docking and MD simulations using *Ca*CS. The analysis of the modeled structure demonstrated a good quality of the proposed model, as well as its validation for use in virtual screening simulations. Among the selected molecules, we identified the compound Zinc4217277 (CaCS02) and demonstrated its ability to bind and inhibit the enzyme in *in vitro* studies. Furthermore, CaCS02 showed promising *in vitro* antifungal activity against the most virulent strain of the human pathogenic fungus *P. brasiliensis* (Pb18). Its low cytotoxicity and the significant synergistic effect with AmB, which is the reference drug in severe cases of this disease, demonstrated a potential use in a combinatorial drug therapy against PCM.

Author contributions

All the authors performed the experiments, analyzed the data and contributed equally in the writing of the manuscript. FAV Seixas, ES Kioshima and P Macheroux designed the study. All authors read and approved the manuscript.

Acknowledgments

The authors thank LNCC for computational facilities and FINEP/COMCAP/UEM for equipment facilities

Summary points

- The modeled structure of the enzyme chorismate synthase (CS) from *Candida albicans* was used in virtual screening simulations.
- Two molecules, Zinc6445805 and Zinc4217277, also termed CS8 and CaCS02, respectively, were identified in-common by different programs.
- The molecules were active against strains of *Paracoccidioides brasiliensis* in assays of minimum inhibitory concentration (MIC) and minimal fungicidal concentration (MCF).
- The MIC/MFC found were 512 and 32 μg·ml^{-1} for CS8 and CaCS02, respectively, suggesting fungicidal profile.
- Assays of enzyme inhibition were performed with the recombinant CS from *P. brasiliensis*.
- The molecule CaCS02 inhibited the enzyme CS with IC$_{50}$ of 20 μM.
- The molecule CaCS02 presented strong synergism in combination with amphotericin B (Σfic = 0.375).
- No cytotoxic effect was observed when CaCS02 was incubated with HeLa or Vero cells up to the limit of 512 μg·ml^{-1}.

References

Papers of special note have been highlighted as ● of interest; ●● of considerable interest

1. Pianalto KM, Alspaugh JA. New horizons in antifungal therapy. *J. Fungi* 2(4), 24 (2016).
2. Giacomazzi J, Baethgen L, Carneiro LC *et al.* The burden of serious human fungal infections in Brazil. *Mycoses* 59(3), 145–150 (2016).
3. Da Silva Dantas A, Lee KK, Raziunaite I *et al.* Cell biology of *Candida albicans*-host interactions. *Curr. Opin. Microbiol.* 34, 111–118 (2016).
4. Kabir MA, Hussain MA, Ahmad Z. *Candida albicans*: a model organism for studying fungal pathogens. *ISRN Microbiol.* 2012, 538694 (2012).
5. Sheehan DJ, Hitchcock CA, Sibley CM. Current and emerging azole antifungal agents. *Clin. Microbiol. Rev.* 12(1), 40–79 (1999).
6. Martinez R. Epidemiology of paracoccidioidomycosis. *Rev. Inst. Med. Trop. Sao Paulo* 57(Suppl. 19), 11–20 (2015).
7. Bocca AL, Amaral AC, Teixeira MM, Sato PK, Shikanai-Yasuda MA, Soares Felipe MS. Paracoccidioidomycosis: eco-epidemiology, taxonomy and clinical and therapeutic issues. *Future Microbiol.* 8(9), 1177–1191 (2013).
● **Highlights the clinical aspects of paracoccidioidomycosis (PCM), the diversity and particularities of the pathogen, as well as the strategies for therapy.**
8. Turissini DA, Gomez OM, Teixeira MM, Mcewen JG, Matute DR. Species boundaries in the human pathogen Paracoccidioides. *Fungal Genet. Biol.* 106, 9–25 (2017).
9. Martinez R. New trends in paracoccidioidomycosis epidemiology. *J. Fungi* 3(1), 1 (2017).
10. Da Costa MM, Da Silva SHM. Epidemiology, clinical, and therapeutic aspects of paracoccidioidomycosis. *Curr. Trop. Med. Rep.* 1(2), 138–144 (2014).
11. Defaveri J, Joaquim A. Acute form of paracoccidioidomycosis: analysis of thirteen autopsies with emphasis on the pulmonary involvement. *Ann. Rev. Biomed. Sci.* Special issue, 111 (2002).
12. Defaveri J, Joaquim A. Chronic form of paracoccidioidomycosis: analysis of forty autopsies with emphasis on the pulmonary pathogeny. *Ann. Rev. Biomed. Sci.* Special issue, 111 (2002).

13. Bellissimo-Rodrigues F, Bollela VR, Da Fonseca BaL, Martinez R. Endemic paracoccidioidomycosis: relationship between clinical presentation and patients' demographic features. *Sabouraudia* 51(3), 313–318 (2013).

14. Naranjo TW, Lopera DE, Diaz-Granados LR, Duque JJ, Restrepo A, Cano LE. Combined itraconazole-pentoxifylline treatment promptly reduces lung fibrosis induced by chronic pulmonary paracoccidioidomycosis in mice. *Pulm. Pharmacol. Ther.* 24(1), 81–91 (2011).

15. Spampinato C, Leonardi D. Candida infections, causes, targets, and resistance mechanisms: traditional and alternative antifungal agents. *Biomed. Res. Int.* 2013, 204237 (2013).

16. Campoy S, Adrio JL. Antifungals. *Biochem. Pharmacol.* 133(1), 86–96 (2016).

17. Kaltdorf M, Srivastava M, Gupta SK *et al.* Systematic identification of anti-fungal drug targets by a metabolic network approach. *Front. Mol. Biosci.* 3, 22 (2016).

18. Zhou H, Singh NJ, Kim KS. Homology modeling and molecular dynamics study of chorismate synthase from *Shigella flexneri*. *J. Mol. Graph. Model.* 25(4), 434–441 (2006).

19. Tapas S, Kumar A, Dhindwal S, Preeti Kumar P. Structural analysis of chorismate synthase from *Plasmodium falciparum*: a novel target for antimalaria drug discovery. *Int. J. Biol. Macromol.* 49(4), 767–777 (2011).

20. Herrmann KM, Weaver LM. The shikimate pathway. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 50, 473–503 (1999).

21. Fernandes CL, Breda A, Santos DS, Basso LA, Souza ON. A structural model for chorismate synthase from *Mycobacterium tuberculosis* in complex with coenzyme and substrate. *Comput. Biol. Med.* 37(2), 149–158 (2007).

22. Macheroux P, Petersen J, Bornemann S, Lowe DJ, Thorneley RN. Binding of the oxidized, reduced, and radical flavin species to chorismate synthase. An investigation by spectrophotometry, fluorimetry, and electron paramagnetic resonance and electron nuclear double resonance spectroscopy. *Biochemistry* 35(5), 1643–1652 (1996).

   • **Describes a detailed characterization of the properties of chorismate synthase (CS).**

23. Quevillon-Cheruel S, Leulliot N, Meyer P *et al.* Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 279(1), 619–625 (2004).

24. Magrane M. UniProt Knowledgebase: a hub of integrated protein data. *Database* 2011, bar009 (2011).

25. Mount DW. Using the basic local alignment search tool (BLAST). *CSH Protoc.* **2007**(7), pdb.top17, (2007).

26. Berman HM, Battistuz T, Bhat TN *et al.* The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.* 58(1), 899–907 (2002).

27. Eswar N, Webb B, Marti-Renom MA *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* Chapter 5(Unit 5), 6 (2006).

28. Laskowski RA, Rullmannn JA, Macarthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8(4), 477–486 (1996).

29. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J. Mol. Graph.* 14(1), 33–38 (1996).

30. Phillips JC, Braun R, Wang W *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26(16), 1781–1802 (2005).

31. Mackerell AD Jr., Feig M, Brooks CL, 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11), 1400–1415 (2004).

32. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32(11), 2359–2368 (2011).

33. Neese F. The ORCA program system. *Comput. Mol. Sci.* 2(1), 73–78 (2012).

34. Morris GM, Huey R, Lindstrom W *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30(16), 2785–2791 (2009).

35. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* 1263, 243–250 (2015).

36. Kim S, Thiessen PA, Bolton EE *et al.* PubChem substance and compound databases. *Nucleic Acids Res.* 44(D1), D1202–D1213 (2016).

37. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 52(7), 1757–1768 (2012).

38. Knox C, Law V, Jewison T *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39(Suppl. 1), D1035–D1041 (2011).

39. Schomburg I, Chang A, Placzek S *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 41(D1), D764–D772 (2013).

40. Netto CF, Vegas V, Sciannaméa I, Guarnieri D. The polysaccharidic antigen from *Paracoccidioides brasiliensis*. Study of the time of cultivation necessary for the preparation of the antigen. *Rev. Inst. Med. Trop. Sao Paulo* 11(3), 177 (1969).

41. Rodrigues-Vendramini FAV, Marschalk C, Toplak M *et al.* Promising New antifungal treatment targeting chorismate synthase from *Paracoccidioides brasiliensis*. *Antimicrob. Agents Chemother.* 63(1), e01097-18 (2019).

   •• **Describes the *in vitro* and *in vivo* activity of a CS inhibitor against PCM.**

42.   Bagatin MC, Pimentel AL, Biavatti DC *et al.* Targeting the homoserine dehydrogenase of *Paracoccidioides* species for treatment of systemic fungal infections. *Antimicrob. Agents Chemother.* 61(9), e00165-17 (2017).

••   **A new target against PCM is described in this paper.**

43.   Mor V, Rella A, Farnoud AM *et al.* Identification of a new class of antifungals targeting the synthesis of fungal sphingolipids. *MBio* 6(3), e00647 (2015).

44.   Schaller A, Windhofer Amrhein N. Purification of chorismate synthase from a cell culture of the higher plant *Corydalis sempervirens* Pers. *Arch. Biochem. Biophys.* 282(2), 437–442 (1990).

45.   Fitzpatrick TB, Killer P, Thomas RM, Jelesarov I, Amrhein N, Macheroux P. Chorismate synthase from the hyperthermophile *Thermotoga maritima* combines thermostability and increased rigidity with catalytic and spectral properties similar to mesophilic counterparts. *J. Biol. Chem.* 276(21), 18052–18059 (2001).

46.   Capoci IRG, Bonfim-Mendonça PDS, Arita GS *et al.* Propolis is an efficient fungicide and inhibitor of biofilm production by vaginal *Candida albicans. Evid. Based Complement. Alternat. Med.* 2015, 287693 (2015).

47.   Wilkins MR, Gasteiger E, Bairoch A *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552 (1999).

48.   Maclean J, Ali S. The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure* 11(12), 1499–1511 (2003).

49.   Tapas S, Kumar A, Dhindwal S, Kumar P. Structural analysis of chorismate synthase from *Plasmodium falciparum*: a novel target for antimalaria drug discovery. *Int. J. Biol. Macromol.* 49(4), 767 (2011).

50.   Rohr AK, Hersleth HP, Andersson KK. Tracking flavin conformations in protein crystal structures with Raman spectroscopy and QM/MM calculations. *Angew. Chem. Int. Ed. Engl.* 49(13), 2324–2327 (2010).

51.   Arthington-Skaggs BA, Rex JH. Resistance to antifungal agents. In: *Antimicrobial Resistance and Implications for the Twenty-First Century.* Fong WI, Drlica K (Eds.). Springer, Boston, MA, USA, 325–369 (2008).

52.   Ford CB, Funt JM, Abbey D *et al.* The evolution of drug resistance in clinical isolates of *Candida albicans. eLife* 4, e00662 (2015).

53.   Karkowska-Kuleta J, Kozik A. Cell wall proteome of pathogenic fungi. *Acta Biochem. Pol.* 62(3), 339–351 (2015).

54.   Borba HHL, Steimbach LM, Riveros BS *et al.* Cost-effectiveness of amphotericin B formulations in the treatment of systemic fungal infections. *Mycoses* 61(10), 754–763 (2018).

55.   Linden PK. Amphotericin B lipid complex for the treatment of invasive fungal infections. *Expert Opin. Pharmacother.* 4(11), 2099–2110 (2003).

56.   Pouliot M, Jeanmart S. Pan assay interference compounds (PAINS) and other promiscuous compounds in antifungal research. *J. Med. Chem.* 59(2), 497–503 (2016).

57.   Abadio AK, Kioshima ES, Leroux V *et al.* Identification of new antifungal compounds targeting thioredoxin reductase of *Paracoccidioides* genus. *PLoS ONE* 10(11), e0142926 (2015).

58.   Jeske L, Placzek S, Schomburg I *et al.* BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47(D1), D542–D549 (2018).

Supplementary Materials

New inhibitors of the enzyme chorismate synthase present antifungal activity against *Paracoccidioides brasiliensis*.

Paulo Sérgio Alves Bueno[1], Franciele Abigail Vulgron Rodriguez-Vendramini[2], Marina Toplak[3], Peter Macheroux[3], Érika Seki Kioshima[2], Flavio Augusto Vicente Seixas[1]*

[1]Department of Technology, Universidade Estadual de Maringá, Av. Ângelo Moreira da Fonseca, 1800, 87506-370 Umuarama, PR, Brazil

[2]Department of Clinical Analysis and Biomedicine, Universidade Estadual de Maringá, Av. Colombo 5790, 87020-900 Maringá, PR, Brazil.

[3]Institute of Biochemistry, Graz University of Technology, Petersgasse 12/2, 8010 Graz, Austria.

*Correspondence author: favseixas@uem.br
Av. Ângelo Moreira da Fonseca, 1800, Jd Universitário
87506-370, Umuarama, PR, Brazil
Fone: (+55) 44 36219337
Fax: (+55) 44 36219326
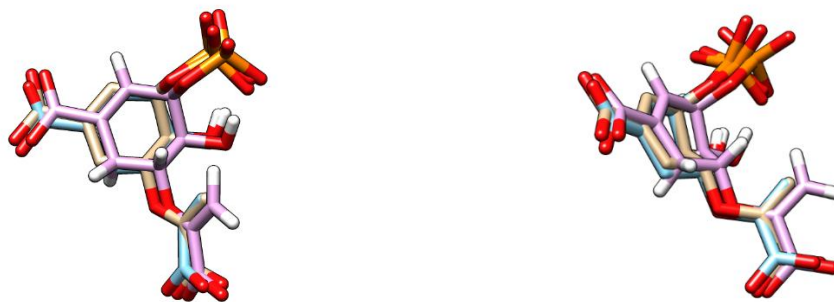
Figure S1. Redocking of the EPSP substrate by the Autodock program. The modeled reference structure is shown in coral; The poses found by redocking are depicted in blue and purple:.
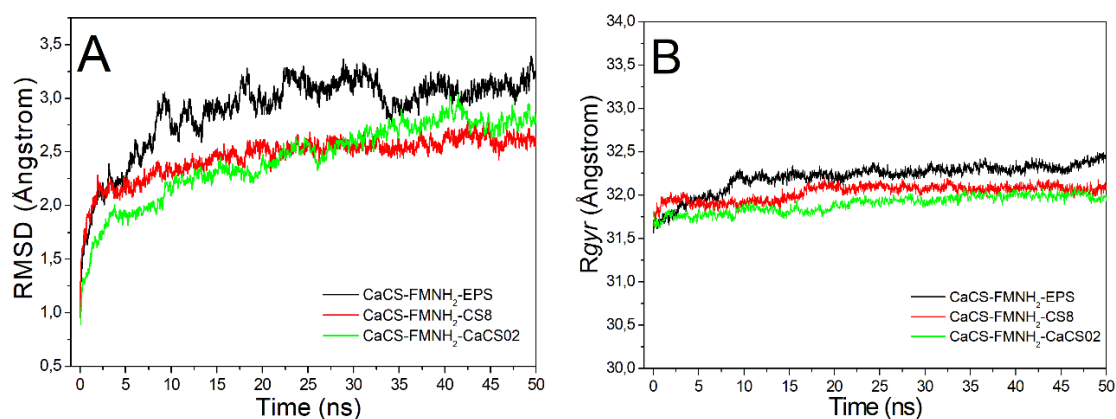
Figure S2: Protein behavior throughout the simulation. The *Ca*CS-FMNH$_2$-EPSP complex (**black**), and the *Ca*CS-FMNH$_2$-CS8 complex (**red**) and the *Ca*CS-FMNH$_2$-CaCS02 complex (**green**). (A) RMSD calculated for the protein main chain atoms. (B) Radius of gyration (R$_{gyr}$) calculated for the protein main chain atoms.
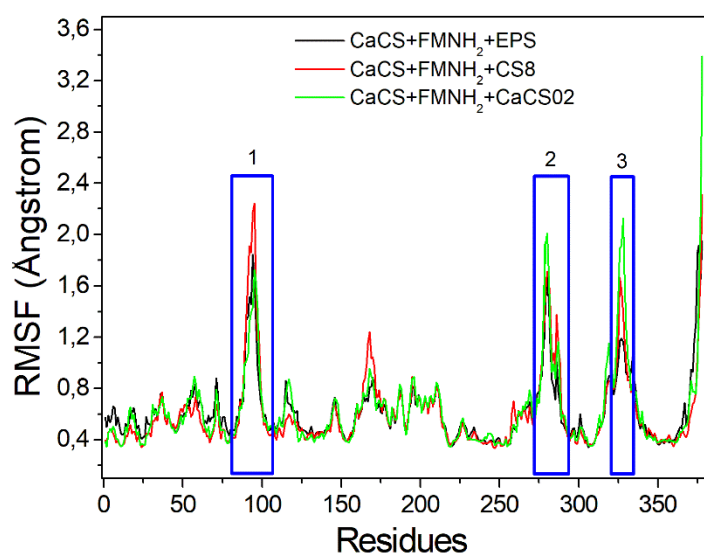
Figure S3. RMSF of the Cα atoms of the *Ca*CS in the last 5 ns of simulation (equilibrium region). The blue boxes highlight the regions with the greatest degree of flexibility.
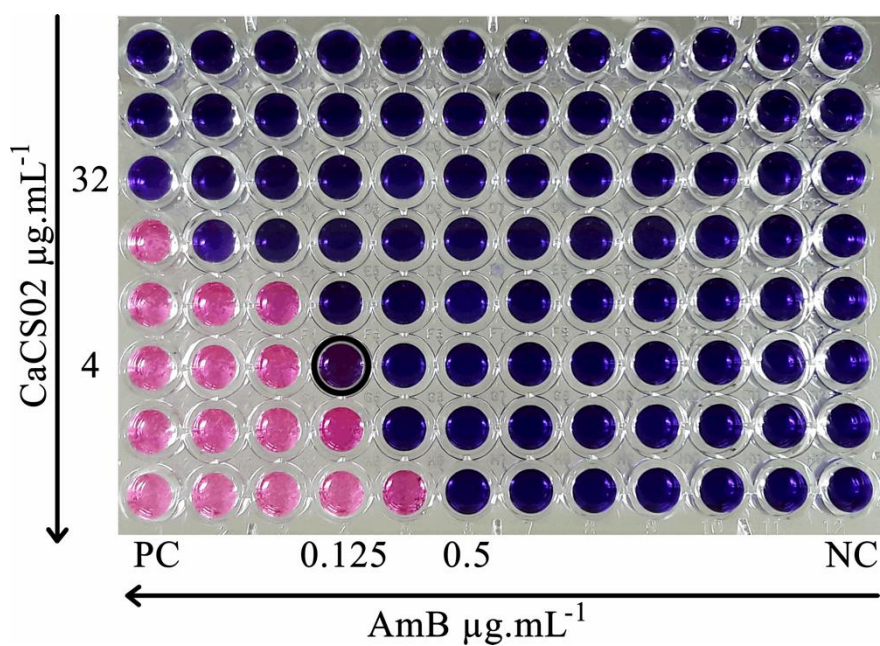
Figure S4. Checkerboard plate of the combination of AmB and CaCS02 assayed in *P. brasiliensis* (Pb18) strains. We found values of $MIC_{AmB}=0.5$ µg·mL^{-1} and $Comb_{AmB}=0.125$; $MIC_{CaS02}=32$ µg·mL^{-1} and $Comb_{CaCS02}=4$ µg·mL^{-1}. The black circle indicates the point of combination between the compounds tested. The 2-fold drop in inhibition concentration for each compound revealed a strong synergistic interaction with $\Sigma FIC=0.375$. The plate was incubated at 35 °C for 7 days, on the 6th day resazurin (0.01%) was added to each well. Blue color indicates fungal non-viability while pink color indicates viable fungal growth. (PC) Positive control; (CN) negative control.

PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS

Pentose phosphate pathway
Photosynthesis

Quinate

7P-2-Dehydro-3-deoxy-D-arabino-heptonate
1.1.5.8 1.1.1.282 1.1.1.24

2-Amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid

D-Erythrose 4-phosphate 2.5.1.54 4.2.3.4 1.4.1.24 2.2.1.10 L-Aspartate 4-semialdehyde

Glycolysis Phosphoenol-pyruvate 3-Dehydroquinate 4.2.1.10 6-Deoxy-5-keto-fructose-1-phosphate 2.2.1.11 D-Fructose-1-phosphate

2.2.1.11 D-Fructose-1,6-diphosphate

Protocatechuate 4.2.1.118 3-Dehydro-shikimate 1.1.1.25 1.1.5.8 1.1.1.282

Benzoxazinone biosynthesis

Indole 4.2.1.20 Indole diterpene alkaloid biosynthesis Shikimate Shikimate pathway

4.2.1.20

L-Tryptophan 4.2.1.20 (3-Indoyl)-glycerolphosphate 2.7.1.71 Shikimate 3-phosphate

Tryptophan metabolism 4.1.1.48

Indole alkaloid biosynthesis 1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5-phosphate 2.5.1.19 5-O-(1-Carboxyvinyl)-3-phosphoshikimate

Staurosporine biosynthesis 5.3.1.24 PRPP 2.4.2.18 Anthranilate 4.2.3.5

N-(5-Phospho-β-D-ribosyl)-anthranilate

Acridone alkaloid biosynthesis 4.1.3.27 Chorismate Biosynthesis of siderophore group nonribosomal peptides Folate biosynthesis

5.4.99.5 4.1.3.45 3.3.2.13

Prephenate 3-Hydroxy-benzoate 3,4-Dihydroxy-cyclohexa-1,5-dienoic acid Biosynthesis of enediyne antibiotics

4.2.1.91 4.2.1.51 2.6.1.57 1.3.1.12 1.3.1.13
2.6.1.78 2.6.1.79 4-Hydroxy-phenylpyruvate

3-Benzylmalate 2-Benzylmalate HphCD HphA Phenyl-pyruvate L-Arogenate Ubiquinone and other terpenoid-quinone biosynthesis

HphB

2.6.1.1 2.6.1.5 2.6.1.1 2.6.1.5
2.6.1.9 2.6.1.57 2.6.1.9 2.6.1.57
1.4.3.2 2.6.1.58 1.4.1.20 2.6.1.58 1.4.1.20 1.4.3.2

2-Oxo-4-phenyl-butanoate

Phenylpropanoid biosynthesis 4.2.1.51 1.3.1.43 Tyrosine
4.2.1.91 1.3.1.78
Phenyl-alanine 1.3.1.79

Homo-phenylalanine Biosynthesis of tropane, pyrrolizidine, nicotine and piperidine alkaloids 1.14.16.1 Puromycin biosynthesis Isoquinoline alkaloid biosynthesis

Glucosinolate biosynthesis Phenylpropanoid biosynthesis

Phynylalanine metabolism Tyrosine metabolism
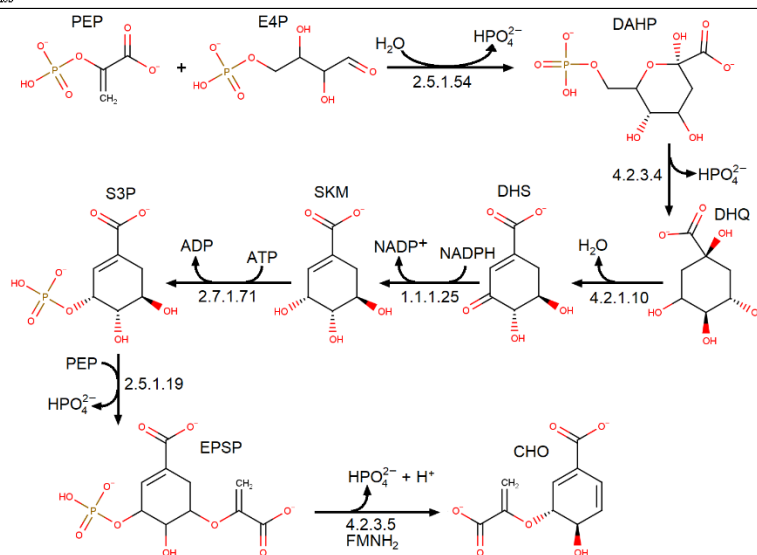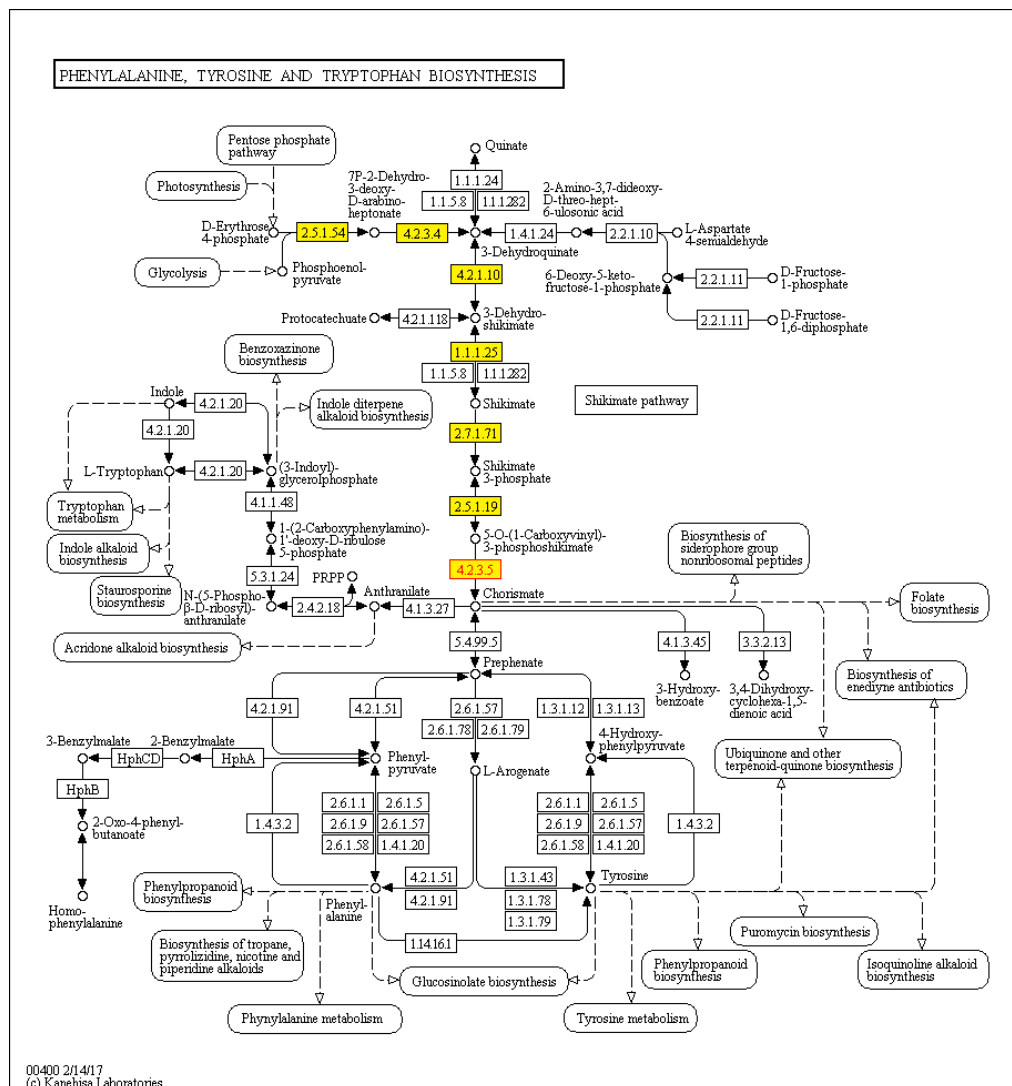
00400 2/14/17
(c) Kanehisa Laboratories

Figure S5. A) The fungal pathway for the biosynthesis of phenylalanine, tyrosine, and tryptophan (KEGG map00400), showing the enzymes of the shikimate pathway (yellow boxes) and the chorismate synthase (EC: 4.2.3.5, red text). Source: KEGG (https://www.genome.jp). B) Detailed view of the shikimate pathway. PEP = phosphoenolpyruvate, E4P = D-erythrose 4-phosphate, DAHP = 3-deoxy-D-arabino-hept-2-ulosonate 7-phosphate, DHQ = 3-dehydroquinate, DHS= 3-dehydroshikimate, SKM = shikimate, S3P = 3-phospho-shikimate, EPSP = 5-O-(1-carboxyvinyl)-3-phosphoshikimate, CHO = chorismate.
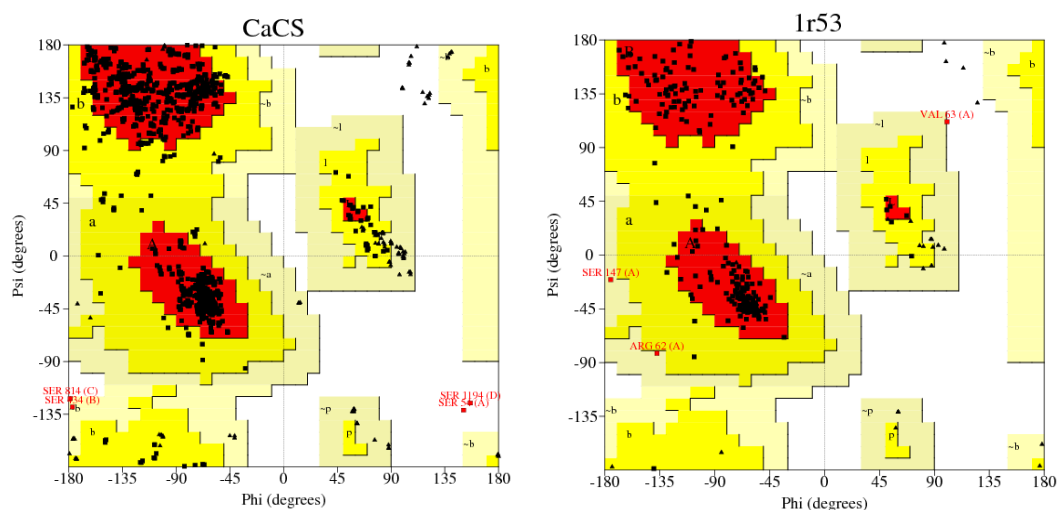
63

Figure S6. Ramachandran plots of the modeled structure of *Ca*CS and of the main template (PDB ID: 1r53). *Ca*CS has 91.7% of the residues in the most favored regions, 8.0% in the additional allowed regions, 0.2% in the generously allowed region, and 0.2% in the disallowed regions. The four residues highlighted in red all refer to Ser54, since the protein was modelled as a homotetramer, which means that there is just one residue outlier per chain. The template 1r53 has 91.4% of the residues in the most favored regions, 7.3% in the additional allowed regions, 0.9% in the generously allowed region and 0.4% in the disallowed regions. The three residues shown in red are not the same, which means that there are three outliers per chain. The comparison indicates that the model has better stereochemical quality than the template.

Table S1: Contact frequency between *Ca*CS residues and the evaluated ligands within the range of up to 4 Å. Residues in bold have a contact frequency with ligand of greater than 60%.

Residues	EPSP	CS8	CaCS02	Residues	EPSP	CS8	CaCS02
Ser-16	**0,71**	**0,67**	**0,66**	Tyr-99	-	-	0,17
His-17	-	0,53	0,57	Arg-101	0,46	0,96	1,00
Arg-46	0,99	0,31	1,00	His-104	-	-	0,06
Arg-47	0,09	-	0,18	Ser-118	-	-	0,08
Lys-52	0,19	-	-	Gly-121	**-**	**-**	0,02
Ser-54	0,62	0,20	0,72	Ser-124	-	0,29	0,01
Thr-55	**0,81**	**0,99**	**0,77**	**Ser-125**	**0,75**	**0,70**	**0,70**
Pro-56	0,61	0,41	0,38	**Ala-126**	**0,97**	**0,80**	**0,89**
Asp-58	-	0,10	-	**Arg-127**	**0,96**	**0,62**	**0,97**
Glu-59	0,08	0,42	0,18	Thr-129	-	0,27	0,31
Lis-86	0,19	-	-	Ile-130	0,06	-	0,02
Asp-87	-	0,26	0,22	Ala-314	-	-	0,02
His-88	0,14	-	0,03	Thr-315	0,12	0,93	0,56
Arg-89	-	0,22	0,03	Ala-334	-	0,19	0,45
Pro-90	-	0,33	0,54	Arg-335	-	0,40	0,17
Tyr-93	0,05	0,31	0,26	Gly-336	-	0,49	0,22
Ser-94	-	-	0,01	**Arg-337**	**0,63**	**1,00**	**0,74**
Glu-95	-	-	0,14	His-338	**-**	**-**	0,19
Thr-96	-	-	0,13	Asp-339	-	-	0,06
Asp-97	-	-	0,35	**Arg-345**	**1,00**	**0,69**	**0,99**
Leu-98	-	-	0,16	**FMN**	**0,77**	**1,00**	**0,99**

Table S2. Main electrostatic contacts between *Ca*CS and the ligands evaluated by molecular dynamics within a range of 4.0 Å and contact frequency greater than 60%.

Protein residues	EPSP		CS8		CaCS02	
	Atom	Distance (Å)	Atom	Distance (Å)	Atom	Distance (Å)
Ser16 Oγ	-	-	-	-	O2	2.4
Ser125 N	O10	3.0	O3	4.0	-	-
Ser125 Oγ	-	-	-	-	O11	2.4
Ala126 N	O10	3.0	-	-	O11	3.1
Arg127 NH1	O8	2.6	CL1	3.1	O3	3.1
Arg337 NH1	O7	2.6	CL1	4.0	O4	3.2
Arg337 NH2	O7	2.7	CL1	3.7	O4	2.9
Arg345 Nε	O2	2.9	O1	2.9	O10	2.7
Arg345 NH2	O2	2.7	O	3.5	O10	2.9
Arg345 NH2	O1	3.0	-	-	-	-
FMNH₂ O9	O10	3.6	O3	4.8	O11	3.5

Table S3. Toxicity predictions for hits found in virtual screening by admetSAR [58].

Toxicity	CaCS02		CS8	
	result	probability	result	probability
Carcinogens	Carcinogens	0.8384	Non-carcinogens	0.6517
Carcinogenicity (Three-class)	Non-required	0.5648	Non-required	0.6256
Honey Bee Toxicity	Low HBT	0.5419	Low HBT	0.7178
Biodegradation	Not ready biodegradable	0.8416	Not ready biodegradable	1.0000
Acute Oral Toxicity	III	0.5527	III	0.6163
Model	value	unity	value	unity
Rat Acute Toxicity (LD_{50})	2.3122	mol/kg	2.4709	mol/kg
Fish Toxicity (pLC_{50})	1.1239	mg/L	0.6797	mg/L
Tetrahymena Pyriformis Toxicity ($pIGC_{50}$)	0.3619	ug/L	1.3225	ug/L

58. Cheng F, Li W, Zhou Y, et al. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* 52(11), 3099-3105 (2012).

Artigo 4: The structure of viral cathepsin from *Bombyx mori Nuclear Polyhedrosis Virus* as a target against grasserie: docking and molecular dynamics simulations.

The structure of viral cathepsin from *Bombyx mori Nuclear Polyhedrosis Virus* as a target against grasserie: docking and molecular dynamics simulations

Paulo Sérgio Alves Bueno, Débora Carina Biavatti, Alex Sandro Gularte Chiarello, Verônica Aureliana Fassina, Maria Aparecida Fernandez & Flávio Augusto Vicente Seixas

Taylor & Francis
Taylor & Francis Group

Check for updates

The structure of viral cathepsin from *Bombyx mori Nuclear Polyhedrosis Virus* as a target against grasserie: docking and molecular dynamics simulations

Paulo Sérgio Alves Bueno[a], Débora Carina Biavatti[a], Alex Sandro Gularte Chiarello[b], Verônica Aureliana Fassina[b], Maria Aparecida Fernandez[b] and Flávio Augusto Vicente Seixas[a] (iD)

[a]Departamento de Tecnologia, Universidade Estadual de Maringá, Umuarama, Brazil; [b]Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Maringá, Brazil

Communicated by Ramaswamy H. Sarma

ABSTRACT

The viral cathepsin from *Bombyx mori* Nuclear Polyhedrosis Virus (*Bm*NPV-*Cath*) is a broad-spectrum protease that participates in the horizontal transmission of this virus in silkworm by facilitating solubilization of the integument of infected caterpillars. When a *B. mori* farm is attacked by *Bm*NPV, there are significant sericultural losses because no drugs or therapies are available. In this work, the structure of viral cathepsin *Bm*NPV-*Cath* was used as a target for virtual screening simulations, aiming to identify potential molecules that could be used to treat the infection. Virtual screening of the Natural Products library from the Zinc Database selected four molecules. Theoretical calculations of $\Delta G_{binding}$ by the molecular mechanics Poisson–Boltzmann surface analysis (MM-PBSA) method indicated that the molecule Zinc12888007 (Bm5) would have high affinity for the enzyme. The *in vivo* infection models of *B. mori* caterpillars with *Bm*NPV showed that treatment with a dose of 100 μg Bm5 dissolved in Pluronic-F127 0.02% was able to reduce the mortality of caterpillars in 22.6%, however, it did not impede the liquefaction of dead bodies. Our results suggest a role of *Bm*NPV-*Cath* in generating a pool of amino acids necessary for viral replication and indicate a mechanism to be exploited in the search for treatments for grasserie disease of the silkworm.

Introduction

Sericulture is an agroindustrial activity that contributes to the economy of several countries (Inserco.org, 2018; Ribeiro, Brancalhao, Torquato, & Fernandez, 2009). This activity revolves around *Bombyx mori*, the silk-producing insect. Because of this, producers face serious problems when pests caused by protozoa, fungi, bacteria, or viruses occur in silkworms, viruses being responsible for the greatest losses in production (Shobahah et al., 2017). Among the viruses that attack (infect) *B. mori*, the most important are baculoviruses such as *Bombyx mori Nuclear Polyhedrosis Virus (Bm*NPV), belonging to the Baculoviridae family, which occurs as two phenotypically distinct virus groups throughout the infection cycle: budded viruses (BVs), which are extracellular viruses, and occluded viruses, also known as polyhedra or granules, depending on the genus (Herniou, Olszewski, Cory, & O'Reilly, 2003; Murphy et al., 1995). This virus is the cause of the disease known as grasserie, responsible for losses of approximately 20% of the worldwide silkworm crop, which can reach between 70% and 100% at the local producer level (Brancalão & Ribeiro, 2003).

The caterpillar is infected by ingestion of polyhedra (occluded viruses) present in mulberry leaves, which are the caterpillars' only food source. Under the alkaline conditions of the caterpillar midgut, the polyhedra release viral particles, which cross the membrane of the columnar cells and are carried to the nucleus of the cell through the cytoskeleton (Volkman & Keddie, 1990). The genetic material is released in the nucleus, where viral gene transcription and genome replication are initiated, culminating in the release of new nucleocapsids to infect cells of the tracheal system and hemolymph (Funk & Consigli, 1992). The other form of infection is the secondary one, characterized by the transport of viral particles directly to the cells of the tracheal system and hemolymph, where infected epithelial cells produce the first viral phenotype, the BVs, which will infect other cells by receptor-mediated endocytosis, in which the receptor recognizes the viral protein GP64. This protein is present in the virus capsid and its recognition by cellular receptors mediates the fusion of the viral membrane with the endocytic membrane after endocytosis, culminating with the entry of the virus into the cytoplasm of the infected cell (Engelhard, Kam-Morgan, Washburn, & Volkman, 1994).

When the caterpillar dies, its integument is disintegrated, thus liberating a great amount of polyhedra (polyhedral viruses) into the environment, which serve as the inoculum for infection of other insect hosts (Volkman & Keddie, 1990). These processes are facilitated by the presence of viral hydrolytic enzymes such as viral cathepsin (V-Cath) and chitinase (chiA) (Ohkawa, Majima, & Maeda, 1994; Slack, Kuzio, & Faulkner, 1995), which are essential for horizontal transmission of the infection (Hom & Volkman, 2000).

CONTACT Flávio Augusto Vicente Seixas ✉ favseixas@uem.br

ⓑ Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2018.1521344.

In studies where the v-cath and chiA genes were silenced in a recombinant baculovirus, Hawtin et al. observed that the deletion of these genes did not affect viral replication in cell culture nor in caterpillars, but the virus lost the ability to liquefy the body of the insect (Hawtin et al., 1995). A similar effect was observed when the v-cath gene was deleted from the genome of the baculovirus *Autographa californica multiple nuclear polyhedrosis virus* (*Ac*MNPV), thus showing the importance of this enzyme in the role of viral transmission (Slack et al., 1995). In another study using protease inhibitor E-64 *[trans-epoxysuccinylleucylamido-(4-guanidino)butane]* in cell cultures, the cysteine protease activity of *Bm*NPV virus was inhibited, and deletion of the *Bm*NPV-*Cath* had no significant effect on viral multiplication or polyhedra production, indicating that *V-Cath* is also not essential for replication of the virus *in vitro* (Ohkawa et al., 1994).

These studies demonstrate that *Bm*NPV-*Cath* plays an important role in viral transmission. The selective inhibition of this enzyme can interrupt the liquefaction of the integument and consequently interfere in the process of horizontal transmission of the virus, reducing the number of infected caterpillars, which facilitates the production management and reduces losses.

Therefore, the objective of this work was to model the structure of the enzyme *Bm*NPV-*Cath* (EC 3.4.22.50) in complex with the cathepsin inhibitor *4-Morpholin-4-Yl-Piperidine-1-Carboxylic Acid [1-(3-Benzenesulfonyl-1-Propyl-Allylcarbamoyl)-2-Phenylethyl]-Amide*, in order to use it as a drug target for virtual screening simulations, aiming to identify inhibitors of this enzyme.

Material and methods

Sequence identification and model building

The amino acid sequence of *Bm*NPV-*Cath* (Uniprot: p41721) was analyzed by using InterProScan (Mulder & Apweiler, 2007) to identify structural motifs. The search for structural models was carried out in two ways, *via* BLASTp (Mount, 2007) and by structural alignment by means of folding recognition *via* pGenThreader (Lobley, Sadowski, & Jones, 2009), which also predicted the secondary structure. Then, the amino acid sequence was submitted to the server EDBCP (Cheng, Saigo, & Baldi, 2005) to evaluate the presence of possible disulfide bonds.

The crystallographic structures of the monomeric forms of cathepsin F (PDBid: 1m6d), cathepsin K (PDBid: 3kwz), and cathepsin L1 (PDBid: 2xu3), all from *Homo sapiens*, served as templates for protein alignment. The ligand of cathepsin F (PDBid: 1m6d), *4-Morpholin-4-Yl-Piperidine-1-Carboxylic Acid [1-(3-Benzenesulfonyl-1-Propyl-Allylcarbamoyl)-2-Phenylethyl]-Amide* (Myp, Zinc14880279), a protease inhibitor, was modeled as the ligand of *Bm*NPV-*Cath*, by means of Modeller v.9.13 (Eswar et al., 2006) software, where 2000 models were generated and the best five were classified by means of the Modeller Dope Score. The best model was selected from these five based on stereochemical evaluation by the program Procheck (CCP4, 1994). From this model, 500 new models were generated, aiming to correct residues that were positioned in non-permitted regions of the Ramachandran plot. The final model was selected and re-evaluated among them using the same criteria (Dope score and Ramachandran plot).

Energy minimization

The final Modeller output model was subjected to 30,000 steps of energy minimization by the conjugate gradient method through the NAMD2 program (Phillips et al., 2005) so that the interactions of the protein with the ligand were maximized. For this, the whole complex was immersed in a periodic box with limits at least 10 Å away from the outermost surface of the molecule and with sufficient sodium counterions to neutralize the system charges. The Charmm C35b2-C36a2 force field (Mackerell, Feig, & Brooks, 2004) was used for the polypeptide chain while the ligand force field was generated in the same format by the SwissParam server (Zoete, Cuendet, Grosdidier, & Michielin, 2011). This minimized structure was used as the input parameter for docking and molecular dynamics (MD) simulations.

Redocking and virtual screening

The minimized structure of *Bm*NPV-*Cath* was used in redocking simulations in order to validate the protocol to be used in virtual screening. In this procedure, the programs Vina (Trott & Olson, 2009) (implemented in Pyrx-0.9 graphical interface (Wolf, 2009)) and Molegro virtual Docker 6.0 (Thomsen & Christensen, 2006) were used. In the Vina program, the standard algorithm for search and rank was used. The search box was centered on ligand with sizes of 15, 20, and 20 in *x*, *y*, and *z*, respectively. The Molegro program uses the search algorithm *Iterated Simplex* in combination with the rank function *Plants Score (Grid)*, and a search radius of 11 Å centered on the ligand. The docking results of the modeled ligand Myp were reproducible in both programs, giving a root mean square deviation (*rmsd*) of 1.7 ± 0.52 Å and mean scores of -8.8 (Vina) and -36.8 (Molegro). After the validation, the docking protocol was applied in the virtual screening simulations, which were performed from two libraries, one composed of compounds with 80% structural similarity to the cathepsin inhibitors and the other formed by the purchasable natural products from the Zinc database (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012). All libraries were initially evaluated by the Vina program, where the molecules ranked higher than Myp were used in four other replicates using both programs. The molecules selected in common with 100% reproducibility were evaluated by the ADMETox criteria through the AdmetSar server (Cheng et al., 2012).

The structure of the cathepsin synthetic substrate known as *Z-Phe-Arg 7-amido-4-methylcoumarin* (Zpa, Zinc14961466) was used as a control in the simulations for comparison purposes, since there are no data available in the literature about the kinetic parameters regarding V-*Cath* with Myp.

MD simulations

The MD simulations of the protein in Apo form and complexed with the ligands were performed with the Gromacs 4.5.7 program (Hess, Kutzner, van der Spoel, & Lindahl, 2008) using a Gromos96 v43a1 force field (Scott et al., 1999) with the residues adjusted to their expected protonation states at pH 7.0. The force fields of the ligands were generated in the same format by the Prodrg2 server (Schuttelkopf & van Aalten, 2004). Next, the protein was immersed in a cubic periodic box with limits at least 10 Å away of the outermost surface of the molecule. A sufficient amount of NaCl was added to neutralize the system charges and adjust the final concentration to 150 mM NaCl. Initially the complexes were minimized by the steepest descent method until convergence to the lower energy state, which was used for simulation of molecular dynamics during 13 ns under NPT conditions where temperature was set to 300 K and pressure set to 1 atm.

The simulations were analyzed from the trajectory file in terms of the *rmsd* and radius of gyration (r_{gyr}), of the main chain atoms of the protein with the minimized initial structure as a reference. The fluctuation of the alpha carbons from each residue (root mean square fluctuation, *rmsf*), as well as the contact frequency of the protein residues with the ligand, were evaluated in the last 2 ns of the simulation (thermodynamic equilibrium region). The free energy of binding between protein and ligands ($\Delta G_{binding}$) was calculated from MD trajectory for each complex using the Molecular Mechanics Poisson–Boltzmann Surface Analysis method implemented in the program g_mmpbsa (Baker, Sept, Joseph, Holst, & McCammon, 2001; Kumari, Kumar, Lynn, & Consort, 2014), where, in general terms, the free energy of binding from the protein with the ligand in a solvated system can be expressed as:

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \tag{1}$$

Biological evaluation of the compounds

Experiments with *B. mori* caterpillars were carried out in the Laboratory of Genetic Improvement of Silkworm at the State University of Maringá. Commercial hybrid caterpillars of *B. mori* were supplied by the company Fiação de Seda Bratac S/A (PR, Brazil). The hybrid caterpillars were bred inside plastic boxes in a controlled temperature room (25 ± 1 °C) with a 14 h:10 h (light:dark) photoperiod and relative humidity around 70% (±10%) until the end of their life cycle, being fed three times a day with fresh mulberry leaves (*Morus sp*).

Caterpillars on the first day of the fifth instar (24 h after ecdysis), which were not yet fed, were used for *in vivo* viral infection assays. On this day, infection was carried out by means of an inoculum of *Bm*NPV, isolated from the geographic region of Paraná-Brazil, to a concentration of 3×10^6 COPs/mL (COPs = polyhedral occlusion bodies per milliliter). A volume of 10 μL viral suspension was dropped onto shaped leaf discs of mulberry (16 cm² diameter) and provided as food for the caterpillars of the infected group. During feeding, the caterpillars remained individually confined within disposable plastic cups to ensure the ingestion of the entire viral suspension (Figure S1).

The inhibitors tested Bm1 (10 mg), Bm2 (20 mg), Bm4 (10 mg), and Bm5 (20 mg) were purchased from InterBioScreen (Russia). Solutions containing inhibitors were prepared in two distinct ways for two different sets of trials. The reference inhibitor Myp is not purchasable, therefore biological assays performed in the presence of this protease inhibitor was not possible.

For the initial trial, all inhibitors were diluted to a final concentration of 10 mg mL⁻¹ in 10% DMSO. For this assay, the caterpillars were divided into four groups with 10 caterpillars each, and the assays were performed in triplicate (or duplicate), totaling 30 (or 20) caterpillars per group. The groups were divided into negative control 1 (no virus inoculation and treated with water), control DMSO (no virus inoculation and treated with DMSO 10%), infected control (infected with *Bm*NPV and treated with water), and treatment (infected with *Bm*NPV and treated with the respective inhibitor dissolved in DMSO 10%). On the fourth day post-infection, which corresponds to fifth day of the fifth instar, all experimental groups received 10 μL (100 μg dose) of the respective solution deposited on mulberry leaves, as previously described.

For the second trial, only Bm5 was tested. The compound was dissolved to a final concentration of 10 mg mL⁻¹ in the presence of 0.02% Pluronic-F127 in distilled water (vehicle). Five groups containing 20 caterpillars each were identified as follows: Negative control 2 (no virus inoculation and treated with water), Pluronic control (no virus inoculation and treated with 0.02% Pluronic-F127 in water), treatment vehicle (infected with *Bm*NPV and treated with vehicle 0.02% Pluronic-F127), infected control 2 (infected with *Bm*NPV and treated with water), and Bm5 (infected with *Bm*NPV and treated with Bm5 in 0.02% Pluronic-F127). On the fourth day after infection, which was the fifth day of the fifth instar, all groups received 10 μL of the respective solution deposited on mulberry leaves, as previously described. The trials were performed in quadruplicate, totaling 80 caterpillars per group. Statistical analyzes were performed using ANOVA.

Results and discussion

The amino acid sequence of the cysteine protease *Bm*NPV-Cath was obtained from the Uniprot database. It was predicted by the EDBCP server to have three disulfide bonds between residues (21–62, 55–95, and 150–198) of a total number of 323 residues, with residues 1–16 referring to a signal peptide for the secretory pathway. The amino acids 17–112 refer to an inhibition peptide, which is removed during the maturation process of the protein, so that the enzyme does not become active in the intracellular medium.

In the interactive modeling process, only residues from 113 to 323 that corresponded to the active tertiary structure of the enzyme were considered. In this way, the crystallographic structures of human cathepsins F (Somoza, Palmer, & Ho, 2002), K (Rankovic et al., 2010), and L1 (Hardegger et al., 2011) (PDBids: 1m6d, 3kwz, and 2xu3, respectively) of which
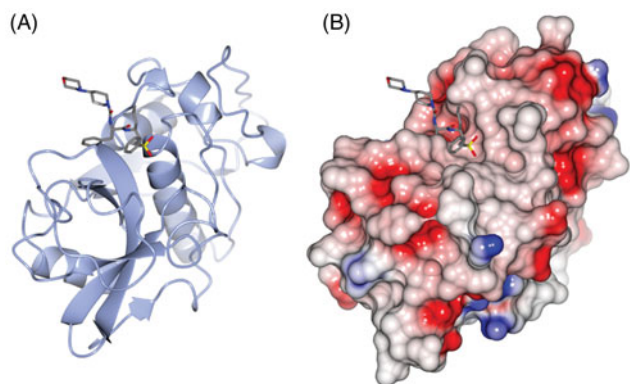
Figure 1. Tertiary structure of *Bm*NPV-*Cath* bound to the inhibitor Myp (gray) represented as ribbons (A) and by electrostatic colored surface (B). Figure generated by CCP4mg program (McNicholas, Potterton, Wilson, & Noble, 2011).

the set identity with *Bm*NPV-*Cath* was 42%, enough to allow homology modeling (Rost, 1999; Yang & Honig, 2000). Alignment with multiple templates allowed the entire protein sequence of the *Bm*NPV-*Cath* to be covered and thus fully modeled from structural element information (Figure S4). This fact is important, since it avoids the introduction of inconsistencies into the model. The structure of *Bm*NPV-*Cath* was modeled in the presence of the inhibitor *vinyl sulfone* (Myp), an irreversible inhibitor of cathepsin F, which covalently binds to catalytic Cys at the active site (Somoza et al., 2002). The structure of the inhibitor was copied into the configuration in which it appeared in the original template, so that the experimental information on its binding mode was conserved, and the inconsistencies arising from a modeling process could be eliminated. Modeling the protein in the presence of a known ligand (substrate or inhibitor) allows the clear identification of the binding site where the docking simulation is to be performed, and also serves as a structural guide for identifying the best pose of an unknown ligand. The ligand Myp was modeled since it is a protease inhibitor which was co-crystallized with the template enzyme. In this way, the experimental information of its docking pose was copied from the template to the model.

The structure of the templates was superimposed to *Bm*NPV-*Cath* and the r.m.s.d. calculated for the main chain atoms from the equivalent regions (excluding regions of insertions and deletions) was 3.24, 3.12, and 3.50 Å for the templates 1m6d, 2xu3, and 3kwz, respectively. From this superimposition, it was possible to observe that the structure around the active site remained well defined (Figure S2). The final model (Figure 1) showed 100% of the residues in the allowed regions of the Ramachandran plot, 90.4% in favorable regions, and 9.6% in the additional allowed regions. However, the modeling program does not have the force field information for the ligand and therefore could not maximize its non-covalent interactions with the protein. Because of this, the modeled structure of the complex *Bm*NPV-*Cath*-Myp was minimized by the conjugate gradient method, in order to leave the structure of the active site as adapted as possible to the presence of the ligand, which we believe is the most suitable conformation for the docking simulations (Zacharias, 2003). In addition, the presence of the
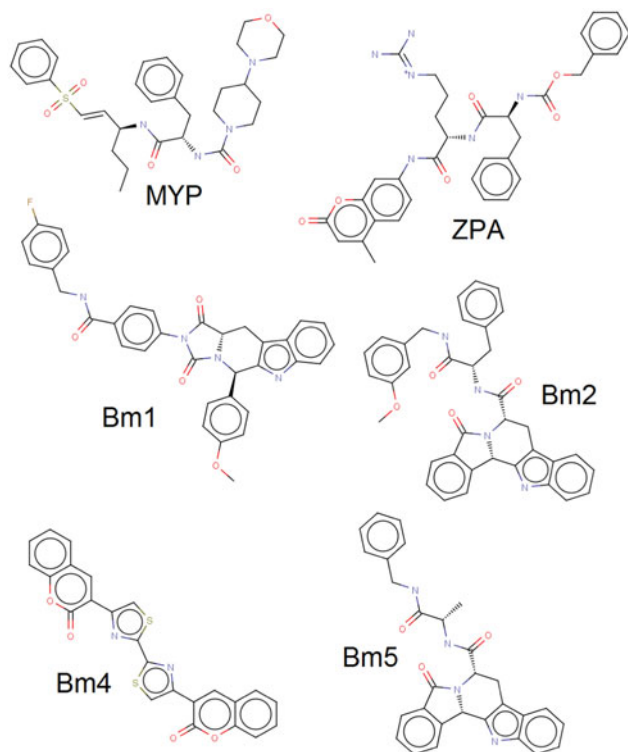

Figure 2. Chemical structures of the reference ligands (Myp and Zpa) and the inhibitors selected by virtual screening.

ligand provides a clear identification of the binding site region, acting as a structural guide for selecting the best pose of an unknown ligand in the virtual screening simulations, which provides additional screening criteria.

Virtual screening

The redocking simulations indicate that the *rmsd* of the ligand Myp oscillates substantially in the region of the radical group *piperidin-morpholin*, which lies on the outer face of the molecule exposed to the solvent (Figure 1). This fact confers on Myp a high degree of freedom of movement, which has raised the mean value of the *rmsd*. However, the major region of the ligand Myp that remains in contact with the protein oscillates less than 0.7 Å. Thus, the protocol was considered validated for use in virtual screening simulations, which was performed using two different programs, since we believe that if the same compound is selected by both programs using different search and ranking algorithms, the chances of selecting a false-positive compound decrease. In this way, the protocol for both programs was validated by redocking of the modeled compound before it was applied to the libraries. The libraries were divided into folders with 2500 molecules each, plus the structure of the reference inhibitor (Myp), and only compounds ranked higher than Myp were selected.

The first library was built from a search of the literature for compounds with cathepsin inhibitory activity, and for each compound found, a new search in the Zinc database for molecules with 80% similarity to it was conducted, totaling 10,159 molecules. None of the molecules in this library were better ranked than the reference ligand Myp. The
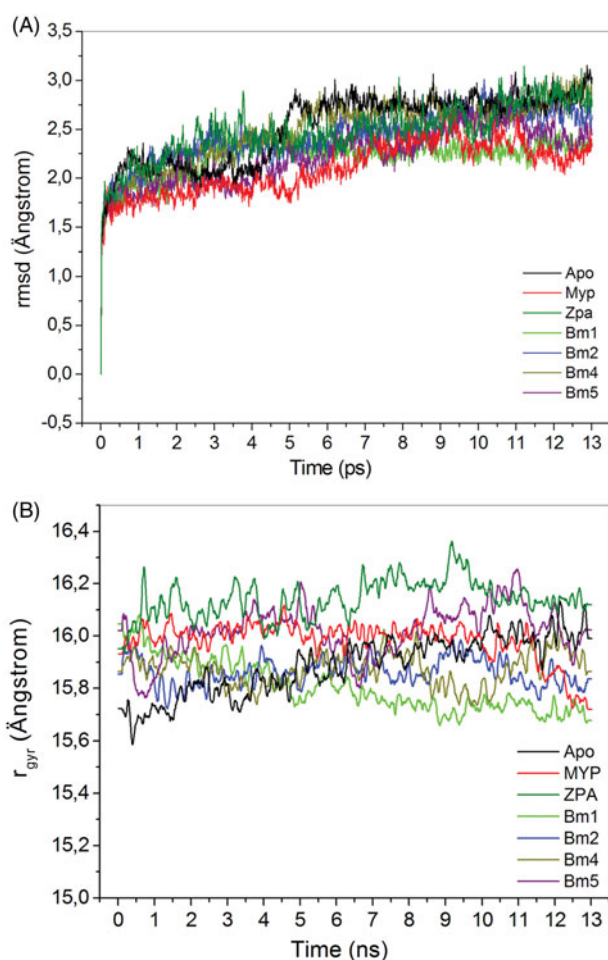
Figure 3. Analysis of the trajectory of the molecular dynamics from the atoms of the main chain in terms of *rmsd* (A) and radius of gyration (B). The colors represent the structure of *Bm*NPV-*Cath* bonded to each ligand. Figures generated by the program SciGraphica v-2.1.0 (Muenchen, Germany).

second library consisted of natural products with purchasable status from the Zinc database, totaling 111,500 molecules. From this library, 137 compounds were initially selected by the Vina program and, after checking the reproducibility of these results with four other screenings, we found 50 compounds better ranked than the reference inhibitor Myp. These 50 compounds were then evaluated by the Molegro program through four other simulations, and finally, only 12 compounds showed 100% reproducibility in all simulations using both programs. All compounds were successfully filtered by applying the ADMETox criteria by AdmetSar server. However, the ADMETox prediction methods are based on mammalian cells and tissues models. Therefore, the application of these criteria to insects may be questionable. Four molecules were selected as a representative sample of the set to be purchased for animal trials. These compounds were Bm1 (Zinc 70705449), Bm2 (Zinc70701476), Bm4 (Zinc 02496946), and Bm5 (Zinc12888007). The structure of each compound selected is shown in Figure 2.

MD evaluation

The best ranked pose for each ligand found by virtual screening was incorporated into the *Bm*NPV-*Cath* structure

and submitted to MD simulations using the Gromacs program. Although the structure had already been minimized by the NAMD2 program, the coordinates of the ligands had not, so a steepest descent minimization was necessary to start MD. As a result, the system quickly converged to the lower energy state in this procedure. The conjugate gradient (CG) method works better and faster to find the structure at its real minimum compared to the steepest descent (Sarkar, 1991) and therefore, we believe that CG application would be the most adequate to obtain the conformation to be used in the docking simulations.

The analysis of the trajectory by means of the *rmsd* shows that all the complexes reached thermodynamic equilibrium after 5 ns of simulation (Figure 3(A)), indicating the time of 13 ns was sufficient, which allowed them to be used in the subsequent calculation of free energy of binding. All the simulated complexes were stable regarding their folding status because their radius of gyration remained stable around 15.8 Å over time (Figure 3(B)). This may reflect the three disulfide bonds present in the molecule. Although the observed oscillation around 0.2 Å is not significant, it is clear that there was a tendency to increase the value of the radius of gyration in the Apo form. Comparatively, this may indicate that the ligands further stabilized the protein structure.

When analyzing the fluctuation of the α-carbons of each residue in the last 2 ns of simulation, which comprises the thermodynamic equilibrium region in all the complexes, it is possible to infer which regions of the molecule presented the greatest flexibility. This analysis can be performed in terms of the *rmsf* and is shown in Figure S3 (Supplementary Material). The highest fluctuation did not exceed 2.1 Å, which reinforces the observation that the structures of the complexes are quite stable. However, the presence of regions with greater fluctuation in all complexes, such as the amino-terminal region (residues 1–5) and the regions of the loops 37–50, 161–174, and 175–191 are evident. The other fluctuations observed for each complex can be attributed to the different contacts that the protein makes with the structure of the different ligands. The information regarding the contact frequency of the protein residues with the evaluated ligands, within a range of 4.0 Å is shown in Table S5 (Supplementary material). Since each ligand interacts with the protein differently it justifies distinct affinities. However, residues Gly22, Ala23, Cys24, Trp25, Cys62, Asn63, Gly64, Gly65, Leu66, Leu155, and Asn156 are the most important since they have a contact frequency greater than 80% with most ligands.

The different levels of contacts that ligands make with the protein, as suggested in Table S5 (Supplementary material), are reflected directly in their binding energy, which can be estimated by calculations of molecular mechanics. The mm-pbsa method was used in these calculations and provides an accurate assessment of the interaction energy of different compounds with the same protein. Once the same simulation conditions are maintained, the values obtained are comparable to each other and can be used to estimate which compound binds with the highest affinity to the protein. Table 1 shows the values for each component used in the calculation of the $\Delta G_{binding}$ (Equation (1)).

Table 1. Parameters obtained by mm-pbsa calculation to estimate the $\Delta G_{binding}$ of each ligand used in this work.

Energy	Energy (kJ mol^{-1})					
	Myp	ZPA	Bm1	Bm2	Bm4	Bm5
van der Waals	−298.6 (17.83)	−309.5 (26.70)	−294.9 (17.98)	−323.5 (28.26)	−257.0 (18.56)	−298.0 (19.41)
Electrostatic	−485.8 (46.02)	−459.7 (23.86)	−46.1 (20.00)	−40.1 (13.73)	−23.3 (11.14)	−62.3 (11.06)
Polar solvation	304.9 (61.37)	340.8 (31.37)	209.1 (25.40)	131.3 (16.25)	95.9 (13.44)	138.0 (11.47)
SASA	−23.2 (1.47)	−25.9 (1.89)	−22.2 (1.71)	−23.1 (1.59)	−17.9 (1.31)	−21.2 (1.09)
$\Delta G_{binding}$	−502.7 (25.10)	−454.2 (23.31)	−154.1 (20.38)	−255.4 (28.15)	−202.2 (19.16)	−243.5 (19.45)
Vina score	−36.8 (0.25)	−36.4 (0.02)	−37.2 (0.02)	−39.1 (0.21)	−41.0 (0.01)	−39.8 (0.01)

The number in parentheses refers to the standard deviation.

According to the Vina program, the five selected ligands have a higher docking score for *Bm*NPV-*Cath* than the reference ligand Myp (Table 1). However, the calculations from the MD show that Myp would bind with almost twice the affinity of the best ranked inhibitor (Bm2). The answer to this contradiction may be in the way these values were calculated. The scoring function of Vina ignores the electrostatic component (Trott & Olson, 2009), whereas the mm-pbsa program considers it, and this component contributed the most to the lowest value of Myp $\Delta G_{binding}$ (Table 1).

The literature reports that the K_M for human cathepsin-F with the Zpa ligand ranges from 0.44 to 17.0 μM (Ho et al., 2002; Wang et al., 1998), whereas, for human V-cathepsin, the K_M for this ligand was 4.8 μM (Yasuda et al., 2004). Because it is a substrate recognized by several cathepsins such as F, O, and V, we believe that it could also be used as a substrate of *Bm*NPV-*Cath*; therefore, it was used in calculations for comparative purposes. No reports citing the binding constants or other kinetic parameters for Myp and proteases were found in the literature; therefore, a comparison between calculated and experimental values was not possible for Myp.

When we substituted the $\Delta G_{binding}$ value displayed in Table 1 into Equation (2), we found a value of K that differed vastly from those measured experimentally

$$\Delta G_{binding} = -RT \ln K \qquad (2)$$

$$\Delta G = \Delta H - T \Delta S \qquad (3)$$

These differences occur for several reasons (Kumari et al., 2014); however, despite the differences between the kinetically measured value and the one calculated by the simulation, the values of $\Delta G_{binding}$ calculated using the mm-pbsa method are comparable (Kumari et al., 2014). In this way, it is possible to predict that in a comparison of the measured and calculated values, the measured value should vary by the same proportion as the calculated one. Thus, the computational simulations carried out in this work showed that the Myp inhibitor would have a slightly higher affinity than the synthetic substrate Zpa, whereas the inhibitor candidate compounds would have a lower affinity than the synthetic substrate.

The structures of the inhibitor compounds (Figure 2) contain alternating NH, C, and C = O, mimicking the main chain of a polypeptide, which is in direct contact with the catalytic site of *Bm*NPV-*Cath*. The aromatic radicals of the compounds, including that from the reference ligand Myp, penetrate the protein pocket, providing additional stability by means of van der Waals contacts.

Most *B. mori* infections occur orally, including *Bm*NPV virus. Treatment options using antivirals should therefore also be administered orally; however, caterpillars tend to reject mulberry leaves when odors or contaminants are present (Porto, Okamoto, Ikuno, Ferreira, & Margatho, 2005), which may include antivirals. This factor may limit treatment options; however, in *in vivo* experiments, compounds deposited on mulberry leaves were ingested without rejection by the caterpillars.

Biological activity of the molecules

The molecules Bm1, Bm2, Bm4, and Bm5 were used in the initial evaluations. In this first trial, all compounds were dissolved in 10% DMSO, since it is the first choice of solvent for solubilization of organic molecules (Ellson et al., 2005). The caterpillars were inoculated with the virus on the first day of the fifth instar and received the treatment on the fourth day after infection, when the symptoms of viral infection were visible (swelling of the intersegmental membranes and diarrhea). The results of these assays are shown in Table 2.

In the treatments with Bm1 and Bm4, only 20 caterpillars were used due to the availability of the compound. Equal letters in the same column represent no difference between the means according to ANOVA ($p > .05$).

The results showed that treatment with Bm1 was more deadly for caterpillars than the virus itself. Treatments with Bm2, Bm4, and Bm5 reduced the mortality of caterpillars by an average of 15%. The mortality rate of the pupae is calculated as a function of the number of cocoons generated by the surviving caterpillars. However, they can die inside the cocoon due to viral infection and therefore do not generate viable matrices. The Bm5 group presented a mortality rate of caterpillars similar to that observed with Bm2 and Bm4; however, no pupae died with this treatment. Calculation of binding free energy by the mm-pbsa method also suggested that Bm5 could have a high affinity for *Bm*NPV-*Cath* among inhibitors (Table 1). These results indicate that Bm5 would be the most promising candidate. The DMSO proved to be toxic to caterpillars (DMSO control), since administration of 10 μL of DMSO 10% solution caused a mortality rate of 10%, similar to those observed in treatments with inhibitors ($p > .05$).

Due to these results, a further trial was performed on a larger number of caterpillars using only the Bm5 molecule dissolved in micelles of Pluronic-F127, another agent used for drug solubilization. It is a hydrophilic non-ionic surfactant

Table 2. Mortality of caterpillars and pupae as a function of treatments.

Treatment	N	Dead caterpillars (n)	Dead caterpillars (%)	Cocoon (n)	Dead pupae (n)	Dead pupae (%)
Negative Ctrl 1	30	0	0.00	30	1	3.33[a]
DMSO Control	20	2	10.00[a]	18	1	5.56[a]
Infected Ctrl 1	30	8	26.67[c]	22	3	13.63[b]
Treatment Bm1	20	8	40.00[b]	12	1	8.33[a]
Treatment Bm2	30	4	13.33[a]	26	3	11.54[b]
Treatment Bm4	20	3	15.00[a]	17	1	5.88[a]
Treatment Bm5	30	5	16.67[a]	25	0	0.00

For these trials, the caterpillars were infected with 10 μL of a BmNPV suspension at a concentration of 3×10^7 COPS/mL.

Table 3. Mortality of caterpillars and pupae as a function of treatments.

Treatment	N	Dead caterpillars (n)	Dead caterpillars (%)	Cocoons (n)	Dead pupae (n)	Dead pupae (%)
Negative Ctrl 2	80	1	1.25[a]	77	0	0.00
Pluronic Ctrl	80	1	1.25[a]	76	0	0.00
Infected Control 2	80	17	21.25[b]	63	6	9.52[a]
Treatment Pluronic	80	16	20.00[b]	64	2	3.12[b]
Treatment Bm5	79	13	16.45[c]	63	3	4.48[b]

For these assays, the caterpillars were infected with 10 μL of a BmNPV suspension at a concentration of 8×10^8 COPS/mL.

capable of forming micelles that aid in the solubilization and absorption of organic compounds (Sezgin, Yuksel, & Baykara, 2006). In this assay, the concentration of inoculated virus was increased to 8×10^8 COPS/mL so that the same DL_{50} of the previous trial was obtained, since some polyhedra are inactivated over time.

The results of this second trial (Table 3) showed that the infected group treated with Pluronic-F127 at 0.02% (Pluronic treatment) had a similar mortality rate to the infected and untreated group (infected control 2) ($p > .05$). The group treated with Pluronic-F127 at 0.02% (Pluronic control) had a mortality rate similar to that of the water-treated group (negative control 2) ($p > .05$). These results indicate that Pluronic-F127, when used as a vehicle for solubilizing Bm5, did not interfere with the treatment, was not toxic, and did not have a curative effect on the caterpillars. On the other hand, the group of caterpillars infected and treated with Bm5 presented a 22.6% reduction in mortality rate, which is quite significant ($p < .07$).

Equal letters in the same column represent no difference between the means according to ANOVA ($p > .07$).

Although this result is quite promising, we must consider that, in both trials, the caterpillars were treated on the fourth day after infection, at a time when they already had the signs of the viral infection. The appearance of these symptoms is the first sign that allows the detection of the infection in silkworm breeding grounds. It is possible that in this period, the infection is so widespread that treatment is less effective. Therefore, it is possible that by treating the caterpillar soon after infection and/or by adding more doses of Bm5 along the fifth instar, the mortality rate could be reduced further.

To date, no substance has been found that, alone, could prevent silkworm infection by BmNPV. As a result, the current method used to combat the virus occurs through management techniques such as disinfection of the environment with chemicals such as 30% formaldehyde and the application of hydrated lime on the floor before starting a new breeding (Okamoto & Rodella, 2004). Despite this, there are

reports that show resistance of pathogens to these agents (Brancalhão et al., 2013). Thus, methods that prevent liquefaction of the dead body of the caterpillar can contribute to management techniques, since they facilitate its removal from the place, preventing the budded viruses from spreading through the environment.

The dead caterpillars and pupae showed treatment-independent body liquefaction. This liquefying effect is not only caused by BmNPV-Cath, but also by viral chitinases (Hawtin et al., 1995). Thus, we attributed this result to the action of viral chitinases and remaining BmNPV-Cath that were not inhibited by Bm5.

Although the initial hypothesis of our project was that inhibition of BmNPV-Cath could prevent liquefaction of the caterpillar's body, this was not observed in the trials. However, unexpectedly, the results showed a significant reduction in the mortality rate of treated caterpillars. Although the literature reports that V-cath is not essential for virus replication (Hawtin et al., 1995; Ohkawa et al., 1994), the results presented here indicate that inhibition of BmNPV-Cath was important for caterpillar survival. In this way, our results also suggest that BmNPV-Cath may play a role in the viral cycle that goes beyond just liquefying the caterpillar's body and causing the spread of polyhedra. Because it is a broad-spectrum protease, BmNPV-Cath, as well as other viral proteases, may play a role in the replication and/or generation of a pool of amino acids that would be available for the synthesis of viral proteins (Brecher, Zhang, & Li, 2013; Tong, 2002).

Regarding future technologies, the use of BmNPV-Cath inhibitors in combination with other compounds that may interfere with viral replication may prove to be a useful strategy in the development of therapies such as antiviral cocktails aiming to control this plague of sericulture.

Conclusions

Based on techniques of biochemistry and computational biophysics, the structure of *Bombyx mori* Nucleopedrovirus

cathepsin (*Bm*NPV-*Cath*) was modeled and used in the identification of the molecule Zinc12888007 (Bm5) by structure-based virtual screening techniques. The use of the Bm5 molecule dissolved in Pluronic-F127 micelles for *in vivo* infection trials of silkworm caterpillars with *Bm*NPV promoted a significant reduction in mortality of fifth instar caterpillars after the onset of viral infection symptoms. The biological role of *Bm*NPV-*Cath* can go beyond just liquefying the body of dead caterpillars for the propagation of the virus; it may also play a role in generating a pool of amino acids to be used in the synthesis of viral proteins. The results of this research have the potential for direct biotechnological application in the fight against one of the most common plagues in sericulture.

Disclosure statement

Funding

ORCID

Flávio Augusto Vicente Seixas (iD) http://orcid.org/0000-0002-0117-6919

References

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), 10037–10041. doi:10.1073/pnas.181342398

Brancalão, R. M. C., & Ribeiro, L. F. C. (2003). Cytopathology of BmNPV infection in the tegument of *Bombyx mori* L., 1758 (Lepidoptera: Bombycidae). *Arquivos de Ciências Veterinárias e Zoologia da UNIPAR*, 6, 15–20. Retrieved from http://revistas.unipar.br/index.php/veterinaria/article/view/788

Brancalhão, R. M. C., Ribeiro, L. F. C., Torquato, E. F. B., Fernandez, M. A., Munhoz, R. E. F., Colombelli, K. T., & da Costa, R. R. G. (2013). Virucidal evaluation of sodium dichloroisocyanurate formulation against nucleopolyhedroviruses of silkworm. *Arquivos do Instituto Biológico*, 80(2), 233–235. doi:10.1590/S1808-16572013000200014

Brecher, M., Zhang, J., & Li, H. (2013). The flavivirus protease as a target for drug discovery. *Virologica Sinica*, 28(6), 326–336. doi:10.1007/s12250-013-3390-x

CCP4. (1994). The CCP4 Suite: Programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography*, 50, 760–763. doi:10.1107/S0907444994003112

Cheng, F. X., Li, W. H., Zhou, Y. D., Shen, J., Wu, Z. R., Liu, G. X., … Tang, Y. (2012). admetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *Journal of Chemical Information and Modeling*, 52(11), 3099–3105. doi:10.1021/Ci300367a

Cheng, J., Saigo, H., & Baldi, P. (2005). Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. *Proteins: Structure, Function, and Bioinformatics*, 62(3), 617–629. doi:10.1002/prot.20787

Ellson, R., Stearns, R., Mutz, M., Brown, C., Browning, B., Harris, D., … Wold, D. (2005). In situ DMSO hydration measurements of HTS compound libraries. *Combinatorial Chemistry & High Throughput Screening*, 8(6), 489–498. doi:10.2174/1386207054867382

Engelhard, E. K., Kam-Morgan, L. N., Washburn, J. O., & Volkman, L. E. (1994). The insect tracheal system – A conduit for the systemic spread of Autographa-California-M Nuclear Polyhedrosis-Virus. *Proceedings of the National Academy of Sciences*, 91(8), 3224–3227. doi:10.1073/pnas.91.8.3224

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., … Sali, A. (2006). Comparative protein structure modeling using modeller. *Curr Protoc Bioinformatics*, Chapter5, Unit 5 6. doi:10.1002/0471250953.bi0506s15

Funk, C. J., & Consigli, R. A. (1992). Evidence for zinc-binding by 2 structural proteins of plodia-interpunctella granulosis-virus. *Journal of Virology*, 66, 3168–3171. Retrieved from http://jvi.asm.org/content/66/5/3168.short

Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., … Diederich, F. (2011). Systematic investigation of halogen bonding in protein–ligand interactions. *Angewandte Chemie International Edition*, 50(1), 314–318. doi:10.1002/anie.201006781

Hawtin, R. E., Arnold, K., Ayres, M. D., Zanotto, P. M. D., Howard, S. C., Gooday, G. W., … Possee, R. D. (1995). Identification and preliminary characterization of a chitinase gene in the Autographa-Californica Nuclear Polyhedrosis-Virus Genome. *Virology*, 212(2), 673–685. doi:10.1006/viro.1995.1525

Herniou, E. A., Olszewski, J. A., Cory, J. S., & O'Reilly, D. R. (2003). The genome sequence and evolution of baculoviruses. *Annual Review of Entomology*, 48(1), 211–234. doi:10.1146/annurev.ento.48.091801.112756

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3), 435–447. doi:10.1021/Ct700301q

Ho, J. D., Meltser, Y., Buggy, J. J., Palmer, J. T., Elrod, K. C., Chan, H., … Somoza, J. R. (2002). Expression, purification, crystallization and preliminary X-ray diffraction studies of human cathepsin F complexed with an irreversible vinyl sulfone inhibitor. *Acta Crystallographica Section D Biological Crystallography*, 58(12), 2187–2190. doi:10.1107/S0907444902017535

Hom, L. G., & Volkman, L. E. (2000). Autographa californica M nucleopolyhedrovirus chiA is required for processing of V-CATH. *Virology*, 277(1), 178–183. doi:10.1006/viro.2000.0586

Inserco.org. (2018, June 01). *International Sericultural Comission: Statistics* [online]. Retrieved from http://inserco.org/en/statistics

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, 52(7), 1757–1768. doi:10.1021/ci3001277

Kumari, R., Kumar, R., Lynn, A., & Consort, O. S. D. D. (2014). g_mmpbsa- A GROMACS tool for high-throughput mm-pbsa calculations. *Journal of Chemical Information and Modeling*, 54(7), 1951–1962. doi:10.1021/Ci500020m

Lobley, A., Sadowski, M. I., & Jones, D. T. (2009). pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics*, 25(14), 1761–1767. doi:10.1093/bioinformatics/btp302

Mackerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 25(11), 1400–1415. doi:10.1002/jcc.20065

McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. M. (2011). Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallographica Section D Biological Crystallography*, 67(4), 386–394. doi:10.1107/S0907444911007281

Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harbor Protocols*, *2007*(7), pdb.top17. doi:10.1101/pdb.top17

Mulder, N., & Apweiler, R. (2007). InterPro and InterProScan: Tools for protein sequence classification and comparison. *Methods in Molecular Biology*, *396*, 59–70. doi:1-59745-515-6:59

Murphy, F. A., Faukquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., … Summers, M. D. (1995). *Virus taxonomy: Classification and nomenclature of viruses: International Committee on Taxonomy of Viruses* (6th ed.) New York, NY: Springer-Verlag Wien.

Ohkawa, T., Majima, K., & Maeda, S. (1994). A cysteine protease encoded by the baculovirus *Bombyx-mori Nuclear Polyhedrosis-Virus. Journal of Virology*, *68*, 6619–6625. Retrieved from http://jvi.asm.org/content/68/10/6619.short

Okamoto, F., & Rodella, R. A. (2004). Silkworm (*Bombyx mori* L.) Cocoon production related to morphological and bromatological characteristics of mulberry leaves (Morus SPP.). *Boletim de Indústria Animal*, *61*, 91–99. Retrieved from http://www.iz.sp.gov.br/pdfsbia/1180467536.pdf

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. doi: 10.1002/jcc.20289

Porto, A. J., Okamoto, F., Ikuno, A. A., Ferreira, V. C. A., & Margatho, L. F. F. (2005). Biological and productivity evaluation of silkworm (*Bombyx mori* L.) feed with mulberry leaves treated with *Mirabilis jalapa* extracts. *Arquivos do Instituto Biológico*, *72*, 445–453. Retrieved from http://www.biologico.sp.gov.br/publicacoes/revista/arquivos-do-instituto-biologico/todos

Rankovic, Z., Cai, J., Kerr, J., Fradera, X., Robinson, J., Mistry, A., … Nicolai, E. (2010). Design and optimization of a series of novel 2-cyano-pyrimidines as cathepsin K inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *20*(5), 1524–1527. doi:10.1016/j.bmcl.2010.01.100

Ribeiro, L. F., Brancalhao, R. M., Torquato, E. F., & Fernandez, M. A. (2009). Susceptibility of the *Bombyx mori* cardia cells to Nucleopolyhedrovirus, multiple subgroup, BmMNPV. *Journal of Invertebrate Pathology*, *100*(3), 195–198. doi:10.1016/j.jip.2009.01.001

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*, *12*(2), 85–94. doi:10.1093/protein/12.2.85

Sarkar, T. K. (1991). *Application of conjugate gradient method to electromagnetics and signal analysis* (Vol. 5). New York, NY: Elsevier Science Limited.

Schuttelkopf, A. W., & van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallographica Section D Biological Crystallography*, *60*(8), 1355–1363. doi:10.1107/S0907444904011679

Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., … van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A*, *103*(19), 3596–3607. doi:10.1021/Jp984217f

Sezgin, Z., Yuksel, N., & Baykara, T. (2006). Preparation and characterization of polymeric micelles for solubilization of poorly soluble anti-cancer drugs. *European Journal of Pharmaceutics and Biopharmaceutics*, *64*(3), 261–268. doi:10.1016/j.ejpb.2006.06.003

Shobahah, J., Xue, S., Hu, D., Zhao, C., Wei, M., Quan, Y., & Yu, W. (2017). Quantitative phosphoproteome on the silkworm (*Bombyx mori*) cells infected with baculovirus. *Virology Journal*, *14*(1), 117. doi:10.1186/s12985-017-0783-8

Slack, J. M., Kuzio, J., & Faulkner, P. (1995). Characterization of v-cath, a cathepsin l-like proteinase expressed by the baculovirus *Autographa-californica* Multiple Nuclear Polyhedrosis-Virus. *Journal of General Virology*, *76*(5), 1091–1098. doi:10.1099/0022-1317-76-5-1091

Somoza, J. R., Palmer, J. T., & Ho, J. D. (2002). The crystal structure of human cathepsin F and its implications for the development of novel immunomodulators. *Journal of Molecular Biology*, *322*(3), 559–568. doi:10.1016/S0022-2836(02)00780-5

Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321. doi:10.1021/jm051197e

Tong, L. (2002). Viral proteases. *Chemical Reviews*, *102*(12), 4609–4626. doi:10.1021/cr010184f

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. doi:10.1002/jcc.21334

Volkman, L. E., & Keddie, B. A. (1990). Nuclear polyhedrosis virus pathogenesis. *Seminars in Virology*, *1*, 249–256. Retrieved from https://www.sciencedirect.com/journal/seminars-in-virology/issues

Wang, B., Shi, G. P., Yao, P. M., Li, Z., Chapman, H. A., & Bromme, D. (1998). Human cathepsin F. Molecular cloning, functional expression, tissue localization, and enzymatic characterization. *Journal of Biological Chemistry*, *273*(48), 32000–32008. doi:10.1074/jbc.273.48.32000

Wolf, L. K. (2009). New software and websites for the chemical enterprise. *Chemical & Engineering News*, *87*(5), 32. doi:10.1021/cen-v087n005.p032

Yang, A. S., & Honig, B. (2000). An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *Journal of Molecular Biology*, *301*(3), 665–678. doi:10.1006/jmbi.2000.3973

Yasuda, Y., Li, Z., Greenbaum, D., Bogyo, M., Weber, E., & Bromme, D. (2004). Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages. *Journal of Biological Chemistry*, *279*(35), 36761–36770. doi:10.1074/jbc.M403986200

Zacharias, M. (2003). Protein–protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Science*, *12*(6), 1271–1282. doi:10.1110/ps.0239303

Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*, *32*(11), 2359–2368. doi:10.1002/jcc.21816

Supplementary Materials

The structure of viral cathepsin from *Bombyx mori Nuclear Polyhedrosis Virus* as a target against grasserie: Docking and molecular dynamics simulations

Paulo Sérgio Alves Bueno[1], Débora Carina Biavatti[1], Alex Sandro Duarte Chiarello[2], Verônica Aureliana Fassina[2], Maria Aparecida Fernandez[2], Flávio Augusto Vicente Seixas[1#]

[1]Department of Technology, Universidade Estadual de Maringá, campus Umuarama, Av. Ângelo Moreira da Fonseca, 1800, 87506-370, Umuarama, Paraná, Brazil

[2]Department of Biotechnology, Genetics and Cell Biology, Universidade Estadual de Maringá, Av. Colombo, 5790, Maringá, Paraná, Brazil

Corresponding author:

favseixas@uem.br

Laboratório de Bioquímica Estrutural

Departamento de Tecnologia, Universidade Estadual de Maringá, Campus Umuarama

Av. Ângelo Moreira da Fonseca, 1800, 87506-370, Umuarama, Paraná, Brasil.
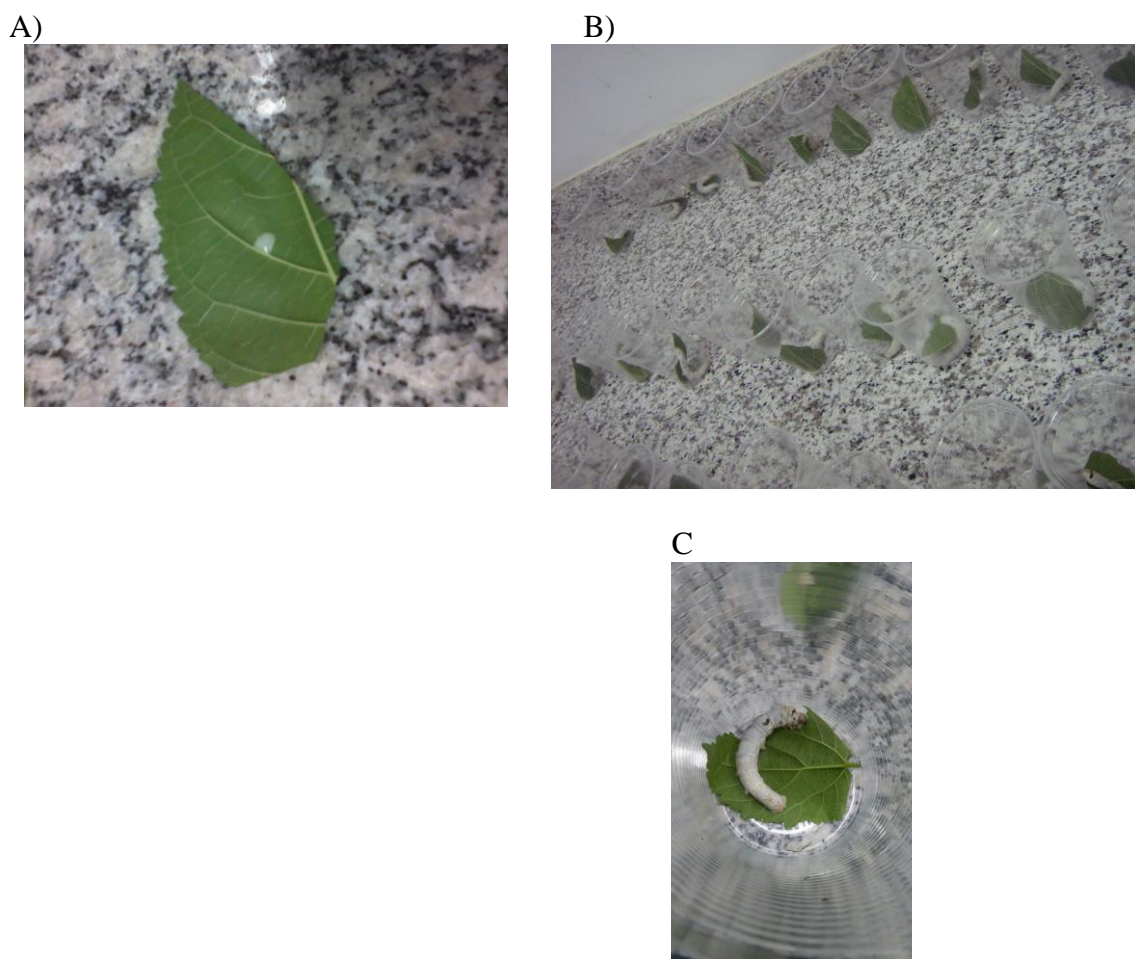
Fone: +55 44 36219337

Figure S1. (A) Substances used in the treatments applied on mulberry leaves. It was expected for the liquid to dry before the treated leaf was used to feed the caterpillars. (B) Fifth instar *Bombyx mori* caterpillar imprisoned in a disposable cup until the ingestion of the treated leaf. (C) Zoon view of (B).
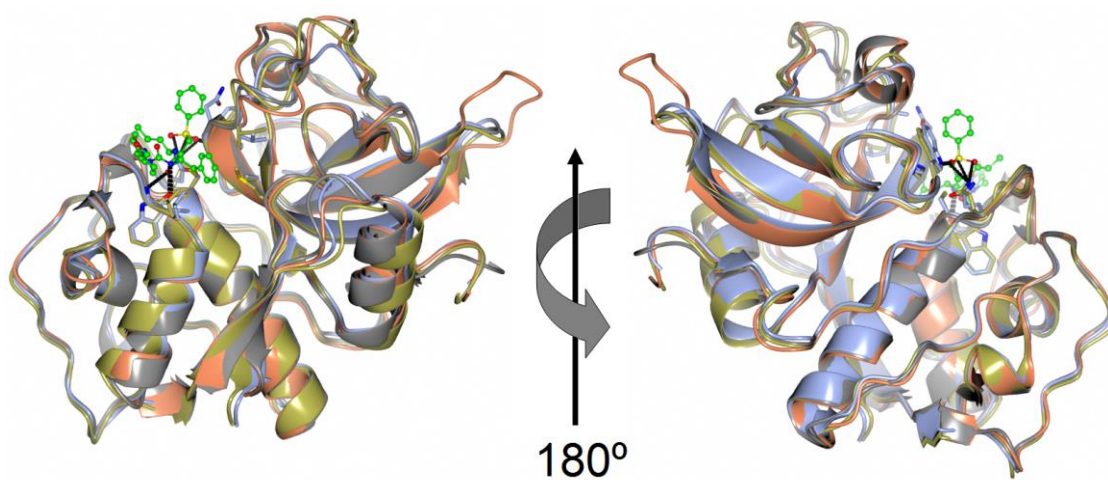

Figure S2. Ribbon model of the template structures 1m6d (olive), 2xu3 (orange), 3kwz (gray), superimposed on BmNPV-Cath (blue). The Myp ligand is in green (ball & stick).
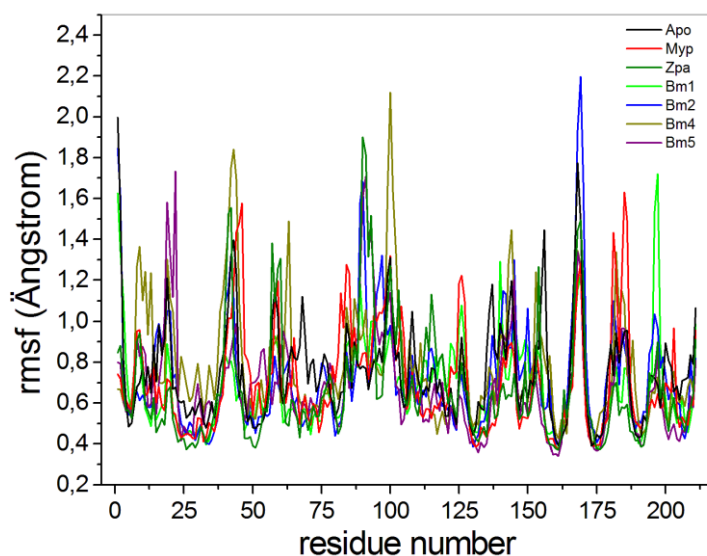
Figure S3. Root mean square fluctuation of the alpha carbons at the last two nano-seconds of simulation. Each color indicates the *Bm*NPV-*Cath* in complex with the simulated ligands. The higher the rmsf value, the greater the oscillation of the residue in the structure.

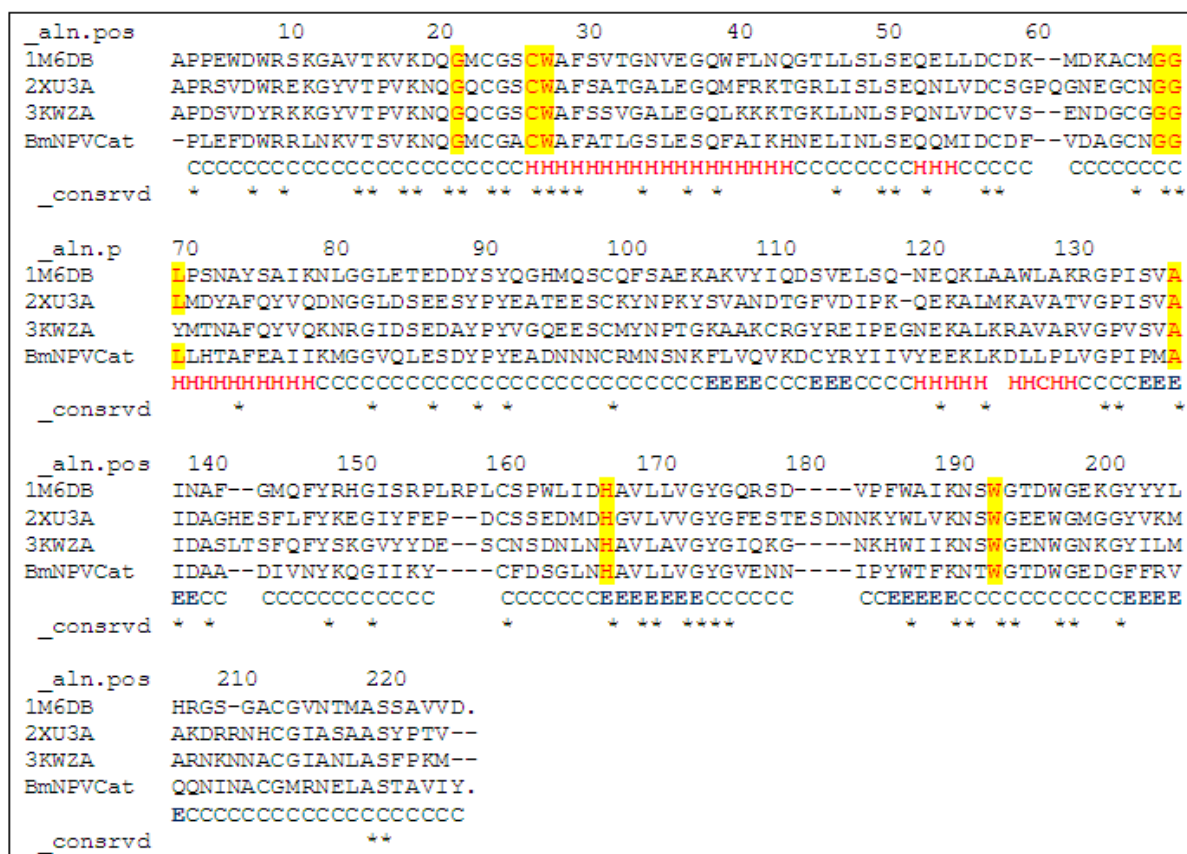**Figure S4.** Clustaw alignment of the amino acid sequences from *Bm*NPV-*Cath* and from templates used in this work. Asterisks indicate identical residues; Residues composing the catalytic site in the BmNPV-Cath and in the templates are highlighted in yellow; The regions corresponding to the alpha-helix structural motifs are in red (H) and the structural motifs of beta-strands are in blue (E).

Table S5. Contact frequency between BmNPV-Cath residues and the ligands identified by virtual screening within 4.0 Å distance at last 2 ns of molecular dynamics. Only contacts with frequency greater than 5% are shown. Bold residues are those with a contact frequency greater than 80% with at least 3 ligands.

Residue	BM1	BM2	BM3	BM4	BM5
Asn 17	17,73				
Gln 18	66,50	99,01	100,00	62,07	0,49
Gly 19	1,48	93,10	73,40	63,05	
Met 20	12,32	87,68	94,09	87,19	
Cys 21		29,56	41,87	90,64	
Gly 22	84,24	100,00	97,54	100,00	46,80
Ala 23		87,19	99,51		99,01
Cys 24	45,81	100,00	100,00	80,30	100,00
Trp 25	98,03	98,52	100,00	91,63	100,00
Ile 53				2,46	
Asp 59			0,49		12,81
Ala 60				3,94	41,38
Gly 61		75,37	93,60	77,34	
Cys 62		98,03	94,09	95,57	
Asn 63	86,21	96,55	88,67	43,35	99,51
Gly 64	47,78	100,00	100,00	98,03	100,00
Gly 65	71,92	100,00	100,00	99,51	100,00
Leu 66	97,04	93,60	91,63	96,06	98,52
Leu 67	51,72	95,57	67,98	47,29	92,12
His 68					1,97
Ala 90				25,12	
Asp 91				2,46	
Arg 111		0,49			
Met 132		34,98			
Ala 133	55,17	76,85	50,25	52,22	18,72
Ala 136			33,50		0,49
Asp 138	94,58				
Ile 139	99,51				87,68
Val 140	80,79	7,39	55,17		44,33
Asn 141	20,20				
Gly 154	0,49		39,41	10,34	
Leu 155	99,51	100,00	100,00	96,55	100,00
Asn 156	100,00	100,00	100,00	92,12	100,00
His 157	78,33	27,09	99,01	88,18	80,30
Ala 158	91,13	82,76	58,62	41,38	56,16
Thr 178		8,37			
Trp 179	100,00	32,51	89,66	0,49	99,51
Trp 183	84,24				2,96
Met 200	3,45				48,28
Leu 204	0,49	52,71			66,50

CONCLUSÃO FINAL

Nesta tese foram descritos interessantes resultados utilizando técnicas de simulação computacional. A partir de diferentes alvos celulares, conseguimos selecionar, especificamente, candidatos a inibidores com sucesso, os quais possuem aplicação biotecnológica direta. Um destes candidatos, o HS9, foi patenteado como formulação farmacêutica contra infecções por fungos do gênero Paracoccidioides, para uso tópico e sistêmico em medicina humana e veterinária.

Os estudos desta tese também demonstraram que os resultados das simulações conseguem apenas apontar para uma molécula candidata à ligante com maior probabilidade. A comprovação de que este candidato pode, de fato, se ligar ao alvo selecionado ainda depende de comprovação experimental. Porém, ao selecionar um número reduzido de candidatos dentro de uma biblioteca com centenas de milhares de opções, reduz-se em muito o número de ensaios de bancada necessários para se observar o efeito esperado. Isso demonstra um caráter sustentável das metodologias *in silico*, pois reduzem os custos com material químico, energia e tempo de experimentos.

O uso da simulação computacional para o entendimento de sistemas biológicos, tem aumentado nos últimos anos em função do aumento da capacidade de processamento dos computadores e do número de informações disponíveis nas bases de dados online. A cada trabalho desenvolvido nesta tese, novas metodologias *in silico* foram aplicadas decorrentes da evolução constante deste campo de pesquisa.