

Paulo Sérgio Alves Bueno

MODELAGEM POR HOMOLOGIA DA ENZIMA CORISMATO SINTASE DE *Candida albicans* COMPLEXADA COM LIGANTES: SIMULAÇÕES DE ANCORAGEM E DINÂMICA MOLECULAR.

Orientador: Prof. Dr. Flavio Augusto Vicente Seixas

Paulo Sérgio Alves Bueno

MODELAGEM POR HOMOLOGIA DA ENZIMA CORISMATO SINTASE DE *Candida albicans* COMPLEXADA COM LIGANTES: SIMULAÇÕES DE ANCORAGEM E DINÂMICA MOLECULAR.

Dissertação 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 Mestre em Ciências Biológicas.

Maringá

Fevereiro – 2017

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

B928m Bueno, Paulo Sérgio Alves, 1990-
Modelagem por homologia da enzima corismato sintase de *Candida albicans* complexada com ligantes : simulações de ancoragem e dinâmica molecular / Nome do autor. -- Maringá, 2017.
ix, 18 f. : figs., tabs.

Orientador: Prof. Dr. Flavio Augusto Vicente Seixas.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Biológicas, Programa de Pós-Graduação em Ciências Biológicas, 2017.

1. Corismato sintase (Enzima) - *Candida albicans*. 2. Modelagem molecular. 3. *Candida albicans*. 4. Ancoramento molecular. 5. Antifúngico. I. Seixas, Flavio Augusto Vicente, orient. II. Universidade Estadual de Maringá. Centro de Ciências Biológicas. Programa de Pós-Graduação em Ciências Biológicas. III. Título.

CDD 23.ed. 572.8


GVS-003716



Universidade Estadual de Maringá
 Centro de Ciências Biológicas
 Programa de Pós-Graduação em Ciências Biológicas



ATA DE DEFESA DA DISSERTAÇÃO DE MESTRADO DO ALUNO PAULO SÉRGIO ALVES BUENO. Aos dezessete dias do mês de fevereiro de dois mil e dezessete, sexta-feira, realizou-se no Bloco G-56, sala 201, no campus universitário, a sessão pública de defesa da dissertação intitulada: **“Modelagem por homologia da enzima corismato sintase de *Candida albicans* complexada com ligantes: simulações de ancoragem e dinâmica molecular”**, apresentada pelo aluno Paulo Sérgio Alves Bueno, Bacharelado em Bioquímica, pela Universidade Estadual de Maringá, que concluiu os créditos exigidos para obtenção do grau de “Mestre em Ciências Biológicas”. Os trabalhos foram instalados às 13:30, pelo Prof. Dr. Flavio Augusto Vicente Seixas, Presidente da Banca Examinadora, constituída pelos professores Dr. Eduardo Borges de Melo e Dra. Gisele de Freitas Gauze como membros. A Banca Examinadora, tendo se decidido a aceitar a dissertação, passou à arguição pública do candidato. Encerrados os trabalhos de arguição às 15:45 horas, os examinadores deram parecer final, considerando a dissertação aprovada. Proclamado o resultado pelo Presidente da Banca Examinadora, foram encerrados os trabalhos e lavrada a presente ata que vai assinada pelos membros da Banca Examinadora. Maringá, dezessete de fevereiro de dois mil e dezessete.


 Prof. Dr. Flavio Augusto Vicente Seixas


 Prof. Dr. Eduardo Borges de Melo


 Profa. Dra. Gisele de Freitas Gauze

ÍNDICE

BIOGRAFIA.....	iv
AGRADECIMENTOS.....	v
APRESENTAÇÃO.....	vi
RESUMO GERAL.....	vii
GENERAL ABSTRACT.....	viii
ARTIGO CIENTÍFICO.....	1
Resumo.....	2
Introdução.....	3
Materiais e métodos.....	4
<i>Obtenção da sequência e modelagem.....</i>	<i>4</i>
<i>Simulações de Dinâmica Molecular.....</i>	<i>4</i>
<i>Simulações de ancoragem.....</i>	<i>5</i>
<i>Varredura Virtual.....</i>	<i>5</i>
<i>Campos de força e cálculo das cargas parciais.....</i>	<i>6</i>
Resultados e Discussão.....	6
<i>Análise da sequência de CaCS.....</i>	<i>6</i>
<i>Modelagem molecular e validação da estrutura da CaCS.....</i>	<i>6</i>
<i>Triagem virtual.....</i>	<i>9</i>
<i>Simulações de Dinâmica Molecular.....</i>	<i>10</i>
<i>Sítio ativo e a ligação do 5-EPSP.....</i>	<i>11</i>
<i>Ligante CS8.....</i>	<i>12</i>
<i>Testes in vitro.....</i>	<i>13</i>
Conclusão.....	13
Agradecimentos.....	14
Referências.....	14
MATERIAL SUPLEMENTAR.....	17
ARTIGO PUBLICADO NO PERÍODO.....	18

BIOGRAFIA

Paulo Sérgio Alves Bueno nasceu em Jandaia do Sul, estado do Paraná em 06 de Junho de 1990. Iniciou a graduação no ano de 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 microorganismos, além de monitor das disciplinas Bioquímica de microorganismos e Bioquímica da sinalização, orientado pela professora Cristina Giatti Marques de Souza. Graduou-se em 2014 e no ano seguinte ingressou no mestrado do Programa de Pós Graduação em Ciências Biológicas (Biologia Celular e Molecular) da UEM. Realizou sua pesquisa sob orientação do professor Dr. Flavio Augusto Vicente Seixas no laboratório de Bioquímica Estrutural, na cidade de Umuarama, onde localiza-se o *campus* de Tecnologia da universidade. Seu trabalho, intitulado “Modelagem por homologia da enzima corismato sintase de *Candida albicans* complexada com ligantes: simulações de ancoragem e dinâmica molecular”, resultou na descoberta de um possível candidato contra candidíase, o que representa uma contribuição significativa para o combate contra doenças infecciosas causadas por fungos.

AGRADECIMENTOS

Agradeço a Deus, que sempre me deu forças para enfrentar os obstáculos da vida e não me deixou jamais desistir, abençoando sempre a minha caminhada.

À Universidade Estadual de Maringá (UEM), em especial ao Programa de Pós Graduação em Ciências Biológicas pela oportunidade de ingresso, e que, pela qualidade de seu corpo docente, e infra-estrutura, me proporcionou uma excelente formação enquanto mestrando em Biologia Celular e Molecular.

Ao Professor Dr. Flavio Augusto Vicente Seixas, pela total paciência, disponibilidade na orientação e ensinamentos que levarei para a vida toda.

Eu o agradeço muito!

A toda minha Família. À minha mãe, Maria, pelo seu amor, dedicação, esforço e sacrifício. Ao meu pai, Osvaldo, pelos conselhos, carinho e princípios repassados. Aos meus irmãos, Anderson, Tatiane, Fabiano e Daiane, por me dar apoio mesmo estando longe. À minha namorada Jaqueline por estar comigo, sempre apoiando minhas pesquisas.

Ao restante dos membros da minha família, que verdadeiramente querem meu bem.

A todos os colegas de laboratório, pela alegria e ajuda nos momentos difíceis.

A todos que de alguma forma contribuíram para a realização desse trabalho.

A todos, muito obrigado! Amo muito vocês!

APRESENTAÇÃO

Esta dissertação é composta de um artigo científico, que em consonância com as regras do Programa de Pós-graduação em Ciências Biológicas da Universidade Estadual de Maringá (UEM), foi redigido de acordo com as normas da revista de interesse.

O artigo intitulado “Modelagem por homologia da enzima corismato sintase de *Candida albicans* complexada com ligantes: simulações de ancoragem e dinâmica molecular” que será submetido à revista **Journal of Molecular Modeling** e inicia com uma revisão sobre a doença causada por *Candida albicans* e segue com a caracterização da enzima alvo do trabalho, a Corismato Sintase, juntamente com a justificativa de sua utilização em todos os processos computacionais realizados, até a identificação de um possível candidato contra candidíase, denominado CS8.

RESUMO GERAL

Infecções fúngicas matam milhares de pessoas todos os anos. O fungo comensal de humanos, *Candida albicans* (*Ca*) é responsável por quase metade dessas mortes. Existem poucos antifúngicos disponíveis para o tratamento e seu uso contínuo pode, inevitavelmente, levar ao desenvolvimento de mecanismos de resistência. Portanto, a descoberta de novas drogas, de preferência com novos alvos, é necessária. Nesse contexto, as enzimas da via do chiquimato, presentes em plantas, fungos e bactérias, representam excelentes alvos para o desenvolvimento de drogas, pois uma vez que estão ausentes em humanos, conferem à droga toxicidade seletiva. Este trabalho apresenta o primeiro estudo já realizado, tendo como alvo uma enzima da via do chiquimato no combate a esta doença. Desta forma, o objetivo deste trabalho foi o de modelar a estrutura da enzima Corismato Sintase (EC.: 4.2.3.5) de *Ca* de modo a utilizá-la em estudos de triagem virtual (utilizando ancoramento molecular) e dinâmica molecular, visando identificar novos compostos de ação específica contra candidíase. A estrutura da *CaCS* foi modelada na forma tetramérica e ligada ao substrato 5-EPSP e ao cofator FMNH₂, tendo as estruturas homólogas de *Saccharomyces cerevisiae* e *Streptococcus pneumoniae* como moldes. A identidade entre as sequências foi de 34-76%. A análise das interfaces de contato entre as subunidades sugeriu que a estrutura tetramérica seja a unidade biológica da proteína em solução. Para a realização da triagem virtual, a estrutura tetrámerica foi inicialmente minimizada por gradiente conjugado para remover possíveis colisões estereoquímicas entre a proteína e os ligantes. Em seguida, a subunidade A da proteína foi utilizada em simulações de *redocking* do 5-EPSP para validação do protocolo de ancoragem. O protocolo final foi então aplicado na triagem virtual que consistia de duas bibliotecas, uma formada por compostos similares aos inibidores já conhecidos da CS (2501 compostos) e outra formada por 120.435 produtos naturais com disponibilidade comercial, todos oriundos da base de dados Zinc. Os resultados identificaram o composto CS8 com um $\Delta G_{\text{ligação}}$ teórico de -11,84 Kcal.mol⁻¹, menor do que o obtido com o 5-EPSP que foi de -11,44 Kcal.mol⁻¹ indicando afinidade aumentada pelo inibidor. Esse composto não possui patente registrada nem aplicabilidade conhecida. Simulações de dinâmica molecular da *CaCS* durante 50 ns ligadas ao 5-EPSP ou ao CS8, confirmaram a estabilidade da unidade biológica como um homotetrâmero e que o CS8 se liga ao sítio ativo por um número maior de contatos em relação ao substrato. O conjunto destes resultados, mais o fato de que o CS8 foi filtrado com sucesso por critérios ADMETox, apontam para sua utilização em futuros estudos *in vitro* como um candidato alternativo no tratamento contra *Ca*.

GENERAL ABSTRACT

Fungal infections are responsible for thousands of people death every year. The fungus *Candida albicans* (*Ca*) is responsible for almost half of these deaths. There are few antifungal agents available for treatment and the continuously use of them will inevitably lead to the development of resistance. The discovery of new drugs, preferably using new targets is necessary. In this context, the enzymes of the shikimate pathway that are present in plants, fungi and bacteria represent excellent targets for the design of drugs, since they are absent in humans it is possible to find a compound with selective toxicity. This work presents the first study already carried out, targeting an enzyme of the shikimate of pathway in the fight against this disease. Thus, the objective of this work was to model the structure of the enzyme chorismate synthase (EC: 4.2.3.5) from *Ca* in order to use it in docking and molecular dynamics simulations, aiming to identify new compounds with specific action against candidiasis. The structure of *Ca*CS was modeled in tetrameric form and bound to the 5-EPSP substrate and FMNH₂ cofactor, using the homologous structures of *Saccharomyces cerevisiae* and *Streptococcus pneumoniae* as templates. The identity between sequences was 34-76%. The analysis of the contact interfaces between the subunits suggested that tetrameric structure was the biological unit of the protein in solution. In order to perform virtual screening, the tetrameric structure was initially minimized by conjugate gradient to remove possible stereochemical collisions between the protein and ligands. Then, the A subunit of the protein was used in the redocking of 5-EPSP aiming the validation of the docking protocol. The final protocol was then applied on the virtual screening using of two libraries, one formed by compounds similar to the known CS inhibitors (2501 compounds) and other formed by 120,435 natural products with commercial availability, all from Zinc database. The results identified the compound CS8 with theoretical $\Delta G_{\text{binding}}$ of -11,84 Kcal.mol⁻¹, which is lower than -11.44 Kcal.mol⁻¹ obtained for 5-EPSP, indicating that CS8 has increased affinity. This compound is there no patent registered neither any known applicability for it. The molecular dynamics simulations from *Ca*CS during 50 ns bound to 5-EPSP or CS8, confirmed the stability of the biological unit as homotetramer and that CS8 binds to the active site by a greater number of contacts than substrate. The set of those results plus the fact that CS8 was successfully filtered by ADMETox allow its use in *in vitro* assays aiming evaluations as a lead candidate for the treatment of *Ca*.

Modelagem por homologia da enzima corismato síntase de *Candida albicans* complexada com ligantes: Simulações de ancoragem e dinâmica molecular.

Paulo Sérgio Alves Bueno, Flavio Augusto Vicente Seixas*

Departamento de Bioquímica, Universidade Estadual de Maringá (UEM). Av. Colombo 5790, Maringá-PR, Brasil.

*Autor para correspondência: favseixas@uem.br

Fone: (+55) 44 36219337 Fax: (+55) 44 36219326

Resumo: Infecções fúngicas matam milhares de pessoas todos os anos e o fungo *Candida albicans* é responsável por quase metade dessas mortes. Os poucos antifúngicos disponíveis para o tratamento tendem a levar ao desenvolvimento de mecanismos de resistência. A descoberta de novas drogas, de preferência com novos alvos, é necessária. A via do chiquimato é essencial para a sobrevivência de fungos sendo ausente em humanos. A enzima corismato sintase (CS) (EC 4.2.3.5) participa na última etapa desta via, catalisando a conversão do 5-EPSP em corismato na presença do cofator FMNH₂. O objetivo deste trabalho foi o de modelar a estrutura da CS de *C. albicans* para ser utilizada em simulações de triagem virtual na busca de novas moléculas candidatas a inibidores. Para isso, a estrutura homotetramérica da enzima CS na presença do substrato 5-EPSP e do cofator FMNH₂ foi modelada por meio do programa Modeller 9v14. O melhor modelo foi utilizado em simulações de triagem virtual utilizando o programa Autodock e simulações de dinâmica molecular no do programa NAMD. Como resultados, a varredura da biblioteca de compostos com 70% de identidade aos inibidores conhecidos da CS identificou a molécula CS8, como sendo a mais bem pontuada (-11,84 kcal.mol⁻¹) em relação ao substrato (-11,44 kcal.mol⁻¹). Simulações de dinâmica molecular de equilíbrio mostraram que o CS8 permaneceu estável em seu sítio de ligação durante 50 ns, não levando a proteína a se desenovelar. Além disso, o CS8 foi filtrado com sucesso nos critérios ADMETox indicando que não seria tóxico às células humanas. Os resultados desta pesquisa poderão levar a identificação de uma molécula candidata a medicamento contra infecções causadas por *Candida* sp. Antes disso, a molécula CS8 deverá ser adquirida para estudos de inibição *in vitro*.

Palavras-chave: Corismato sintase, modelagem molecular, *Candida albicans*, ancoramento molecular, antifúngico.

Introdução

Infecções fúngicas invasivas caracterizam graves ameaças para a saúde humana, ocasionando pelo menos 1,5 milhões de mortes todos os anos no mundo [1]. Desse número, quase metade é causada pelo fungo da espécie *Candida albicans* (*Ca*) [2]. Ele vive de forma comensal no intestino, trato urogenital e outras superfícies mucosas, na proporção de um a cada dois indivíduos [3]. Suas vítimas são pacientes imunossuprimidos como receptores de transplantes, pacientes com câncer, vítimas de HIV, dentre outros que passam a ter a população do microorganismo aumentada, o que leva à doença infecciosa chamada candidíase [4,5]. Esta doença pode ser superficial e afetar a pele, as membranas mucosas ou, de forma mais agressiva, pode entrar na corrente sanguínea e invadir órgãos internos [6]. O arsenal de fármacos contra esse microorganismo é limitado a agentes antifúngicos como a anfotericina B, azólicos (itraconazol e fluconazol) e equinocadinas (caspofungina, micafungina). No entanto há relatos na literatura descrevendo cepas resistentes a cada um destes fármacos [6], o que limita ainda mais a sua utilização na medicina [7]. Portanto, há uma necessidade urgente da descoberta de novos agentes antifúngicos, com preferência para aqueles que possam atuar em diferentes alvos metabólicos.

Atualmente, o desenvolvimento de um agente antifúngico é um desafio, devido a poucos alvos de ação em potencial que não sejam compartilhados pelo fungo e a espécie humana (hospedeiro). Por ser uma via ausente em mamíferos [8], enzimas da via do chiquimato são promissores alvos para o desenvolvimento de novos antifúngicos [8], além de agentes antimicrobianos [9], antiparasitários [10] e herbicidas [11]. Ela converte, através de sete etapas catalisadas por enzimas, dois compostos presentes no metabolismo de carboidratos, o fosfoenolpiruvato (PEP) e a eritrose-4-fosfato (E4P) no produto final corismato. A partir desse, são produzidos metabólitos como ácido fólico, vitamina K, ubiquinona, fenilalanina, tirosina e triptofano, os quais são essenciais para a sobrevivência de fungos, bactérias, algas, plantas e parasitas do filo apicomplexa [12]. Desta forma, as enzimas que compõem essa via são excelentes alvos de drogas que permanecem inexplorados em *Ca*.

A enzima corismato síntase (CS) (EC.: 4.2.3.5) atua no último passo da via do chiquimato, catalisando a conversão do substrato 5-enolpiruvilchiquimato-3-fosfato (5-EPSP) em corismato na presença do cofator flavina mononucleotídeo reduzida (FMNH₂) [13,12]. Embora a estrutura tridimensional de *Ca*CS ainda não tenha sido determinada por nenhum método experimental, sua sequência de aminoácidos já possui status confirmado na base de dados Uniprot [14]. Além disso, já existem estruturas cristalográficas da CS depositadas no Protein Data Bank (PDB) [15], como a de *Saccharomyces cerevisiae*, a qual compartilha 76% de identidade com a enzima homóloga de *Ca*. Estas características permitem a modelagem por homologia da CS de *Cândida albicans*, de modo a

utilizá-la em estudos de modelagem e triagem virtual, com o intuito de descobrir novas moléculas candidatas a medicamentos contra infecções causadas por este fungo.

Materiais e Métodos

Obtenção da sequência e modelagem

A sequência de aminoácidos da CaCS escolhida foi a mais genérica, com base em um alinhamento pelo Clustal W [16], de todas aquelas existentes do banco de dados Uniprot [14]. Esta sequência foi utilizada na busca de moldes estruturais por meio de BLASTp [17] com filtro para estruturas cristalográficas no PDB. Em seguida, as sequências do molde e modelo foram alinhadas, e o resultado foi utilizado como parâmetro de entrada para o programa Modeller 9v14 [18]. Foi imposta uma restrição de simetria entre os C α das quatro cadeias para serem gerados 2000 modelos, onde os dez melhores foram escolhidos com base na ferramenta do Modeller, o DOPE Score. Destes, o melhor foi escolhido com base na qualidade estereoquímica pelo programa PROCHECK [19]. Resíduos em regiões proibidas do gráfico de Ramachandram foram corrigidos por meio de modelagens pontuais através do script *loop* do Modeller. Foi utilizado o programa Coot [20] para visualização de estruturas. As superfícies de contato entre as cadeias foram avaliadas pelo servidor PISA [21].

Simulações de Dinâmica Molecular

As simulações de dinâmica (MD) foram realizadas com o pacote de programas NAMD2/VMD [22,23]. A estrutura do homotetrâmero obtida com a modelagem foi solvatada com água TIP3P em uma caixa periódica com limites de 10 Å de distância a partir da superfície mais externa da proteína. Íons Na⁺ foram adicionados em quantidade suficiente para neutralização das cargas do sistema.

O processo de DM ocorreu em cinco etapas. Na primeira etapa, os átomos do ligante e do cofator foram fixados no espaço, enquanto que os da proteína e demais átomos do sistema foram mantidos livres para se movimentar. Neste ponto, o sistema foi submetido a 20.000 passos de minimização por Gradiente Conjugado (CG). Na segunda etapa, todos os átomos do sistema foram novamente minimizados por 10.000 passos de CG. A estrutura resultante dessa etapa foi utilizada em estudos de *redocking* e ancoragem.

Na terceira etapa, os átomos da proteína e ligantes foram mantidos fixos enquanto que as águas e sais foram submetidos a 60 ps de DM de equilíbrio. Na quarta etapa, todos os átomos foram submetidos a novo ciclo de CG por 20.000 passos. Na quinta e última etapa, todo o sistema foi

submetido a 50 ns de DM de equilíbrio em condições NPT (constantes: temperatura 300K, pressão 1 atm e número de átomos).

As análises foram realizadas em termos de raiz quadrada média da distância (RMSD), pelo raio de giro (R_{gyr}), bem como a raiz quadrada média de flutuação (RMSF) dos carbonos alfa para cada resíduo. Foram avaliados também a frequência de contatos pelos resíduos do sítio ativo da proteína em contato com os ligantes numa distância de até 4,0 Å. Todas essas análises foram realizadas a partir da estrutura de referência (quarta etapa da DM) pelo seu arquivo de trajetória.

As simulações de longa duração exigiram um maior poder de processamento, sendo realizadas em 20 nós com processadores Intel Xeon E5-2695v2 Ivy Bridge, 2,4GHZ (480 núcleos) do supercomputador SDumont no LNCC, Brasil.

Simulações de Ancoragem.

O protocolo de ancoragem foi estabelecido por estudos de *redocking* do 5-EPSP no complexo proteína-cofator minimizado (segunda etapa da DM) usando o programa AutoDock 4.2.3 [24] implementado na interface gráfica do programa Pyrx-0.9 [25]. Os parâmetros usados foram algoritmo padrão de busca (Lamarckian GA) e grid de 0,375 Å, com centro de busca em -19, 9, 18, tamanho da caixa 50, 50, 50 em x , y e z respectivamente. Foram restringidos os ângulos de torção da enzima e do cofator, mas permitiu-se a flexibilidade dos ligantes testados. Todas as moléculas de água e os íons foram removidos da estrutura.

Os protocolos foram considerados validados quando a pose do 5-EPSP se sobrepôs à pose modelada com RMSD menor que 0,8 Å. De acordo com a literatura, é aceitável um RMSD de até 2.0 Å [26].

Varredura Virtual

Para as simulações de varredura virtual foram construídas duas bibliotecas. A primeira foi montada por meio da busca de compostos comercializáveis, semelhantes aos inibidores conhecidos de CS (índice Tanimoto de 70%) obtidos na literatura, nos bancos de dados PubChem [27], Zinc [28], DrugBank [29] e Brenda enzymes [30]. A segunda biblioteca utilizada foi a de produtos naturais da base de dados Zinc [28]. Foi incluído nas bibliotecas a estrutura do 5-EPSP de modo que ao final da simulação, somente os compostos melhores ranqueados (menores $\Delta G_{\text{ligação}}$ teóricos) que o encontrado para esse ligante fossem considerados. Todas as moléculas estavam no formato **.sdf*.

Os resultados de três novas varreduras com estas bibliotecas (repetições) foram utilizados para selecionar os compostos com reprodutibilidade, de modo a eliminar resultados falso-positivos. Os compostos selecionados foram filtrados pelos parâmetros de Absorção, Distribuição,

Metabolismo, Excreção e Toxicidade (ADMETox) por meio do programa *Osiris Property Explorer* [31].

Campos de força e cálculo das cargas parciais.

O campo de força (FF) Charmm c35b2-c36a2 [32] foi utilizado para os átomos da proteína. Os arquivos com a estrutura dos ligantes 5-EPSP (Zinc4095612), CS8 e FADH₂ (Zinc31976681) foram obtidos do banco de dados Zinc [28], ajustados para seu estado de protonação em pH 7,0 [28] e submetidos ao servidor Swisparam [33], para gerar os respectivos campos de força no mesmo formato. As cargas de Müliken foram obtidas por meio de cálculos quânticos com o nível de teoria B3LYP/6-311G, utilizando o modelo de solvatação COSMO do programa ORCA [34] e aplicados aos campos de força.

Resultados e Discussão

Análise da sequência de CaCS.

Vinte e seis sequências de aminoácidos de CS de diferentes cepas de *Cândida albicans* foram obtidas no Uniprot e alinhadas. A cepa P75010 (Uniprot id: A0A0A6JZ95) foi escolhida como a representante mais genérica. A sequência de aminoácidos da CaCS possui 378 resíduos e, de acordo com o servidor Protparam [35], possui peso molecular de 40,895 kDa e pI teórico de 6,14. Em fungos essa enzima e a 2-ceto-3-desoxi D-arabinoeptulose-7-fosfato sintase, a primeira da via, são monofuncionais, enquanto que as outras cinco formam um polipeptídeo pentafuncional, responsável pelas outras reações [14].

Modelagem molecular e validação da estrutura da CaCS.

A análise da sequência pelo *Blastp* forneceu dois moldes estruturais. Para a cadeia polipeptídica foi utilizada a CS de *Saccharomyces cerevisiae* (PDB ID: 1R53) [36] com 76% de identidade. As coordenadas do substrato 5-EPSP e do cofator FMNH₂ foram obtidas na estrutura da CS de *Streptococcus pneumoniae* (PDB ID: 1QXO) [37] com 34% de identidade. Estudos demonstram que sequências de proteínas com mais de 25% de identidade possuem similaridade na estrutura tridimensional [38,39]. O molde 1R53 representa a unidade biológica da enzima com quatro cadeias (homotetrâmero), porém, muitos resíduos ao redor do sítio ativo estavam faltando, por falta de densidade eletrônica na região. A cadeia B do molde 1QXO foi utilizada para modelagem destas regiões. A CS possui duas conformações distintas do sítio ativo. Na forma Apo, algumas das alças ao redor do sítio estão em uma conformação aberta, enquanto que na presença do

ligante, estas alças estão na conformação fechada (ativa). A utilização do molde 1QXO permitiu que a CaCS fosse modelada na forma fechada e na presença do cofator FMNH₂ e do substrato 5-EPSP.

O alto grau de liberdade de movimento da proteína na forma monomérica poderia comprometer a configuração do sítio ativo, estruturado essencialmente por segmentos randômicos (alças e voltas) de estrutura. Estudos já realizados pelo método de DM em CS de *Mycobacterium tuberculosis* [40], demonstram que a enzima na forma tetramérica é mais estável do que as forma monomérica e dimérica.

Esse aspecto foi avaliado também por meio do programa Pisa [21]. O arranjo quaternário mostra que um número maior de contatos estabiliza preferencialmente os dois dímeros AB e CD, os quais em seguida se agrupam para formar o tetrâmero, com um menor número de contatos (tab. 1) Os valores de Δ^iG de associação calculado para as estruturas dos moldes e do modelo confirmam esta afirmação. Estes valores indicam a energia livre de solvatação (em Kcal.mol⁻¹) ganha com a formação da interface. Os valores são calculados com a diferença no total de energias de solvatação das estruturas isoladas e interfaceadas. Valores negativos de Δ^iG correspondem a interfaces hidrofóbicas ou afinidade positiva. Este valor não inclui o efeito das ligações de hidrogênio nem salinas ao longo da interface [21].

Apesar da estrutura do molde 1R53 ter muitos resíduos omitidos devido a falta de densidade eletrônica nestas regiões, a proporção dos contatos intermoleculares não mudou em relação a quantidade total de contatos, bem como a proporção da área de interface entre as subunidades em relação aos dímeros (tab. 1). O conjunto desses resultados confirma a similaridade estrutural do modelo com a unidade biológica tetramérica do molde.

Tab. 1 – Análise dos contatos de interface entre as estruturas molde e modelo.

Estrutura	Interfaces de monômero para dímero					Interfaces de dímero para tetrâmero				
	A para B e C para D					A para C e B para D				
	ⁱ N _{res}	N _{HB}	N _{SB}	Área da interface (Å ²)	Δ^iG (kcal/mol)	ⁱ N _{res}	N _{HB}	N _{SB}	Área da interface (Å ²)	Δ^iG (kcal/mol)
CaCS	105	39	6	3700	-32,2	44	21	4	1485	-17,2
1R53	68	28	4	2283	-24,2	30	11	5	1073	-10,6
1QXO	115	48	1	4348	-55,6	56	33	22	1932	-9,9

O fato do Δ^iG entre as subunidades não ser igual, indica que o contato entre os dímeros é mais estável que o contato entre os tetrâmeros, ou seja, os monômeros se associam primeiramente em dímeros que por sua vez, se associam em tetrâmeros. Apesar disso, não há nenhuma evidência na literatura que sugira um fenômeno de cooperatividade com a ligação do substrato.

Com essas evidências, a proteína foi então modelada como um homotetrâmero, sua mais provável unidade biológica, para que, dessa forma, fossem impostas restrições que diminuíssem as possíveis interferências ao objetivo principal deste trabalho, que foi a modelagem da proteína para estudos de varredura virtual. Depois de modelada e minimizada, somente uma das subunidades foi efetivamente utilizada nos estudos de ancoragem.

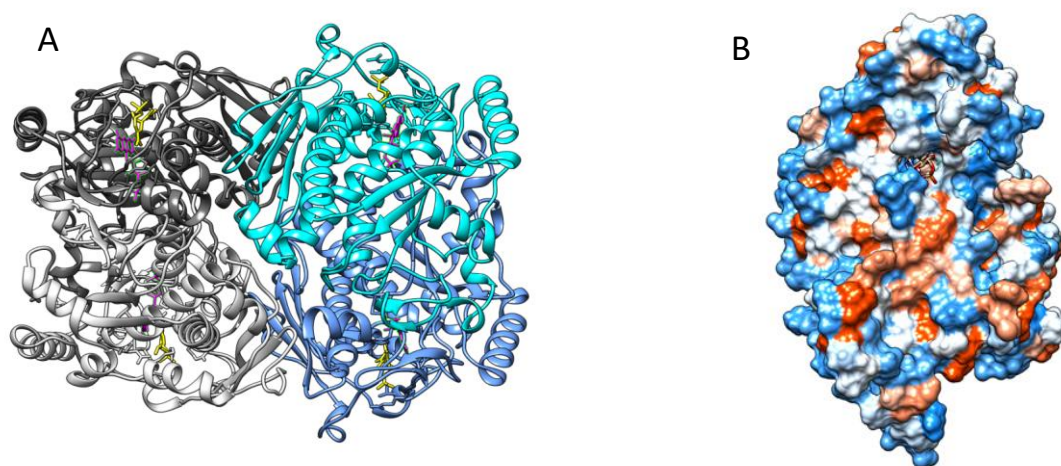


Fig. 1: A) Representação em fita da estrutura tetramérica da CaCS ligada ao cofator FMNH₂ e ao substrato 5-EPSP (amarelo) . B) Estrutura monomérica colorida por potencial eletrostático, com destaque para o sítio do FMNH₂ e ao e 5-EPSP com potencial predominantemente positivo (azul).

O melhor modelo ranqueado pelo Modeller *DOPE Score*, apresentava nove resíduos em posições não favoráveis do gráfico de Ramachandran, os quais foram corrigidos por modelagens pontuais. O modelo final apresentou apenas quatro resíduos de serina em regiões não permitidas do gráfico (um em cada monômero) (Fig. 1S). Estes resíduos não estão em contato direto com o substrato nem com o cofator. A estrutura final modelada apresentou 99,7% dos seus resíduos em regiões permitidas, apresentando excelente qualidade estereoquímica (Fig. 1).

Apesar da excelente qualidade estereoquímica, o cofator FMN presente no molde 1QXO estava na forma oxidada, o que deixava a enzima em um estado redox não ativo. Como o objetivo do trabalho é identificar um inibidor para a forma ativa da enzima (não há sentido em inibir uma enzima inativa) foi necessário corrigir o estado redox do cofator. Além disso, a estrutura global não se encontrava em seu nível de energia mais baixo. Para fazer essas correções, foram realizadas duas

etapas de CG da CaCS, onde o campo de força do cofator estava na forma reduzida (FMNH₂). Ao final da CG verificou-se que o cofator encontrava-se reduzido.

A viabilidade de utilização da CaCS em estudos de triagem virtual, bem como a validação do protocolo de docking foi realizada pelo método de *redocking*, utilizando o programa Autodock, o qual forneceu um RMSD de 0,762 Å (média de quatro simulações). Uma vez que os resultados foram reprodutíveis, a estrutura da CaCS e o protocolo de ancoragem foram validados e puderam ser aplicados nas simulações triagem virtual.

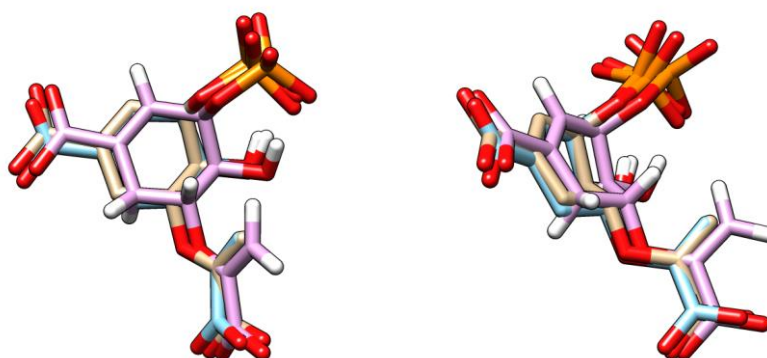


Fig 2: *Redocking* do 5-EPSP pelo programa Autodock. Bege: estrutura de referência modelada; Azul e Roxo: poses encontradas por *redocking*.

Triagem Virtual.

Para a triagem virtual, foram construídas duas bibliotecas de moléculas. A primeira continha 2501 moléculas similares aos inibidores conhecidos da CS, reportados na literatura. A segunda biblioteca utilizada foi a de produtos naturais do banco de dados Zinc, com 120.435 moléculas com disponibilidade comercial. Os candidatos foram subdivididos em pastas com aproximadamente 2.000 compostos cada, mais a estrutura do 5-EPSP (exógeno). O protocolo final validado no *redocking* foi então aplicado na triagem virtual das bibliotecas criadas. Na primeira triagem foram selecionadas 634 moléculas melhores ranqueadas que o 5-EPSP, as quais foram chamadas de *Best results*. Destas moléculas, 530 apresentaram reprodutibilidade em três novas triagens e, em todas as repetições, a molécula que a chamamos de CS8 sempre apareceu entre as cinco melhores ranqueadas, com um valor médio de $\Delta G_{\text{ligação}}$ de -11,84 Kcal.mol⁻¹, menor do que o obtido com o 5-EPSP que foi de -11,44 Kcal.mol⁻¹. O CS8 foi avaliado pelos parâmetros ADMETox, o qual indicou que o composto não teria efeitos tóxicos.

Simulações de Dinâmica Molecular

De modo a investigar a estabilidade estrutural da CaCS, bem como obter informações a respeito dos resíduos envolvidos na estabilização dos ligantes, foram realizadas simulações de DM por 50 ns. Nestas simulações o modelo final com o cofator FMNH₂ ligada ao EPSP ou a melhor pose do CS8 foram minimizadas e depois equilibradas conforme descrito na metodologia. O equilíbrio termodinâmico foi considerado atingido, quando o RMSD da cadeia principal da proteína, extraído da trajetória alinhada da dinâmica, atingiu um patamar constante por pelo menos 5 ns.

De acordo com a figura 3A, observa-se que as estruturas convergiram para o equilíbrio em a partir de 20 ns. Para a estrutura com o 5-EPSP o RMSD obtido foi de $3,0 \text{ \AA} \pm 0,4$. Este valor está próximo ao encontrados em estudos realizados com CS das bactérias *Mycobacterium tuberculosis* [40], *Shigella flexneri* [9] e do parasito *Plasmodium falciparum* [10]. Observa-se na figura 3A, que a CaCS ligada ao CS8 possui uma estabilidade maior em relação a estrutura ligada ao 5-EPSP, com um RMSD mais baixo em torno de $2,5 \text{ \AA} \pm 0,3$.

A estabilidade das estruturas foi evidenciada também em termos de raio de giro (R_{gyr}). Ele é definido como a distância quadrática média do conjunto de átomos a partir de seu centro de gravidade comum. A Fig. 3B também mostra que os complexos se estabilizaram após um tempo de aproximadamente 20 ns, e que a presença dos ligantes não levou a proteína a se desenovelar.

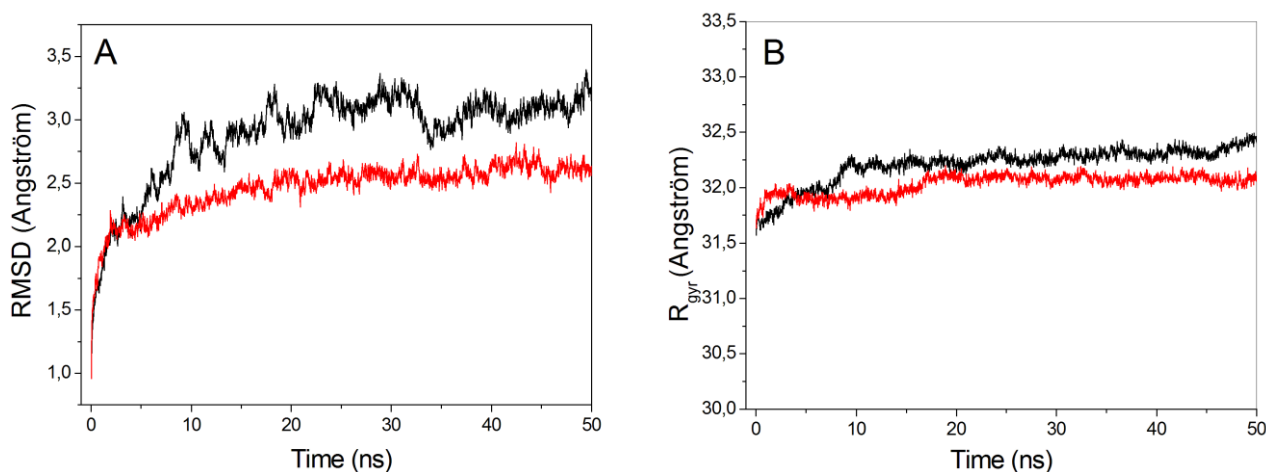


Fig 3: Comportamento das proteínas ao longo da simulação. Em **preto** o complexo CaCS-FMNH₂-EPSP e em **vermelho** o complexo CaCS-FMNH₂-CS8. (A) RMSD calculado para os átomos da cadeia principal. (B) Raio de giro calculado para os átomos da cadeia principal.

A partir da DM, foi possível avaliar regiões da proteína com maior flexibilidade por meio da raiz quadrada média da flutuação (RMSF) dos C α de cada resíduo na presença dos ligantes 5-EPSP

e CS8 (Fig. 4). O RMSF foi calculado nos últimos 5 ns de simulação, estando a proteína já em seu estado de equilíbrio. Os resíduos de maior flexibilidade foram os das regiões C-terminais em todas as cadeias, o que é comum na maioria das proteínas, e no caso da *CaCS*, remetem a um segmento sem conformação definida.

Verifica-se que existem três regiões que possuem uma maior flexibilidade quando qualquer um dos dois ligantes está complexado. As regiões 1, 2 e 3 são verificadas também em outro trabalho [40], o qual justifica que esse padrão é uma tendência de movimentação dos resíduos da proteína quando complexada a um ligante. Contudo, estas regiões compreendem longas alças que possuem, por natureza, um alto grau de liberdade de movimento.

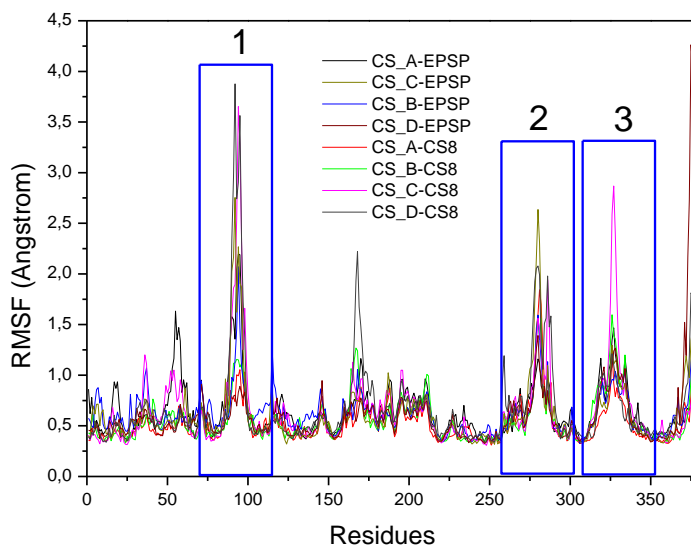


Fig 4: RMSF dos átomos do Ca da de *CaCS* nos últimos 5 ns de simulação (região de equilíbrio). As caixas em azul destacam as regiões com maior grau de flexibilidade.

Sítio ativo e a ligação do 5-EPSP

Foi avaliada a frequência de contatos entre os resíduos da *CaCS* com os ligantes numa distância de até 4,0 Å, ao longo da DM. O cálculo de frequência de contato integrou todo o tempo de simulação (50 ns). Foi delimitada uma frequência maior igual a 60%, para identificar os resíduos mais importantes na estabilização dos ligantes. Cada ligante possui geometria e composição específica, que os fazem interagirem de forma diferente com a proteína e, portanto, estruturando sítios distintos. Sendo assim, uma lista completa dos resíduos que compõe o sítio ativo para cada um dos ligantes 5-EPSP e CS8 é mostrada na tabela 2, na forma de probabilidade de ligação. Ela

mostra claramente que os resíduos Ser-16, Thr-55, Ser-125, Ala-126, Arg-127, Arg-337, Arg-345 são os mais importantes na ancoragem do 5-EPSP e do CS8, o que significa que qualquer mutação nestas posições poderia afetar significativamente a atividade enzimática e/ou provocar resistência a presença do ligante.

Tab. 2: Frequência de contato em até 4 Å entre a CaCS e os ligantes avaliados. Os resíduos em negrito possuem frequência de contato com o ligante maior que 60%.

Resíduos	5-EPSP	CS8	Resíduos	5-EPSP	CS8
Ser-16	0,71	0,67	Tyr-93	0,05	0,31
His-17	-	0,53	Arg-101	0,46	0,96
Arg-46	0,99	0,31	Ser-124	-	0,29
Arg-47	0,09	-	Ser-125	0,75	0,70
Lys-52	0,19	-	Ala-126	0,97	0,80
Thr-55	0,81	0,99	Arg-127	0,96	0,62
Pro-56	0,61	0,41	Thr-129	-	0,27
Asp-58	-	0,10	Ile-130	0,06	-
Glu-59	0,08	0,42	Thr-315	0,12	0,93
Lis-86	0,19	-	Ala-334	-	0,19
Asp-87	-	0,26	Arg-335	-	0,40
His-88	0,14	-	Gly-336	-	0,49
Arg-89	-	0,22	Arg-337	0,63	1,00
Pro-90	-	0,33	Arg-345	1,00	0,69

Ligante CS8

O ligante CS8 (Fig. 3B) não possui patente registrada nem aplicabilidade conhecida. A análise usando o índice de Tanimoto mostra que o CS8 tem 75% de similaridade com o ligante Zinc6445857 (CP1) (Fig.3C). O CP1 é um potente inibidor de CS do fungo *Paracoccidioides brasiliensis* descoberto por nosso grupo de pesquisa e patenteadado recentemente (Patente: BR 10 2016 021164 6), exibindo atividade fungicida e com atividade *in vivo* melhor que o medicamento de referência itraconazol. Os ligantes CS8 e CP1 possuem um peso molecular de 516,43 e 448,86 g/mol e *logP* de 4,09 e 2,98, respectivamente, indicando propriedades físicas parecidas. O maior número de contatos (tab.2) do CS8 em comparação com o 5-EPSP indica um maior preenchimento da cavidade do sítio ativo. Este pode ter sido o motivo do programa Autodock ter ranqueado o CS8 com um $\Delta G_{\text{ligação}}$ melhor do que o 5-EPSP. Essas características fazem do ligante CS8 um candidato em potencial para a inibição específica de CaCS.

Agradecimentos:

CAPES, CNPq, Fundação Araucária (147/14), LNCC.

Referências

1. Pianalto KM, Alspaugh JA (2016) New horizons in antifungal therapy. *J Fungi* 2 (4):24. doi:10.3390/jof2040026
2. Giacomazzi J, Baethgen L, Carneiro LC, Millington MA, Denning DW, Colombo AL, Pasqualotto AC (2016) The burden of serious human fungal infections in Brazil. *Mycoses* 59 (3):145-150. doi:10.1111/myc.12427
3. da Silva Dantas A, Lee KK, Raziunaite I, Schaefer K, Wagener J, Yadav B, Gow NA (2016) Cell biology of *Candida albicans*-host interactions. *Curr Opin Microbiol* 34:111-118. doi:10.1016/j.mib.2016.08.006
4. Kabir MA, Hussain MA, Ahmad Z (2012) *Candida albicans*: A Model Organism for Studying Fungal Pathogens. *ISRN Microbiol* 2012:538694. doi:10.5402/2012/538694
5. Sheehan DJ, Hitchcock CA, Sibley CM (1999) Current and emerging azole antifungal agents. *Clin Microbiol Rev* 12 (1):40-79
6. Spampinato C, Leonardi D (2013) *Candida* infections, causes, targets, and resistance mechanisms: traditional and alternative antifungal agents. *Biomed Res Int* 2013:204237. doi:10.1155/2013/204237
7. Campoy S, Adrio JL (2016) Antifungals. *Biochem Pharmacol pii: S0006-2952(16)30422-1*. doi:10.1016/j.bcp.2016.11.019
8. Kaldorf M, Srivastava M, Gupta SK, Liang C, Binder J, Dietl AM, Meir Z, Haas H, Oshero N, Krappmann S, Dandekar T (2016) Systematic Identification of Anti-Fungal Drug Targets by a Metabolic Network Approach. *Front Mol Biosci* 3:22. doi:10.3389/fmolb.2016.00022
9. Zhou H, Singh NJ, Kim KS (2006) Homology modeling and molecular dynamics study of chorismate synthase from *Shigella flexneri*. *J Mol Graph Model* 25 (4):434-441. doi:10.1016/j.jmgm.2006.02.013
10. Tapas S, Kumar A, Dhindwal S, Preeti, Kumar P (2011) Structural analysis of chorismate synthase from *Plasmodium falciparum*: a novel target for antimalaria drug discovery. *Int J Biol Macromol* 49 (4):767-777. doi:10.1016/j.ijbiomac.2011.07.011
11. Herrmann KM, Weaver LM (1999) The Shikimate Pathway. *Annu Rev Plant Physiol Plant Mol Biol* 50:473-503. doi:10.1146/annurev.arplant.50.1.473
12. Fernandes CL, Breda A, Santos DS, Basso LA, Souza ON (2007) A structural model for chorismate synthase from *Mycobacterium tuberculosis* in complex with coenzyme and substrate. *Comput Biol Med* 37 (2):149-158. doi:10.1016/j.combiomed.2006.01.001
13. Macheroux P, Petersen J, Bornemann S, Lowe DJ, Thorneley RN (1996) 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. doi:10.1021/bi951705u
14. UniProt: a hub for protein information (2015). *Nucleic Acids Res* 43 (Database issue):D204-212. doi:10.1093/nar/gku989

15. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58 (Pt 6 No 1):899-907. doi.org/10.1107/S0907444902003451.
16. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics Chapter 2:Unit 2.3*. doi:10.1002/0471250953.bi0203s00
17. Mount DW (2007) Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* 2007:pdb top17. doi:10.1101/pdb.top172007/7/pdb.top17 [pii]
18. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics Chapter 5:Unit 5.6*. 10.1002/0471250953.bi0506s15.
19. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8 (4):477-486
20. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60 (Pt 12 Pt 1):2126-2132. doi:S0907444904019158 [pii]10.1107/S0907444904019158
21. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372 (3):774-797. doi:10.1016/j.jmb.2007.05.022
22. 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. 10.1002/jcc.20289
23. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14 (1):33-38, 27-38
24. 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
25. Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. *Methods Mol Biol* 1263:243-250. doi:10.1007/978-1-4939-2269-7_19
26. Claussen H, Buning C, Rarey M, Lengauer T (2001) FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 308 (2):377-395. doi:10.1006/jmbi.2001.4551
27. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH (2016) PubChem Substance and Compound databases. *Nucleic Acids Res* 44 (D1):D1202-1213. doi:10.1093/nar/gkv951
28. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Inf Model*. doi:10.1021/ci3001277
29. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 39 (Database issue):D1035-1041. doi:gkq1126 [pii]10.1093/nar/gkq1126
30. Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, Scheer M, Schomburg D (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* 41 (Database issue):D764-772. doi:10.1093/nar/gks1049

31. Miteva MA, Violas S, Montes M, Gomez D, Tuffery P, Villoutreix BO (2006) FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res* 34 (Web Server issue):W738-744. doi:10.1093/nar/gkl065
32. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD, Jr. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* 31 (4):671-690. doi:10.1002/jcc.21367
33. 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
34. Neese F (2012) The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2 (1):73-78. doi:10.1002/wcms.81
35. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531-552
36. Quevillon-Cheruel S, Leulliot N, Meyer P, Graille M, Bremang M, Blondeau K, Sorel I, Poupon A, Janin J, van Tilbeurgh H (2004) Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. *J Biol Chem* 279 (1):619-625. doi:10.1074/jbc.M310380200
37. Maclean J, Ali S (2003) The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure* 11 (12):1499-1511. doi:10.1016/j.str.2003.11.005
38. Yang AS, Honig B (2000) An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J Mol Biol* 301 (3):691-711. doi:10.1006/jmbi.2000.3975
39. Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng* 12 (2):85-94
40. Rocha KB (2011) Virtual screening and molecular dynamics for the identification of inhibitors of the enzymes chorismate synthase of mycobacterium tuberculosis. Dissertation.

MATERIAL SUPPLEMENTAR

Ramachandran Plot
CaCS.BL02680001

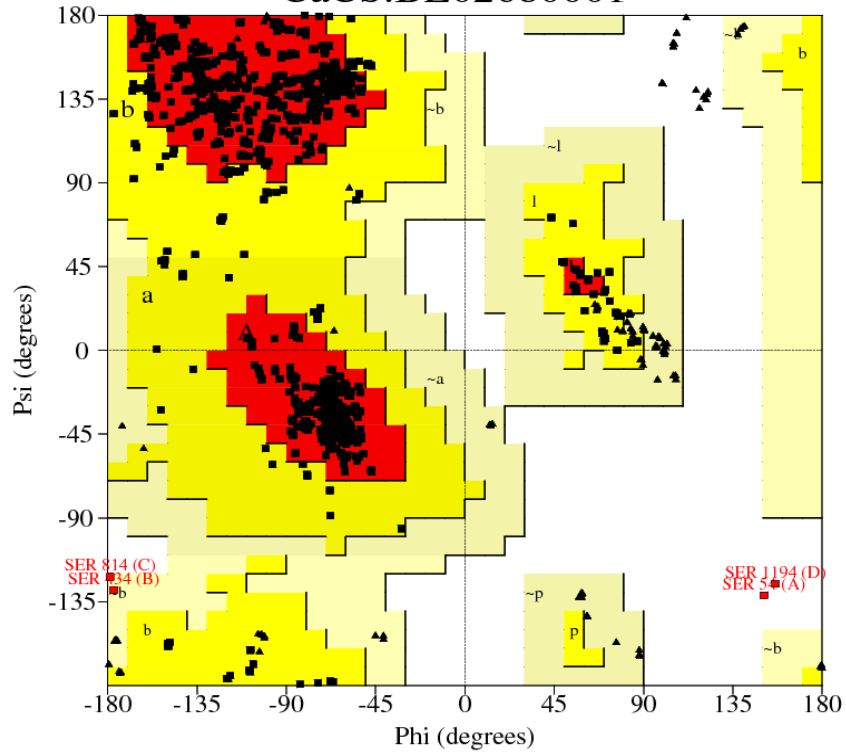


Fig. 1S: Gráfico de Ramachandran do modelo final da CS de *Candida albicans* obtido pelo programa Procheck.

RESEARCH ARTICLE

Open Access



Structural and functional analysis of four non-coding Y RNAs from Chinese hamster cells: identification, molecular dynamics simulations and DNA replication initiation assays

Quirino Alves de Lima Neto^{1*}, Francisco Ferreira Duarte Junior¹, Paulo Sérgio Alves Bueno², Flavio Augusto Vicente Seixas², Madzia Pauline Kowalski³, Eyemen Kheir³, Torsten Krude³ and Maria Aparecida Fernandez¹

Abstract

Background: The genes coding for Y RNAs are evolutionarily conserved in vertebrates. These non-coding RNAs are essential for the initiation of chromosomal DNA replication in vertebrate cells. However thus far, no information is available about Y RNAs in Chinese hamster cells, which have already been used to detect replication origins and alternative DNA structures around these sites. Here, we report the gene sequences and predicted structural characteristics of the Chinese hamster Y RNAs, and analyze their ability to support the initiation of chromosomal DNA replication in vitro.

Results: We identified DNA sequences in the Chinese hamster genome of four Y RNAs (*chY1*, *chY3*, *chY4* and *chY5*) with upstream promoter sequences, which are homologous to the four main types of vertebrate Y RNAs. The *chY1*, *chY3* and *chY5* genes were highly conserved with their vertebrate counterparts, whilst the *chY4* gene showed a relatively high degree of diversification from the other vertebrate Y4 genes. Molecular dynamics simulations suggest that *chY4* RNA is structurally stable despite its evolutionarily divergent predicted stem structure. Of the four Y RNA genes present in the hamster genome, we found that only the *chY1* and *chY3* RNA were strongly expressed in the Chinese hamster GMA32 cell line, while expression of the *chY4* and *chY5* RNA genes was five orders of magnitude lower, suggesting that they may in fact not be expressed. We synthesized all four *chY* RNAs and showed that any of these four could support the initiation of DNA replication in an established human cell-free system.

Conclusions: These data therefore establish that non-coding *chY* RNAs are stable structures and can substitute for human Y RNAs in a reconstituted cell-free DNA replication initiation system. The pattern of Y RNA expression and functionality is consistent with Y RNAs of other rodents, including mouse and rat.

Keywords: Non-coding RNAs, Chinese hamster Y RNAs, DNA replication, Nucleic acid simulations