Pontifícia Universidade Católica do Rio Grande do Sul

Faculdade de Biociências

Programa de Pós-Graduação em Biologia Celular e Molecular

Luís Fernando Saraiva Macedo Timmers

Orientador: Dr. Walter Filgueira de Azevedo Jr.

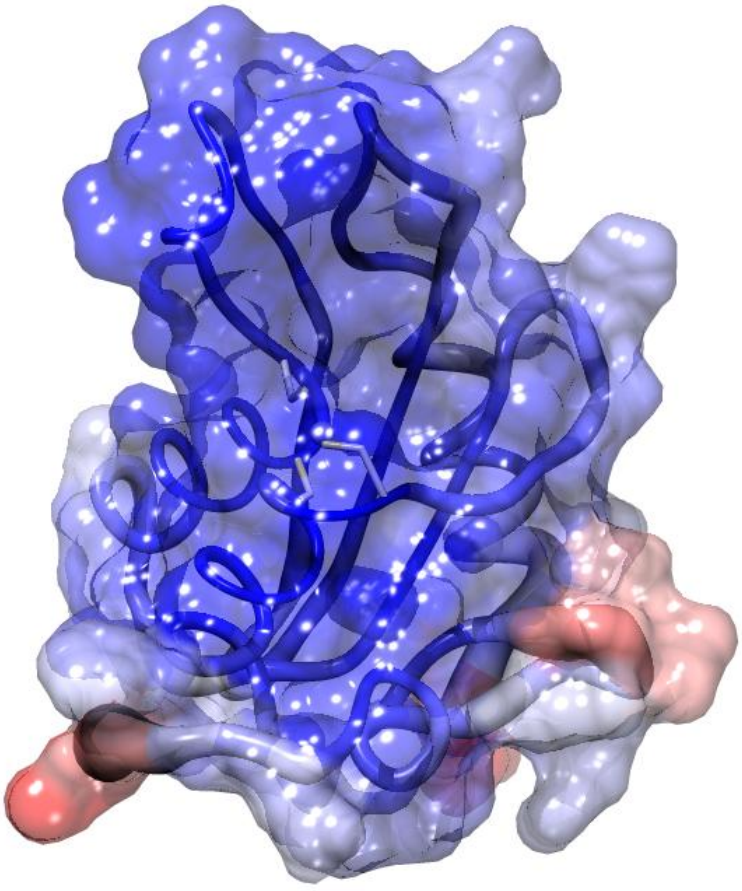
Estudos cristalográficos e aplicação da técnica de triagem virtual de ligantes para a busca de novos inibidores contra a enzima Citidina Deaminase (CDA; EC 3.5.4.5) de *Mycobacterium tuberculosis* H37Rv

Porto Alegre

2011

Luís Fernando Saraiva Macedo Timmers

Estudos cristalográficos e aplicação da técnica de triagem virtual de ligantes para a busca de novos inibidores contra a enzima Citidina Deaminase (CDA; EC 3.5.4.5) de *Mycobacterium tuberculosis* H37Rv



Dissertação apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular da Pontifícia Universidade Católica do Rio Grande do Sul como requisito para a obtenção do título de Mestre em Biologia Celular e Molecular.

Porto Alegre

2011

AGRADECIMENTOS

Ao Prof. Dr. Walter Figueira de Azevedo Jr., agradeço pela oportunidade de integrar seu grupo de pesquisa, Laboratorio de Bioquímica Estrutural (LaBioQuest), do qual faço parte desde 2006, pelo apoio dado desde o início da minha formação cientifica e por todo o conhecimento compartilhado. Ao Prof. Dr. Diógenes Santiago Santos, agradeço por fazer possível que eu fizesse o curso de Pos-Graduação na PUCRS. Ao Prof. Dr. Luiz Augusto Basso, agradeço por todas as sugestões dadas durante o andamento deste trabalho e pelas incessantes correções dos artigos.

Ao Guy Barros Barcellos, agradeço pela amizade durante todo o curso de graduação e por me apresentar a área de biologia estrutural, na qual iniciei a minha formação cientifica.

A Ivani Pauli, agradeço por todo apoio, companhia e carinho em todas as horas, fazendo possível para que este trabalho pudesse ser concluído.

Ao Ms.C Rafael Andrade Caceres, agradeço por todos os ensinamentos durante esta longa jornada, pelos conselhos em horas de desespero que fizeram eu seguir em frente, pelo apoio, amizade em todas as horas. E que certamente sem ele este trabalho não seria possível de ser concluído.

A minha mãe, Magda Saraiva Macedo, pelo amor incondicional, por não medir esforcos para que eu chegasse ate aqui, por todo apoio nas horas difícieis. Por mais que eu tentasse transcrever em palavras, estas não seriam suficientes para demonstrar o quanto sou agradecido a ela.

Ao meu irmao, Joao Augusto Saraiva Macedo Timmers, agradeço pelo apoio em todas horas e pelos momentos de descontração proporcionados.

RESUMO

A tuberculose tem sido um flagelo da humanidade desde os tempos antigos. Esta doença tem cobrado um preço alto em vidas humanas, por esta razão muitos esforços tem sido feitos com o intuito de derrotar o *Mycobacterium tuberculosis*. A sequência genômica do *Mycobacterium tuberculosis* deu início a uma nova era e foi de grande impacto para o presente conhecimento deste patógeno como também na pesquisa de drogas anti-microbiais. Com estes avanços foi possível identificar genes e mais tarde validar vias metabólicas. O melhor entendimento das vias metabólicas presentes no *Mycobacterium tuberculosis* corresponde ao primeiro passo e configura uma possibilidade de identificação de novos alvos moleculares que possam erradicar o bacilo. Entretanto, ainda há muito a fazer, desde a descoberta da Rifampicina, a droga de primeira linha efetivamente utilizada contra TB, nenhuma nova droga quimioterápica foi identificada. Nos dias de hoje, a TB é responsável pela morte de dois milhões de pessoas por ano, sendo considerada uma emergência global, a qual precisa ser urgentemente erradicada. O presente projeto tem o objetivo de realizar a análise estrutural da enzima Citidina Deaminase (CDA; EC 3.5.4.5) de *Mycobacterium tuberculosis* associada com seus substratos e produtos, aliando técnicas de Dinâmica e Docagem molecular.

ABSTRACT

The *consumption* has been a scourge of mankind since ancient times. This illness have been charged a high price in human lives, so many efforts have been made in order to defeat the *Mycobacterium tuberculosis.* The complete genomics sequence of *Mycobacterium tuberculosis* opened a new era and had a major impact in our understanding of this pathogen and also in research for microbial drugs. The researches could identify genes and later validate the metabolic pathways. The better understanding of metabolic pathways present in *Mycobacterium tuberculosis* corresponds to a first step and configures a possibility of identification of new molecular targets that could be derail the bacilli. However, there is much more to do, since of the discovery of rifampicin, a first-line drug used effectively against TB, none new chemotherapy has been released. TB is nowadays responsible for two million of deaths/year and is a global emergency which needs to be urgently repressed. The present work aim to realize structural analysis of the enzyme Cytidine Deaminase (CDA; EC 3.5.4.5) from *Mycobacterium tuberculosis* associated with substrates and products, allying techniques such as molecular docking and dynamics simulations.

LISTA DE ILUSTRAÇÕES

LISTA DE SIGLAS E ABREVIATURAS

TB: Tuberculose.

MtCDA: Citidina Deaminase de *Mycobacterium tuberculosis*

Mtb: *Mycobacterium tuberculosis*

HIV: Vírus da Imunodeficiência Humana

AIDS: Síndrome da Imunodeficiência Adquirida ou SIDA

OMS: Organização Mundial da Saúde

DOTS: *Directly Observed Treatment, Short Course*, tratamento padrão de "curta duração preconizado contra a TB pela OMS.

MDR-TB: TB multi-resistente à drogas

XDR-TB: TB extensivamente resistente à drogas

INCT-TB: Instituto Nacional de Ciência e Tecnologia em Tuberculose

LaBioQuest: Laboratório de Bioinformática Estrutural

SUMÁRIO

1. INTRODUÇÃO

1.1 Tuberculose – A peste cinzenta

O *Mycobacterium tuberculosis,* também conhecido por bacilo de Koch foi identificado em 24 de março de 1882 por Robert Koch, como sendo o agente causador da tuberculose (TB). As doenças infecciosas ainda são responsáveis pelo sofrimento e morte de centenas de milhões de pessoas e 90% das mortes causadas por esse tipo de enfermidade ocorrem em áreas tropicais e subtropicais, regiões nas quais estão presentes 80% de toda população mundial [Trouiller *et al.*, 2001].

A TB é uma das principais causas de morte devido a apenas a um agente infeccioso que ameaça a saúde global. De acordo com a Organização Mundial da Saúde (OMS) aproximadamente dois bilhões de pessoas, um terço da humanidade, estão infectadas. De acordo com o último relatório da OMS, foram estimados 9,27 milhões de novos casos de TB em 2007 [Donald & van Helden, 2009; Frieden *et al.*, 2003]. Apesar de todos estes números, a TB é considerada uma doença negligenciada, tendo em vista que a maioria dos casos reportados é de países emergentes. Entretanto, desde 1980 a TB tem sido reportada em países como Estados Unidos e Inglaterra.

1.1.1 Ressurgência da Tuberculose

A ressurgência da TB nestes países está intimamente relacionada com os casos da Síndrome da Imunodeficiência Adquirida (SIDA). Este sinergismo

da SIDA/TB tem sido um desafio para as indústrias farmacêuticas na busca de novas drogas. Alguns casos relatados de coinfecção de SIDA/TB têm dificultado o tratamento de ambas às doenças, visto que as terapias antirretrovirais não são compatíveis com as drogas utilizadas para o tratamento da TB. Esta incompatibilidade ocorre pelo fato das terapias compartilharem enzimas metabólicas, as quais são de extrema importância para os mecanismos celulares. Um dos casos mais sérios de interação medicamentosa ocorre quando, a família das iso-enzimas conhecidas como citocromo P450 (CYP) são ativadas ou inibidas [Burman *et al.*, 1999].

1.1.2 Tratamento da Tuberculose

A busca por drogas potenciais para o tratamento da TB iniciou aproximadamente em 1940. Selman A. Waksman recebeu o Prêmio Nobel em 1952, quando, seu laboratório descobriu a estreptomicina (EST). Waksman anunciou que esta droga era o caminho para a erradicação da “Grande praga branca” [Jassal & Bishai, 2009]. Não obstante, o pronunciamento foi prematuro e meses depois do início da utilização da EST como medicamento, cepas de Mtb resistentes a EST foram encontradas. Outras diversas moléculas para o tratamento da TB foram identificadas, como, por exemplo, ácido para-aminosalicílico, rifampicina (RIF), isoniazida (INH), pirazinamida (PYZ), etambutol (EMB), fluoroquinolonas, aminoglicosídeos, polipeptídeos, tioamidas e cicloserina. Contudo, estas drogas não podem ser utilizadas como únicos agentes, devido ao aparecimento de cepas resistentes a diversas drogas.

Passados seis anos após a OMS declarar a TB como uma emergência global, em 1999 foi estabelecido um tratamento padrão contra a TB, o qual foi denominado de Tratamento Diretamente Supervisionado da Tuberculose (da sigla em inglês “DOTS”) [WHO, 1999]. O tratamento quimioterápico apresenta uma duração de seis meses, o qual é divido em duas fases. A primeira fase do tratamento é de dois meses combinando quarto tipos de drogas de primeira linha, que são a INH, a RIF, a PYZ e o EMB ou a EST, com o intuito de exercer uma ação bactericida contra o bacilo ativo. Já na segunda etapa do tratamento, são utilizadas as duas principais drogas de primeira linha a INH e a RIF por mais quatro meses.

1.1.3 Resistência as drogas existentes

A TB pode ser tratada, entretanto, o longo tempo que é requerido para a terapia, o número de doses que necessitam ser administradas e os diversos efeitos colaterais provindos das drogas são um agravante para que os infectados não cumpram o tempo necessário da terapia. O uso contínuo das drogas nos primeiros dois meses de tratamento, já são o suficiente para o desaparecimento dos sintomas, proporcionando ao paciente um falso positivo de cura. Esta “pseudo” cura faz com que o indivíduo pare o tratamento antes do final do período determinado. Esta displicência é uma das principais causas para o surgimento de cepas Resistentes a Múltipals Drogas (MDR-TB) e Extensivamente Resistentes as Drogas (XDR-TB).

As cepas MDR-TB são denominadas desta forma, por serem resistentes a duas drogas de primeira linha (INH e RIF). O tratamento tem uma duração

maior do que a terapia padrão, tendo que frequentemente combinar drogas de segunda linha, aumentando desta maneira a toxicidade [Rivers & Mancera, 2008; Johnson *et al.*, 2006].

O outro tipo de cepa resistente ao tratamento padrão são as XDR-TB, as quais apresentam resistência a INH e RIF, assim como a drogas de segunda linha, como Ciprofloxacina, Moxifloxacina, Fluoroquinolonas e a mais uma das três dorgas injetáveis (Kanamicina, Capreomicina e Aminkacina) [Kinnings, 2009]. Estas cepas são consideradas virtualmente incuráveis, pelo fato de que a classe das drogas que precisariam ser administradas apresenta uma baixa potência e uma alta toxicidade [Rivers & Mancera, 2008; Kinnings *et al.*, 2009].

1.2 Biossíntese de Nucleotídeos Pirimídicos

A vida depende da capacidade das células de armazenar, resgatar e traduzir as informações genéticas necessárias para a manutenção do organismo. Essas funções são desempenhadas pelos ácidos nucléicos cujas unidades constitutivas são os nucleotídeos [Alberts *et al.*, 2002]. Os nucleotídeos estão envolvidos em todas as facetas da vida celular, participando de reações de oxidação-redução, transferência de energia, sinalização intracelular e reações biossintéticas. Seus polímeros, os ácidos nucléicos DNA e RNA, são os participantes básicos no armazenamento e na decodificação da informação genética. Os nucleotídeos e os ácidos nucléicos também desempenham funções estruturais e catalíticas nas células [Voet & Voet, 2008].

A biossíntese de pirimidinas representa uma das vias mais antigas presentes nas células e sua sequencia se manteve intacta durante a hierarquia

evolucional [Shambaugh, 1979]. Existem duas principais vias para a síntese de nucleotídeos pirimídicos, via *de novo*, a qual apresenta um gasto energético mais elevado, inicia a síntese de nucleotídeos a partir de precursores simples, e a via de salvamento, que reutiliza bases pirimídicas e nucleosídeos derivados de nucleotídeos pré-formados para a síntese de ribonucleotídeos e desoxirribonucleotídeos [Xu *et al.*, 2008].

1.2.1 Síntese de novo de Pirimidinas

A via de novo de nucleotídeos de pirimidina é composta de seis reações enzimáticas, tendo como produto final a formação de uridina 5'-trifosfato (UTP) e citidina 5'-trifosfato (CTP) [Islam *et al.*, 2007]. A rota inicia com a Carbamoil Fosfato Sintetase, a qual utiliza como substrato ATP dependente de Mg, glutamina e bicarbonato para a formação do Carbamoil fosfato. Este produto então é combinado com aspartato formando o carbamoil aspartato, reação esta catalizada pela Aspartato Transcarbamilase. O carbamoil aspartato é agora convertido a dihidroorotato pela Dihidrooratase [Shambaug, 1979]. A quarta reação da via é a conversão do dihidroorotato a orotato pela Dihidroorotato Desidrogenase. Na quinta reação da via é adicionado ao orotato o grupo fosforribosil do 5-fosforribosil-1-pirofosfato (PRPP) gerando a orotidina 5'-monofosfato (OMP) a qual é descarboxilada formando uridina 5'-monofosfato (UMP), o primeiro nucleotídeo pirimidínico (**Figura 1**). A formação até UMP, na quinta etapa da via, é catalizada por duas reações sucessivas, pelas enzimas Orotato-Fosforribosil Transferase e a Orotidina-5'-Fosfato Descarboxilase. O produto destas reações, o UMP, é fosforilado por quinases até UTP. A formação da CTP ocorre após a aminação do UTP, reação esta catalizada pela

CTP sintetase, a qual utiliza como produto UMP, glutamina e ATP [Moffatt & Ashihara, 2002].

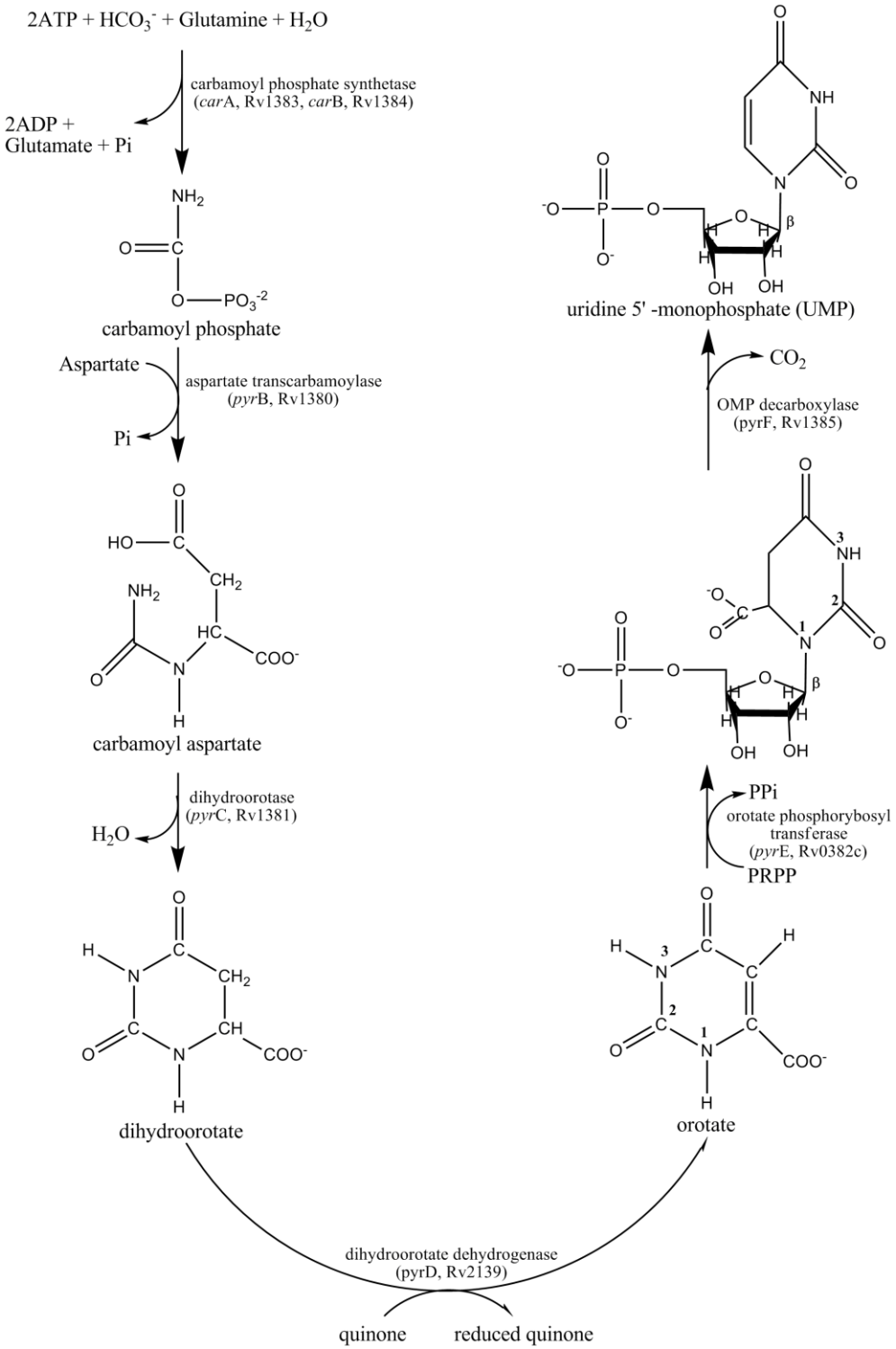


Figura 1. Rota de síntese de novo de pirimidinas. A imagem foi gerada com o ChemDraw (Mills, 2006).

1.2.2 Via de Salvamento de Pirimidinas

A via de salvamento permite que o organismo adquira suporte exógeno, como, por exemplo, bases pirimídicas e nucleosídeos, os quais não são intermediários na síntese *de novo* de pirimidinas [Johansson *et al.*, 2002]. Esta via é preferencialmente utilizada pelo fato de não demandar tanta energia como na via *de novo*. Em relação às bases, apenas a uracila é diretamente reutilizada pela uracil fosforribosil transferase (UPRT), entretanto os nucleosídeos pirimídicos, como, uridina, citidina e deoxicitidina são reutilizados exclusivamente na síntese seus respectivos nucleotídeos, UMP, CMP e dCMP [Moffatt & Ashihara, 2002]. Umas das enzimas chave da via de salvamento de pirimidinas é a Citidina Deaminase (CDA). A **Figura 2** apresenta a via de salvamento de nucleotídeos de pirimidina.

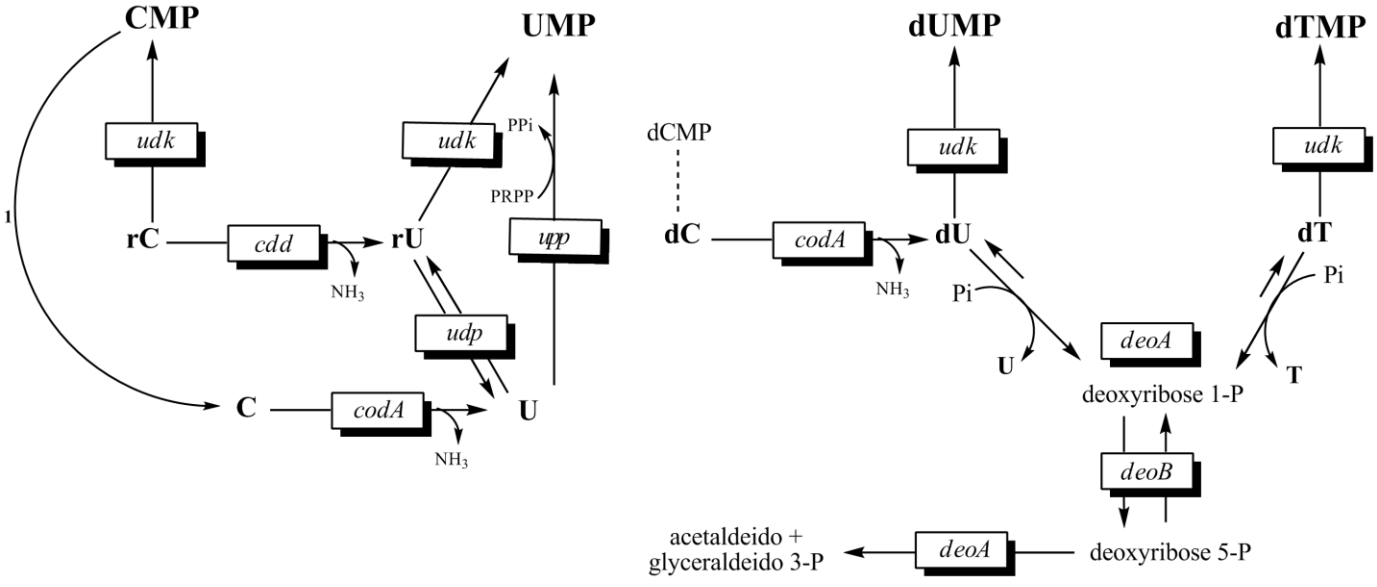


Figura 2. Rota de salvamento de pirimidinas. As enzimas envolvidas nesta rota são identificadas pelo símbolo de seu gene: cdd, citidina (desoxicitidina) deaminase; codA, citosina deaminase; deoA, timina (desoxiuridina) fosforilase; deoB, fosfopentomutase; deoC, deoxiriboaldose; tdk, timidina (desoxiuridina) quinase; thyA, TSase; udk,uridina (citidina)

quinase; udp, uridina fosforilase; upp, UPRTase; 1, CMP glicosilase. A imagem foi gerada com o ChemDraw (Mills, 2006).

1.3 Citidina Deaminase

A CDA é a enzima responsável pela deaminação hidrolítica, irreversível, da citidina e desoxicitidina gerando como produto uridina e desoxiuridina, respectivamente (**Figura 3**).

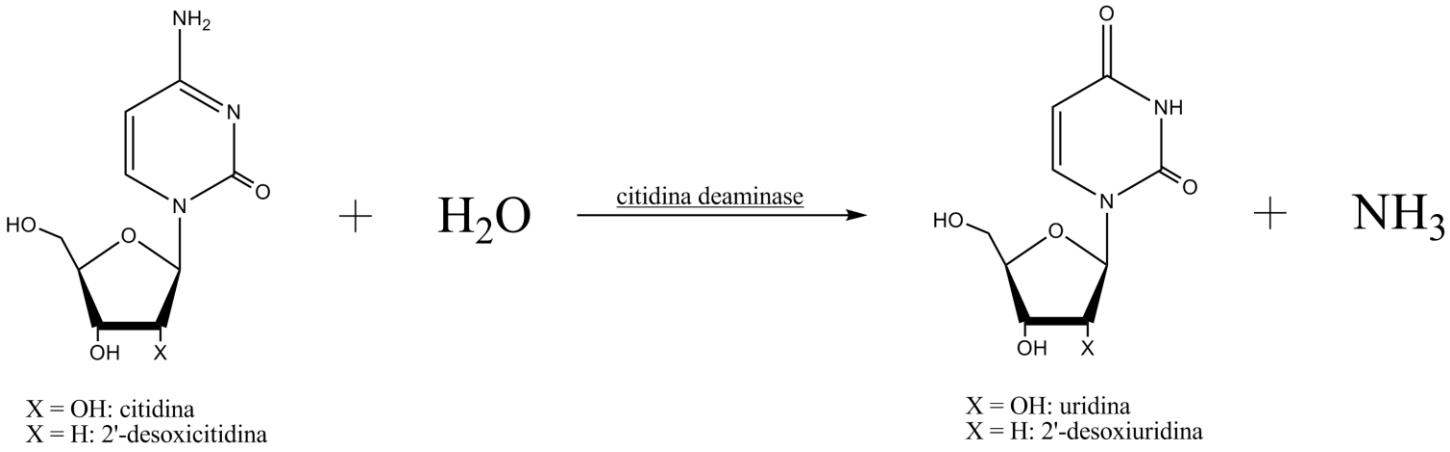


Figura 3. Reação catalisada pela CDA. A imagem foi gerada com o ChemDraw (Mills, 2006).

Estudos sobre esta enzima em outros organismos, como, por exemplo, *E.coli*, *Bacillus subtilis* [Johansson *et al.*, 2002; Xiang *et al.*, 1996] sugerem que o zinco (Zn^{+2}), presente no sítio ativo, apresenta uma papel central no mecanismo catalítico proposto para a CDA, ativando uma molécula de água, formando um íon hidróxido o qual fará um ataque nucleofílico no carbono 4 (C4) do anel da citidina pela protonação do adjacente na posição do nitrogênio

3 (N3). Esta reação gera um composto tetraédrico, o qual irá se decompor pela eliminação da amônia (NH_3), gerando a uridina ou desoxiuridina, dependendo do substrato (**Figura 4**).

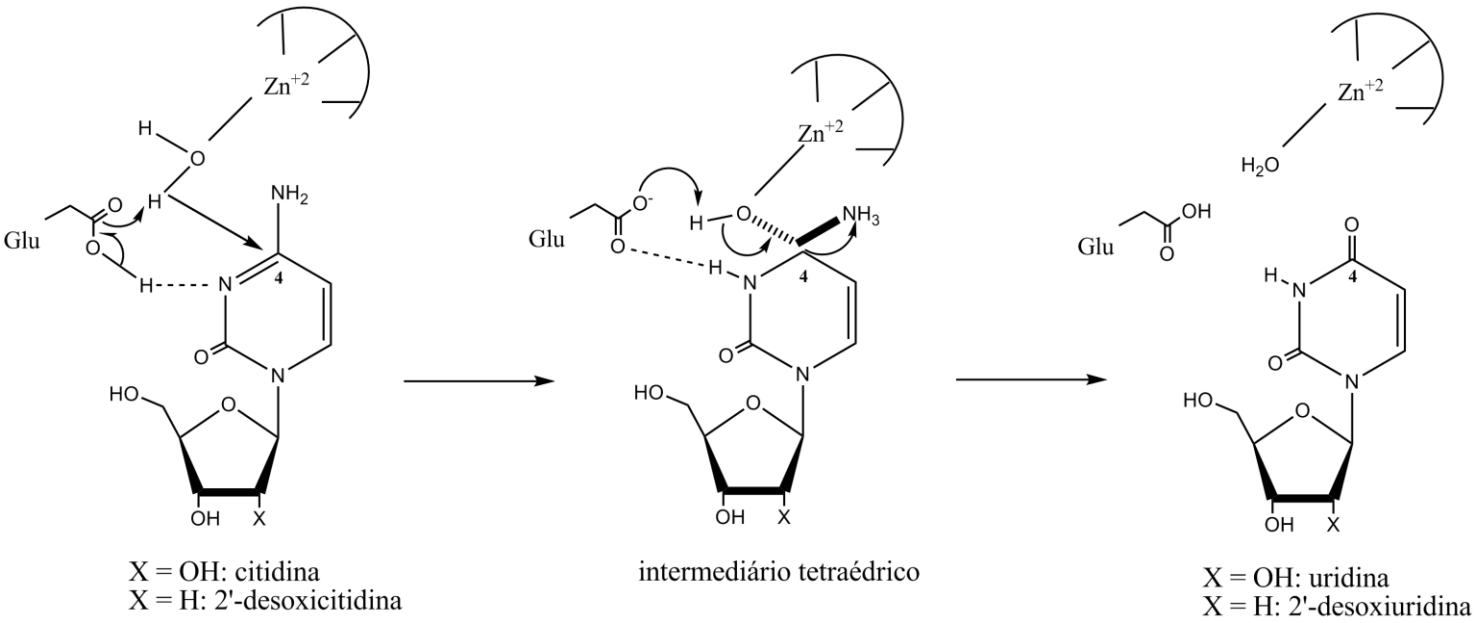


Figura 4. Mecanismo de reação proposto para a CDA. A imagem foi gerada com o ChemDraw (Mills, 2006).

Na natureza, até agora, foram identificadas duas formas da CDA. As homodiméricas (D-CDA) que são caracterizadas por apresentar massa molecular de aproximadamente 32kDa cada subunidade. As homodiméricas também apresentam três resíduos conservados, uma histidina (His) e duas cisteínas (Cys), que são responsáveis pelo ancoramento do Zn^{+2}. Estas D-CDA estão presentes em organismos como, por exemplo, *E. coli* [Betts *et al.*, 1994] e *Arabidopsis thaliana* [Vincenzetti *et al.*, 1999]. A outra forma da CDA são as homotetraméricas (T-CDA), as quais apresentam uma massa molecular, por subunidade, de aproximadamente 15kDa. As T-CDA também apresentam três resíduos conservados, três Cys, os quais exercem a mesma função que os

resíduos das D-CDA. Este tipo de CDA pode ser encontrado em organismos como, por exemplo, *B. subtilis*, humana e de *Mycobacterium tuberculosis* [Johansson *et al.*, 2002; Chung *et al.*, 2005; Sánchez-Quitian *et al.*, 2009].

Em 2009, Sanchez-Quitian e colaboradores elucidaram a estrutura tridimensional da CDA de *Mycobacterium tuberculosis* na sua forma *apo* (sem ligante), apenas com o íon de Zn^{+2} com uma resolução de 1.99 Å (**Figura 5**).

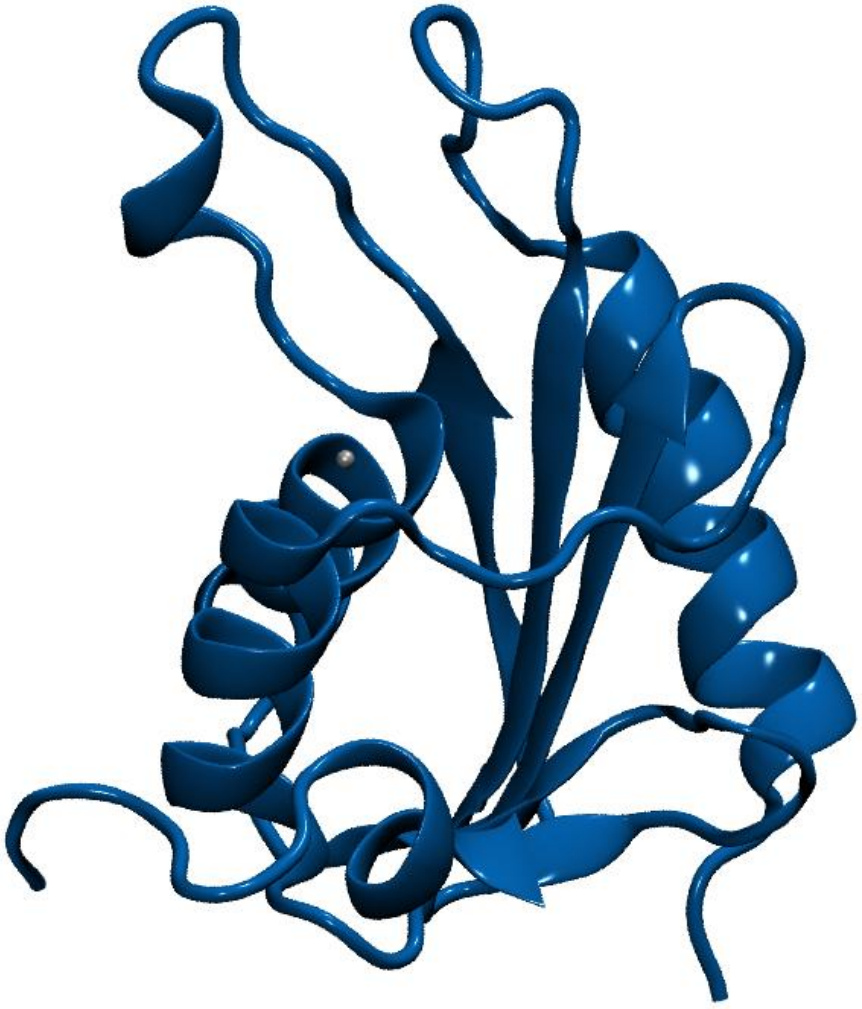


Figura 5. Estrutura tridimensional da CDA de *Mycobacterium tuberculosis* na forma apo. Em cinza está representado o íon de zinco (representação "*CPK*"). A imagem foi gerada com o VMD (Humphrey *et al.,* 1996).

A MtCDA faz parte da família das T-CDA, apresentando os três resíduos de Cys característicos para o ancoramento do íon de Zn^{+2}. O íon de Zn^{+2} está envolvido na orientação destes três resíduos de Cys, tendo um papel crucial na catálise da reação (**Figura 6**). Nesta estrutura também podem ser observadas duas tirosinas (Tyr) altamente conservadas nas sequências das T-CDA, as quais não estão presentes nas D-CDA. O enovelamento da MtCDA é caracterizado como $\alpha/\beta/\alpha$, sendo composta de cinco fitas-beta (Ala28-Val35, Arg39-Asn45, Arg72-Gly81, Leu104-Asp107 e Pro112-Leu115) e cinco hélices-alfa (Trp6-Ala18, Ser50-Thr54, Cys56-Thr68, Cys89-Gly100 e Leu115-Leu119) [Sánchez-Quitian *et al.*, 2009].

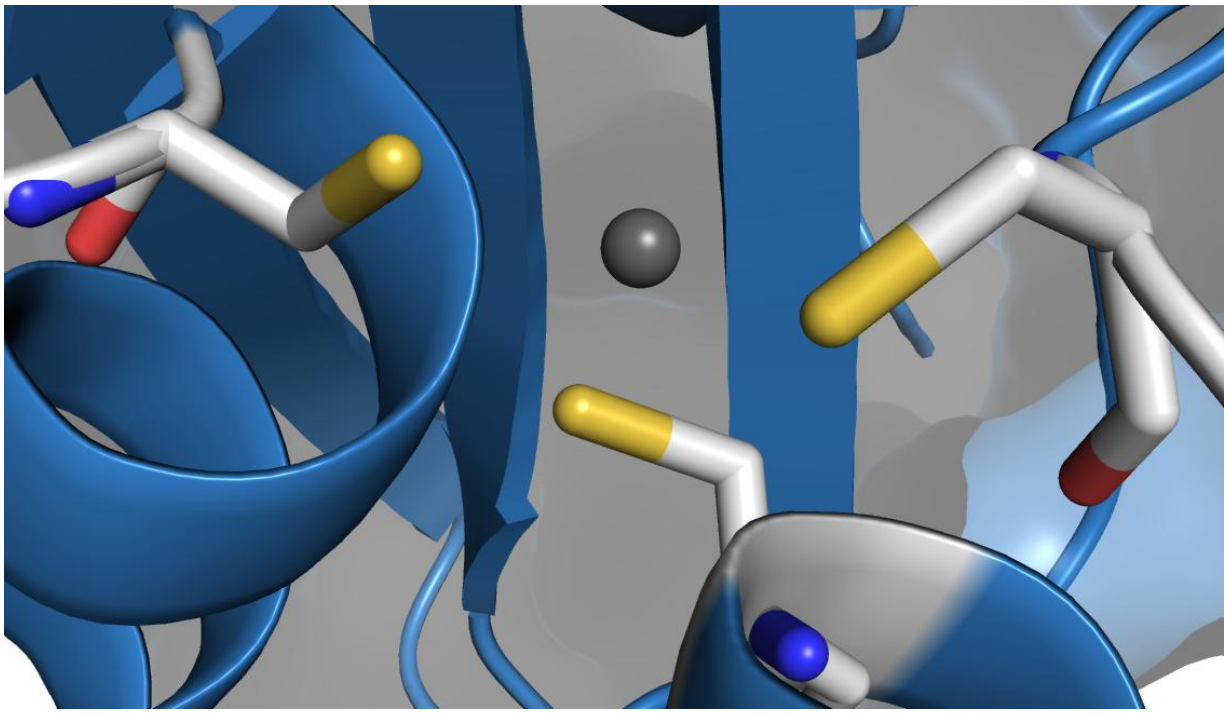


Figura 6. Figura ilustrando a região de ligação entre o íon de zinco, representado no centro (representação "*CPK*"), e as três cisteínas (representação "*sticks*") responsáveis pela estabilização do íon. Em azul claro está representada a enzima (representação "*new cartoon*"), assim como, a superfície molecular. A imagem foi gerada com o Pymol (DeLano, 2002).

O estudo das estruturas associadas com o substrato e produto da MtCDA, poderão aprimorar o conhecimento da interação receptor-ligante, fornecendo informações pivotais para a busca de potenciais alvos para o desenvolvimento de drogas e vacinas.

2. OBJETIVOS

2.1 Objetivo geral

Este estudo faz parte das pesquisas do Instituto Nacional de Ciências e Tecnologia em Tuberculose (INCT-TB), o qual tem como finalidade o desenvolvimento de drogas para tratar, vacinas para prevenir e ferramentas diagnósticas para detectar mais rapidamente o *Mtb* em pessoas infectadas. O objetivo geral deste trabalho é realizar o estudo estrutural da enzima Citidina Deaminase de *Mycobacterium tuberculosis* H37Rv, envolvida na via de salvamento de pirimidinas.

2.2 Objetivos específicos

- Realizar uma TVL, baseados em análogos de substrato, com o intuito de encontrar novos inibidores. As bibliotecas de análogos serão obtidas pelo *webserver* PubChem;
- Realizar simulações por DM dos complexos obtidos por cristalografia e dos três melhores ligantes referentes à triagem virtual, visando à melhor compreensão dos mecanismos de interação receptor-ligante.

3. JUSTIFICATIVA

É iminente a necessidade de novos agentes quimioterápicos para o tratamento da TB, os quais visem diminuir a toxicidade e a duração do tratamento atual. Estes são alguns dos problemas cruciais para o surgimento de cepas resistentes, como a MDR-TB e XDR-TB, às drogas comercialmente disponíveis. Outro agravante para o tratamento da TB são as interações medicamentosas, principalmente com as drogas anti-HIV. Portanto torna-se urgente a busca por novos alvos para o tratamento desta doença.

As enzimas envolvidas na biossíntese de nucleotídeos pirimídicos são alvos interessantes no desenvolvimento de novos agentes antimicrobianos, por estarem possivelmente envolvidos na fase de latência do bacilo. Além disso, alvos que possam agir nesta fase, apresentam grandes possibilidades de prevenir a progressão e reativação da doença.

A MtCDA, uma enzima evolutivamente conservada da rota de salvamento das pirimidinas, é um alvo atrativo para o desenho racional de drogas. Apesar desta enzima estar presente tanto no Mtb como em humanos, a possibilidade de realizar estudos de *chemical biology*, fazendo avaliações minuciosas dos sítios ativos de ambas as enzimas é uma abordagem viável e já comprovada em estudos anteriores [Ho *et al*., 2010]. Esta via é preferencialmente utilizada, pelo fato de demandar menos energia [Moffatt & Ashihara, 2002]. A obtenção das estruturas associadas com seus substratos e

produtos poderão contribuir para a compreensão do mecanismo de interação receptor-ligante, visando o desenho racional de drogas.

4. METODOLOGIA

4.1 Local de execução e duração

As etapas de refinamento dos dados, docagem molecular, TVL, simulações por dinâmica molecular (DM) serão realizados no Laboratório de Bioquímica Estrutural (LaBioQuest) da Faculdade de Biociências – PUCRS. Foi utlizada uma *workstation* Dell *Precision* T-3500 Intel *Xeon Quad Core*, para a realização das análises dos resultados obtidos.

O presente trabalho consiste em uma dissertação de mestrado com previsão de dois anos de duração.

4.2 Docagem molecular

Conceitualmente a docagem molecular pode ser definida como um procedimento computacional utilizado para tentar predizer a posição, orientação e conformação nativa de uma pequena molécula (ligante) associada dentro do sítio ativo de uma macromolécula alvo [Alonso *et al.*, 2006].

Pioneira na década de 80, a docagem molecular continua sendo até hoje uma das ferramentas mais utilizadas para o desenho de drogas *in silico* e é o componente principal em diversos programas de descobrimento de drogas [Zoete *et al.*, 2009]. A docagem pode ser superficialmente descrita como uma combinação de um algoritmo de busca, o qual pretende sugerir diversas possibilidades de conformações para um ligante, e uma função escore que objetiva a identificação do verdadeiro modo de ligação da molécula. Existem vários métodos para avaliar os resultados dos processos de docagem, como, por exemplo, algoritmo genético (AG) [Jones *et al.*, 1997], algoritmo genético

lamarckiano (AGL) [Morris *et al.*, 1998], minimização por Monte Carlo (MC) [Metropolis & Ulam, 1949; Totrov & Abagyan, 1997], similaridade baseada na superfície molecular [Jain, 2003], entre outros. Nesta dissertação foi dado enfoque ao AGL, o qual está implementado no programa AutoDock [Goodsell & Olson, 1990; Goodsell *et al*., 1996; Morris *et al*., 1996; Morris *et al*., 1998] que foi utilizado para fazer as simulações de docagem molecular.

4.2.1 Algoritmo Genético Lamarckiano

Diferentemente dos AG tradicionais, os quais mimetizam as principais características da evolução Dawiniana, aplicando também a genética Mendeliana, o AGL faz alusão à asserção (desacreditada) de Jean Batiste Lamarck, onde características fenotípicas adquiridas durante a vida de um indivíduo podem ser passadas hereditariamente [Morris *et al.*, 1998].

Como todos os AG, o AGL utiliza a idéia baseada na linguagem genética natural e na evolução biológica. Na docagem molecular, o arranjo do ligante em relação à proteína pode ser descrito por um conjunto de valores de translação, orientação e conformação. Esse conjunto é definido como os “estados variáveis”, para os AG estes correspondem a um “gene”. Os estados do ligante correspondem ao genótipo, às coordenadas atômicas são denominadas de fenótipo. Morris e colaboradores definem o “*fitness*”, na docagem molecular, como a energia total de interação de um ligante com a proteína alvo, a qual é mensurada por uma função de energia. Pares fortuitos de indivíduos são acoplados pelo processo de “*crossover*”, onde cada indivíduo herdará genes parentais. Além disso, alguns descendentes estarão sujeitos a “mutação” aleatória, onde um gene pode se modificar em níveis aleatórios. A “seleção”

dos descendentes da “geração” atual ocorre baseada no “*fitness*” de cada indivíduo. Tendo em vista estes dados, as melhores soluções (conformações de menor energia) são mantidas enquanto que as outras são descartadas.

A principal diferença entre o AGL e os outros AG, é a introdução de uma busca local por uma solução de menor energia. Esta busca é feita baseada no fenótipo para o genótipo, quando essa procura encontra uma solução de menor energia o indivíduo anterior é substituído. Enquanto que nos AG darwinianos esta busca é feita apenas do genótipo para o fenótipo. A **Figura 7** ilustra os dois tipos de AGs, evidenciando suas diferenças.

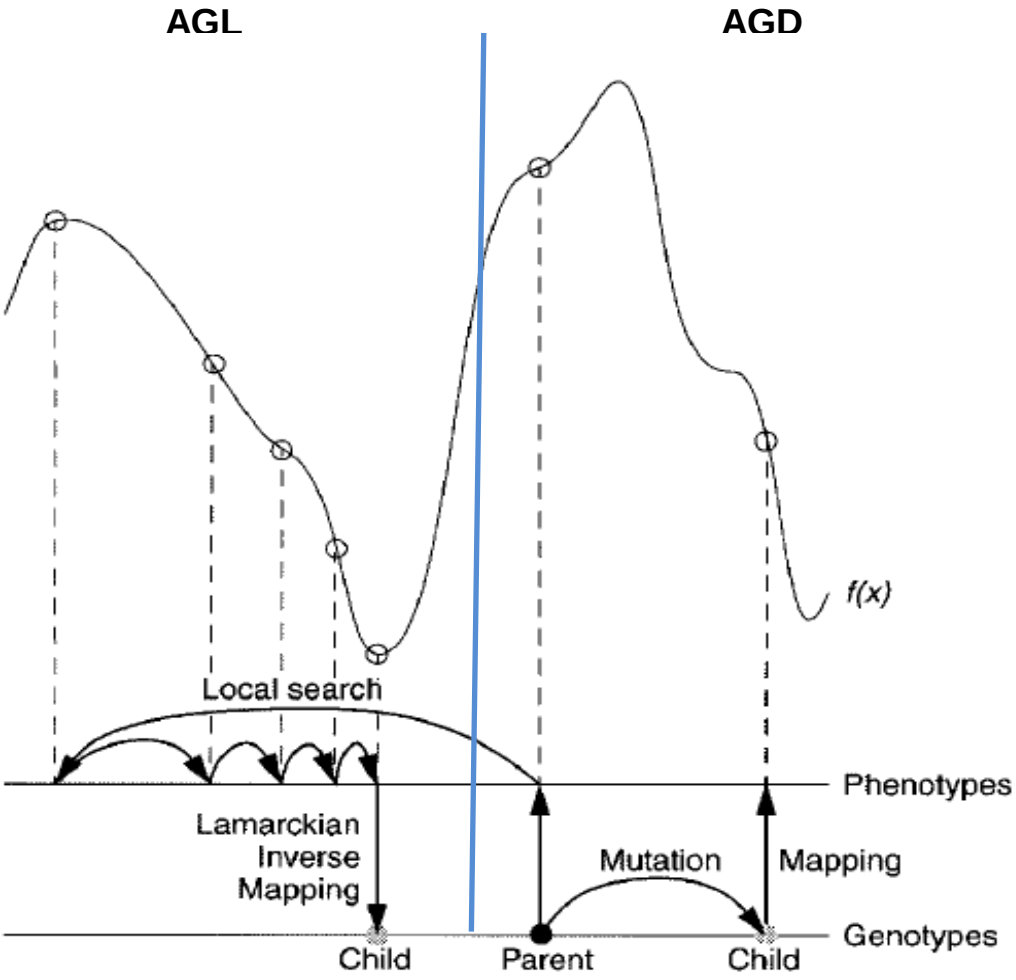


Figura 7. Esta figura ilustra como ocorrem as buscas por melhores soluções entre os dois tipos de AG, o AGL representado no lado esquerdo e o AGD no lado direito (figura adaptada do artigo Morris *et al.*, 1998).

4.3 Triagem virtual de ligantes

A Triagem virtual de ligantes (TVL) pode ser definida como, uma seleção de compostos que satisfazem características específicas utilizando ferramentas computacionais, como, por exemplo, a docagem molecular. A TVL permite a triagem *in silico* de vastas bibliotecas de pequenas moléculas para a seleção de potenciais inibidores. Basicamente, existem duas abordagens para a busca por inibidores *in silico*. A primeira abordagem é a “triagem virtual baseada no ligante”, utilizada normalmente quando a estrutura alvo (macromolécula) a ser estudada ainda não foi elucidada, a estratégia então a ser utilizada é a partir das informações de compostos conhecidos, que apresentem ação inibitória contra o alvo, identificar outros compostos que apresentem propriedades similares. A segunda abordagem é a “triagem virtual baseado no receptor”, a qual envolve explicitamente a docagem de cada ligante dentro de um sítio ativo conhecido, estabelecendo um modo de ligação para cada composto de um banco de dados. Essas informações servirão para selecionar os melhores ligantes, formando um pequeno conjunto de compostos, os quais poderão ser testados *in vitro*.

Apesar da técnica de docagem molecular apresentar-se uma abordagem rápida e de baixo custo computacional, seu ponto fraco está em não permitir que a molécula docada se ajuste no sítio ativo. Tendo isto em vista, é necessária a utilização de outras técnicas para que este problema seja amenizado. A Dinâmica Molecular (DM) pode tratar tanto o ligante quanto a proteína de forma flexível, permitindo desta forma uma melhor avaliação da interação ligante-receptor.

4.4 Dinâmica molecular

Simulações por DM é uma das técnicas mais versáteis para o estudo de macromoléculas biológicas, no âmbito das técnicas *in silico*. Conceitualmente a DM é uma abordagem computacional, na qual são empregadas equações Newtonianas para a resolução de representações atomísticas, de um sistema molecular baseado na Mecânica Clássica, com o intuito de obter informações a respeito de suas propriedades em função do tempo [Alonso *et al.*, 2006]. Os algoritmos utilizados nos programas de DM consistem da solução numérica destas equações de movimento fornecendo uma trajetória (coordenadas e momentos conjugados em função do tempo) do sistema em estudo.

Em 1977, McCammon e colaboradores realizaram a primeira simulação de DM envolvendo proteínas. Esta simulação foi realizada *in vacuo* e o tempo de simulação foi de 8,8 x 10^{-12} s [McCammon *et al.*, 1977]. Desde então, a técnica de DM foi se aprimorando, deixando os sistemas construídos *in silico* mais realísticos. Este progresso deve-se tanto a avanços na área da química, com o melhoramento dos parâmetros dos campos de força, quanto da computação, com o desenvolvimento de máquinas mais robustas, o que permite a realização de simulações mais longas, chegando a 10^{-9} e 10^{-8} s.

A DM é empregada em várias áreas, desde o refinamento de estruturas cristalográficas, predição de estruturas protéicas, otimização de parâmetros geométricos, avaliação da interação ligante-receptor, entre outras.

Como mencionado, o problema da técnica de docagem molecular, utilizada para TVL, é não permitir que ocorra um ajuste na associação receptor-ligante sem que o custo computacional fique extremamente oneroso. Com isso, a TVL serve como um filtro, para que os melhores compostos sejam

submetidos à técnica de DM, podendo assim ter uma avaliação mais realista do modo de associação do complexo. Permitindo também que seja calculada a energia livre de ligação de forma acurada, aprimorando o processo de descobrimento de drogas.

Nesta dissertação a DM foi empregada para o refinamento dos melhores resultados da TVL e também para a melhor compreensão dos mecanismos de interação entre proteína-substrato e proteína-produto. Para isso, foi utilizado o pacote de programas GROMACS [van der Spoel *et al.,* 2005], versão do GROMACS 4.0.5 disponível como *freeware* sob licença GNU (*General Public License*).

5. ARTIGO CIENTÍFICO

J Mol Model
DOI 10.1007/s00894-011-1045-0

ORIGINAL PAPER

Combining molecular dynamics and docking simulations of the cytidine deaminase from *Mycobacterium tuberculosis* H37Rv

Luís Fernando Saraiva Macedo Timmers · Rodrigo Gay Ducati · Zilpa Adriana Sánchez-Quitian · Luiz Augusto Basso · Diógenes Santiago Santos · Walter Filgueira de Azevedo Jr.

Received: 22 December 2010 / Accepted: 16 March 2011

Abstract Cytidine Deaminase (*CD*) is an evolutionarily conserved enzyme that participates in the pyrimidine salvage pathway recycling cytidine and deoxycytidine into uridine and deoxyuridine, respectively. Here, our goal is to apply computational techniques in the pursuit of potential inhibitors of *Mycobacterium tuberculosis CD* (*MtCDA*) enzyme activity. Molecular docking simulation was applied to find the possible hit compounds. Molecular dynamics simulations were also carried out to investigate the physically relevant motions involved in the protein-ligand recognition process, aiming at providing estimates for free energy of binding. The proposed approach was capable of identifying a potential inhibitor, which was experimentally confirmed by IC_{50} evaluation. Our findings open up the possibility to extend this protocol to different databases in order to find new potential inhibitors for promising targets based on a rational drug design process.

Keywords Free energy of binding · IC_{50} determination · Molecular docking simulation · Molecular dynamics simulation

L. F. S. M. Timmers · W. F. de Azevedo Jr.
Faculdade de Biociências, Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Laboratório de Bioquímica Estrutural (LaBioQuest), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS),
Av. Ipiranga 6681,
Porto Alegre, RS 90619–900, Brazil

L. F. S. M. Timmers · Z. A. Sánchez-Quitian · L. A. Basso · D. S. Santos (✉) · W. F. de Azevedo Jr. (✉)
Programa de Pós Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil
e-mail: diogenes@pucrs.br

W.F. Azevedo
e-mail: walter@azevedolab.net

R. G. Ducati · Z. A. Sánchez-Quitian · L. A. Basso · D. S. Santos
Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF), Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS),
Av. Ipiranga 6681,
Porto Alegre, RS 90619–900, Brazil

Introduction

Mycobacterium tuberculosis (*Mt*), the causative agent of tuberculosis (TB), is one of the main causes of human death due to a single infectious agent. According to the World Health Organization (WHO) approximately two billion people (one third of the world population) are infected with *Mt*, and, based on the last WHO report, there was an estimation of 9.27 million new TB cases in 2007 [1–4]. TB is considered a neglected disease, as the majority of the cases are reported in emergent countries. However, since the 1980s, TB has returned to Europe and to the United States of America. The resurgence of TB in developed countries has been tightly related to the emergence of co-infection with the Human Immunodeficiency Virus (HIV/AIDS) [5]. Thereby, the TB/HIV synergism has become a challenge for pharmaceutical companies.

New anti-TB drugs should, ideally, present the following characteristics: lower toxicity, decrease the length of the treatment thereby improving patients' compliance, and be effective against the latent form of *Mt* allowing prevention of reactivation of the disease. These criteria have been used in structural bioinformatics approaches, mainly when the

Published online: 04 May 2011

aim is to find potential compounds that have activity against specific protein targets.

In order to find novel targets to pursuit new inhibitors against *Mt*, the pyrimidine metabolism appears to be an interesting alternative. Pyrimidine nucleotides have a pivotal role in many cellular functions, such as DNA replication and RNA transcription. These functions are regulated by the two major routes, the de novo and the salvage pathway [6]. The de novo pathway of pyrimidine biosynthesis is composed of six enzymes, responsible for the conversion of glutamine, by a series of steps, into uridine-5'-monophosphate (UMP). This set of reactions is characterized by a demand of higher amounts of energy when compared to the salvage pathway. On the other hand, the pyrimidine salvage pathway allows organisms to make use of exogenous pyrimidine bases and nucleosides that are not intermediates in the de novo pyrimidine synthesis [7–10]. One of the key enzymes in the pyrimidine salvage pathway is the Cytidine Deaminase (*CD*). *CD* is an evolutionarily conserved enzyme that participates in the recycling of cytidine and deoxycytidine into uridine and deoxyuridine, respectively (Fig. 1). The deamination reaction is coordinated by a zinc atom, which activates a water/hydroxide molecule in the initial hydrolytic attack on the C4 of the substrate. The *MtCDA* in the free form, was previously described as a homotetramer [11], as in *Bacillus subtilis*. Sang Chung and collaborators observed that, for a proper recognition of the inhibitor, the human *CDA* homologue's active site is made up of residues from three or four subunits [12]. Accordingly, detailed analysis of the enzyme quaternary structure is pivotal to rational-based drug design.

In the present work, our goal was to apply computational techniques in the pursuit of potential inhibitors of *MtCDA* enzyme activity. Molecular docking simulation was employed to both find hit compounds and to rank the best fit of the ligands. Molecular dynamics (MD) simulations were performed in two steps: First, to verify the stability of the quaternary structure previously published [11], comparing it with the structure generated by the symmetry operators. Second, extract physically relevant motions that are essential and presumably meaningful for the protein-ligands recognition processes in the binding pocket, using *principal component analysis* (PCA). We have also calculated the free energy of binding to correlate in silico and in vitro results.

Materials and methods

Validation of the molecular docking simulation approach

Molecular docking simulation was performed to find potential inhibitors of *MtCDA* enzyme activity. Prior to this, however, the search algorithm and the scoring function, implemented in the AutoDock 4.0.2, were tested in order to confirm that this program is capable of ranking the best potential inhibitors. Initially, a redocking was carried out with the structure 3LQP (PDB access code) aiming at verifying whether the program could reproduce the ligand location found in the crystallographic structure. Then, we tested the ability of the program to identify, qualitatively, the best inhibitors previously described [12–14]. Considering that there is no inhibitor described for *MtCDA*, we used the human *CDA* (PDB access code: 1MQ0) [12]. Table 1 presents the result for the qualitative validation of the human *CDA* based on the free energy of binding calculated with the AutoDock 4.0.2 and the constants determined in vitro.

Database construction

The database consists of 91 compounds, where 84 molecules are analogs of the substrates, cytidine and

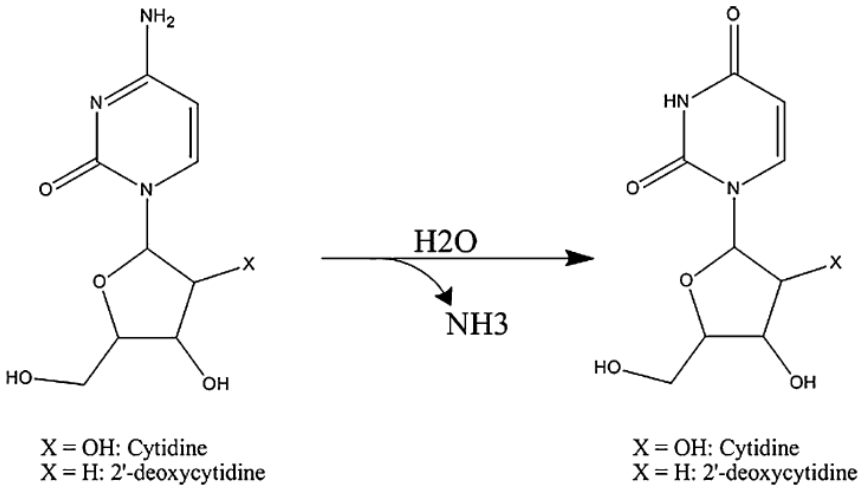


Fig. 1 *MtCDA* chemical reaction

Table 1 The result for the qualitative validation of the human *CDA* based on the free energy of binding calculated with the AutoDock 4.0.2 and the constants determined *in vitro*

Inhibitor	$K_{i(mM)}$	$*\Delta G_{exp}$	ΔG_{calc}
Diazepinone	2.5×10^{-8}	-10.78	-6.41
Tetrahyuridine	9.5×10^{-7}	-8.54	-5.81
Zebularine	6.7×10^{-6}	-7.34	-5.79
CMP	0.39×10^{-3}	-4.83	-5.15

$*\Delta G = R \times T \times \ln K_i$

deoxycytidine. Among these analogs, 49 were obtained from Pubchem, 25 from DrugBank [15], and 10 from ZINC [16] databases. Tanimoto coefficient cuttof of 90% was used to search the analogs. The other seven compounds were constructed based on modifications of the substrates and products, where the amine and hydroxyl groups were substituted by methyl, and sulfur, and the pentose ring was modified for an aliphatic chain.

Molecular docking simulation

The flexible docking simulations were performed using PyrX-0.5, where the AutoDock 4.0.2 is implemented. The AutoDock software uses an empirical scoring function based on the free energy of binding. Among the stochastic search algorithms offered by the AutoDock suite, we chose the Lamarckian genetic algorithm (LGA) that is a hybrid approach [17], which combines genetic algorithm (as global search) [18] and Solis and West algorithm (as local search) [19].

The *MtCDA* structure used for the simulations was 3LQP (PDB access code). A grid box was created with 36×31×38 points and a resolution of 0.375 Å in order to include solely the protein's active site, aiming to reduce the computational cost. The molecular docking process was carried out with 100 independent runs for each docking simulation, an initial population of 150, a maximum number of 500,000 energy evaluation, and a maximum number of 27,000 generations. Mutation and crossover were applied to the population at rates 0.02 and 0.80, respectively.

Molecular dynamics simulations

All MD simulations were carried out using the GROMACS 4.0.5 package [20] with the 53a6 GROMOS force field. We performed MD simulations for the enzyme in the free form (PDB access code: 3IJF), with the products (uridine and deoxyuridine - PDB access code: 3LQP and 3LQT, respectively), and the ten best results obtained from the molecular docking simulations, in order to investigate the dynamic nature of the interactions between the protein-ligand complexes, and water molecules.

The MD simulations were carried out using the particle mesh Ewald method [21] for the electrostatic interactions. The van der Walls and Coulomb cutoff were 14 Å and 10 Å, respectively. The integration time step was 2.0 fs, with the neighbor list being updated every ten steps by using the grid option and a cutoff distance of 12 Å. The simple point charge extended (SPC/E) [22] water model was used. Periodic boundary condition has been used with constant number of particles, pressure, and temperature (NPT) in the system. The V-rescale thermostat was applied using a coupling time of 0.1 ps to maintain the systems at a constant temperature of 298.15 K. The Berendsen barostat was used to maintain the systems at a pressure of 1 bar, and values of the isothermal compressibility were set to 4.5×10^{-5} bar^{-1} for water simulations. The temperature of the systems were increased from 50 to 300 K in five steps (50–100, 100–150, 150–200, 200–250, and 250–300 K), and the velocities at each step were reassigned according to the Maxwell-Boltzmann distribution at that temperature and equilibrated for 10 ps, except for the last part of the thermalization phase, which was of 40 ps. The systems were submitted to a steepest descent followed by conjugated gradient energy minimizations up to a tolerance of 1000 kJ mol^{-1}. An MD simulation with position restraints was carried for a period of 20 ps in order to allow the accommodation of the water molecules in the system. Finally, 10 ns MD simulations were performed to all systems. The topologies files and other force field parameters, except the charges of ligands, were generated using the PRODRG program [23]. The partial atomic charges to the ligands were calculated using the Gaussian03 package [24], being submitted to single-point ab initio calculations at DFT/B3LYP/6-311 G (2d, p) level in order to obtain ESP charges.

Principal component analysis

Essential dynamics (ED), also known as PCA, is an interesting method for the identification of the main conformational changes, which often have importance in biological process for a protein during MD simulation. The ED analysis is a technique that reduces the complexity of the data and extracts the concerted motion in simulations that are essentially correlated and presumably meaningful for biological function [25]. In the ED analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. A set of eigenvectors and eigenvalues was identified by diagonalizing the matrix. The eigenvalues represented the amplitude of the eigenvectors along the multidimensional space, and the displacement of atoms along each eigenvector showed the concerted motions of protein along each direction. An

Table 2 The free energy of binding calculated in silico compared to experimentally determined

Inhibitors	Lennard-Jones^{vdW}	Coulomb^{el}	ΔG_{bind}	ΔG_{bind} Exp
Zebularine	-22,972	-94,240	-8,954	-8,541
Dihydrouridine	-35,051	-23,660	-6,377	-6,237
Fluorozebularine	-30,946	-25,720	-5,923	-9,251
Diazepinone	-60,552	-54,550	-11,846	-10,782
Tetrahydrouridine	-40,392	-9,130	-6,249	-7,338

*The ΔG_{bind} values are represented in *Kcal*

assumption of ED analysis is that the correlated motions for the function of the protein are described by eigenvectors with large eigenvalues. The movements of protein in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis.

Free energy of binding calculation and determination of the α and β coefficients

The linear interaction energy approximation (LIE) method [26] was employed to calculate the free energy of binding to all protein-ligand complexes. The LIE method is a semiempirical approach that combines advantages of free energy perturbation (FEP) and thermodynamics integration (TI) in order to obtain the free energy of binding. Often, two MD simulations are required to calculate the free energy of binding with the LIE method: one for the free ligand in solution and another for the protein-ligand in solution. The LIE equation is as follows:

$$\Delta G_{bind} \approx \alpha \Delta \left\langle V_{l-s}^{vdW} \right\rangle + \beta \Delta \left\langle V_{l-s}^{el} \right\rangle + \gamma, \quad (1)$$

where ⟨⟩ denotes MD averages of the non-bonded van der Walls (*vdW*) and electrostatic (*el*) interactions between the ligand and its surrounding environment (*l-s*). The α and β values are coefficients that represent nonpolar and polar contributions, respectively, and γ is a constant term.

We performed four simulations for each protein-ligand complex: two simulations with the partial atomic charges of the ligand, as described above, and another two simulations without charges to the ligands in order to obtain the long-range electrostatic values. Before initiating the LIE calculations with the ten best docking results, we calibrated the coefficients based on the structure of the human *CDA* (PDB access code: 1MQ0) [12], which has inhibitory constants experimentally calculated for five different inhibitors. The coefficients applied to the LIE calculations were α=0.59, β=0.43, and γ=0, which obtained a correlation coefficient of 0.73 between experimental and calculated values, and the average to the estimated error was 13.29% or $\pm 3 kcal \times mol^{-1}$ (Table 2), similar to that considered in the AutoDock [17].

IC_{50} determination

In vitro studies with the human *CDA* have shown that tetrahydrouridine inhibits the enzyme-catalyzed chemical reaction with a K_i value of 6.7 μM [27]. In the present study, attempts have been made to investigate the level of tetrahydrouridine inhibition of *MtCDA* enzyme activity in vitro. The evaluation of IC_{50} value, concentration of inhibitor required to effect 50 % reduction en enzyme activity, is a good approach to assess relative inhibitor potency. Thereby, tetrahydrouridine IC_{50} for *MtCDA* was determined by measuring initial rates at fixed non-saturating levels of cytidine (K_M=1004 μM [11]) in either absence or presence of tetrahydrouridine (0–1400 μM) in the reaction mixture, being the maximal reaction rate condition determined in the absence of inhibitor. It should be pointed that each individual initial rate datum was the average of duplicate or triplicate measurements. The IC_{50} value, obtained by fitting the data to the appropriate equation (Eq. 2) using SigmaPlot 2004 (Systat Software, Inc.), defines the concentration of inhibitor required to half-saturate the enzyme population; v_i and v_o are, respectively, the reaction velocity in the presence and absence of inhibitor, and v_i/v_o represents the fractional activity remaining at a given inhibitor concentration (fraction of free enzyme) [28].

$$\frac{v_i}{v_o} = \frac{1}{1 + \left(\frac{[I]}{IC_{50}}\right)} \quad (2)$$

Table 3 The minimum, mean, maximum, and standard deviation of the RMSD and RG between structures A and B

Analysis		Structure A	Structure B
RMSD	Min	0.001202	0.001248
	Max	0.501063	1.18391
	Mean	0.380708	0.605804
	Stdev	0.0680289	0.236939
RG	Min	2.12377	2.48344
	Max	2.3302	3.15195
	Mean	2.29459	2.80797
	Stdev	0.024201	0.150226

*The values are represented in "*nm*"

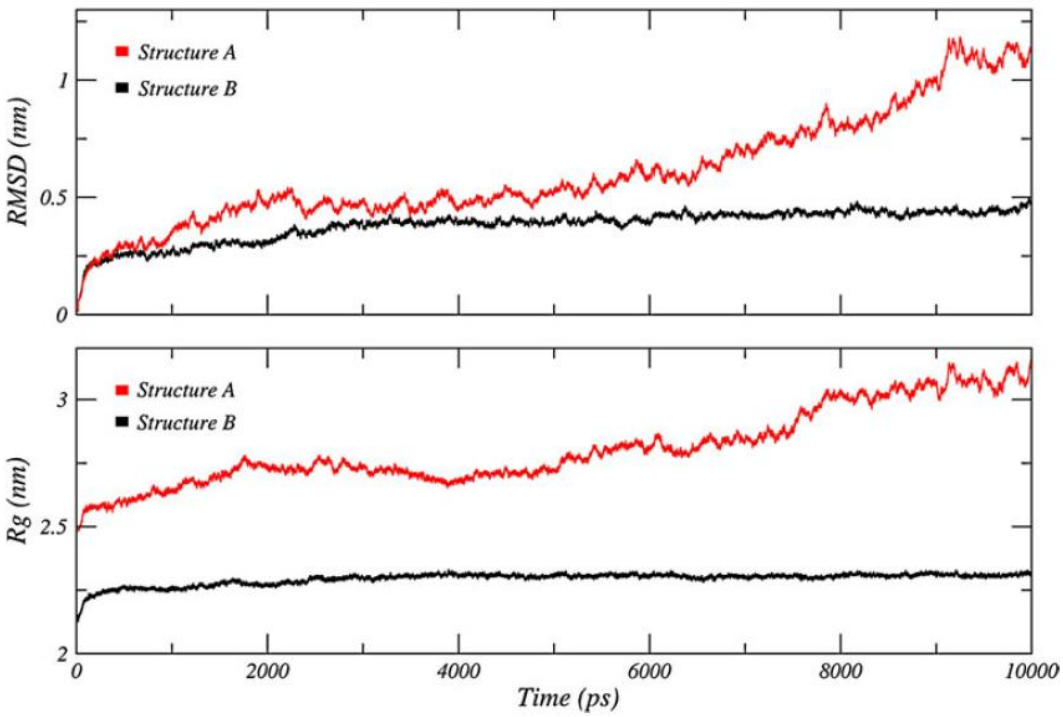


Fig. 2 Graphical representation of RMSD and RG of each *MtCDA* tetrameric structure (**a** and **b**) as a function of time

Results and discussion

Stability of the quaternary structure of the MtCDA

In the first work that elucidated the crystal structure of the *MtCDA* [11], it was identified that the quaternary structure was formed by two-fold rotational axes against atomic coordinates of the content of the asymmetric unit. However, alternative quaternary structure is also possible [11]. In order to verify the stability of the structures in solution, we analyzed the root mean square deviation (RMSD) and the radius of gyration (RG) from both structures. Structure A is the quaternary packaging previously identified and structure B is the alternative one. Taking into account the values observed for the RMSD and RG of the structures present in Table 3 and Fig. 2, we can suggest that structure A does not have a reasonable stability in solution when compared to structure B. Furthermore, analysis of the first eigenvectors represent 69% and 51.9% of the crucial motions during the MD simulations to the structures A and B, respectively. Thereby, the main motions of the structure A are correlated with the dissociation of the quaternary structure, differently from that observed in the structure B (Fig. 3). According to these results, we chose the structure B as the most probable structural packaging for *MtCDA*. A visual inspection of the homotetrameric *CD* from other organisms corroborates with these results, as they present the same quaternary structure found in the present MD simulation. Considering the highest conserved interface regions, which characterize the homotetrameric *CD*, the residues Tyr24, Phe27, Tyr51, Cys59, Arg94, Gln95, Leu119, and Phe123 were analyzed, which correspond to the human *CDA* Phe36, Tyr60, Arg68, Arg103, Gln104, Leu133, and Phe137 residues. These residues represent the most important protein-protein interactions between

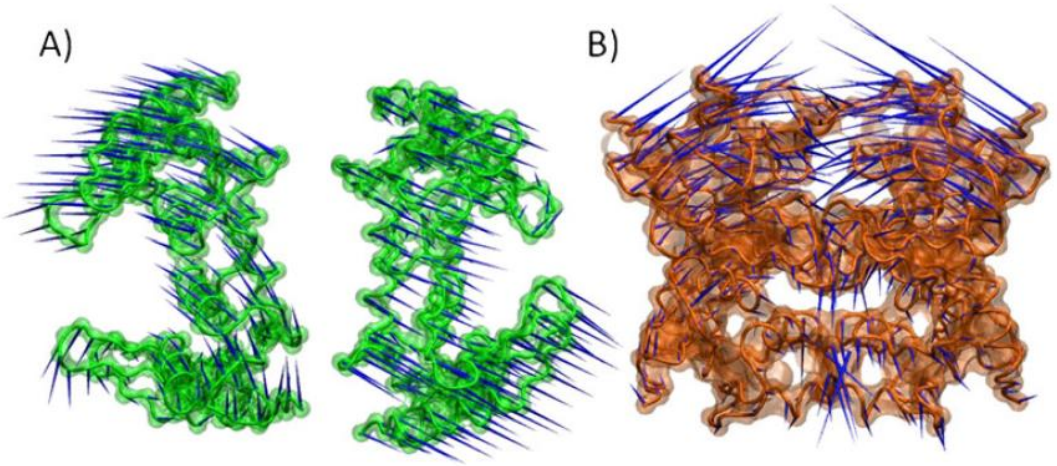
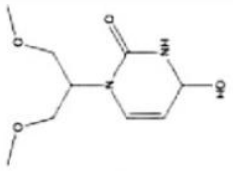
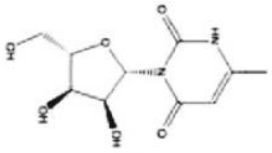
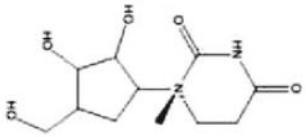
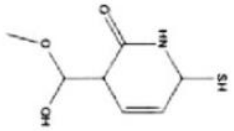
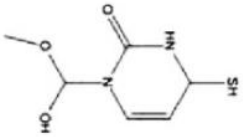
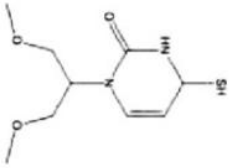
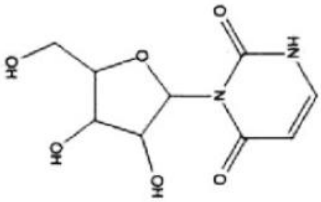
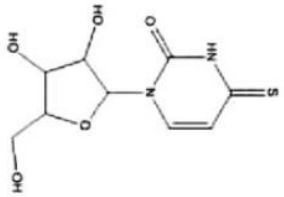
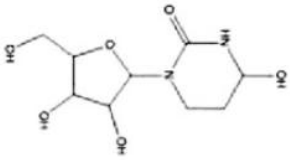
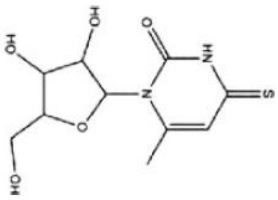


Fig. 3 Displacements of PCA model of the structure **a** (previously published) and **b** (proposed structure)

Table 4 The ten best molecules, which were obtained from molecular docking simulations

Id	Structure	Lipinski's rule of five				ΔG*
		HB_{donor}	$HB_{acceptor}$	MW	LogP	
Compound 1		1	5	245,32	0,593	-6,2
Compound 2		4	8	260,25	0,394	-9,2
Compound 3		5	8	263,27	-4,98	-8,6
Compound 4		1	4	219,31	0,534	-8,8
Compound 5		1	5	220,29	0,309	-8,8
Compound 6		-	4	261,39	1,546	-6,6
Compound 7		4	8	246,22	-1,16	-8,1
Compound 8		4	7	264,3	-0,19	-8,1
Compound 9		5	8	248,24	-1,14	-7,7
Compound 10		4	5	274,29	-1,3	-8,9

*The values are represented in *kcal*

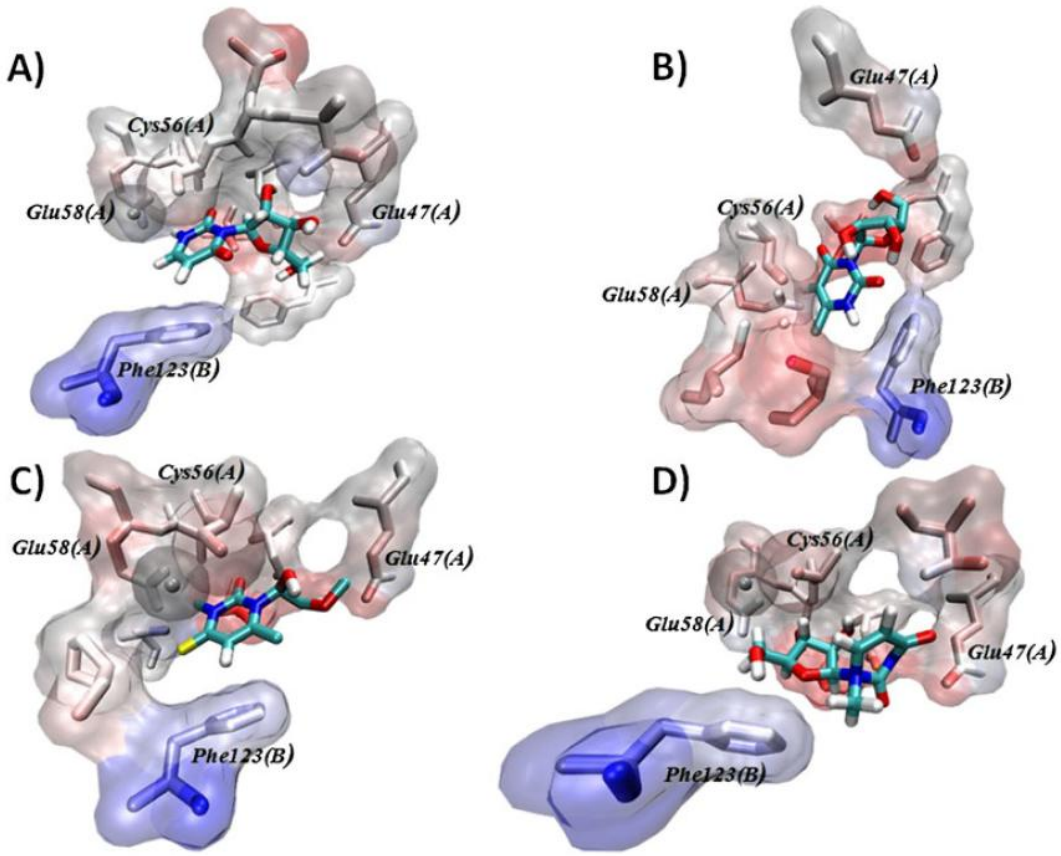


Fig. 4 The four best molecules encountered by LGA. (**a**) Compound 7, (**b**) Compound 2, (**c**) Compound 5, and (**d**) Compound 3

Table 5 The residues making interactions with each of the ten molecules, and the other five crystal structures interacting with their respective ligands

Ligands	HB	vdW	Ligands	HB	vdW	Ligands	HB	vdW
Comp_1		Val29(A)	*Comp_2*		Val29(A)	*Comp_3*		Val29(A)
		Ser50(D)		Phe27(A)	Cys56(A)		Ser50(A)	Asn45(A)
	Asn45(A)	Tyr51(D)		Glu47(A)	Glu58(A)		Thr54(A)	Glu47(A)
	Glu47(A)	Gly52(D)		Ser50(D)	Pro88(A)		Ala57(A)	Tyr51(D)
	Cys56(A)	Ala57(A)		Tyr51(D)	Cys89(A)		Glu58(A)	Cys56(A)
		Glu58(A)		Ala57(A)	Cys92(A)			Phe123(B)
		Cys89(A)			Phe123(B)			
		Phe123(B)						
Comp_4		Val29(A)	*Comp_5*		Val29(A)	*Comp_6*		Val29(A)
		Asn45(A)			Asn45(A)			Glu47(A)
	Cys56(A)	Glu47(A)		Cys56(A)	Glu47(A)		Asn45(A)	Ser50(D)
	Ala57(A)	Ser50(D)		Ala57(A)	Ser50(D)		Gly52(D)	Thr54(A)
		Tyr51(D)			Tyr51(D)		Zn	Leu55(A)
		Glu58(A)			Glu58(A)			Cys56(A)
		Cys78(A)			Cys78(A)			Glu58(A)
		Pro88(A)			Pro88(A)			Phe123(B)
		Phe123(B)			Phe123(B)			
Comp_7	Ser25(A)		*Comp_8*			*Comp_9*		
	Phe27(A)	Val29(A)		Ser25(A)				Ser25(A)
	Glu(47)	Asn45(A)		Phe27(A)	Ser50(D)			Phe27(A)
	Ser50(A)	Cys56(A)		Glu47(A)	Glu58(A)			Val29(A)
	Thr54(A)	Phe123(B)		Val49(A)	Pro88(A)			Asn45(A)

Table 5 (continued)

Ligands	HB	vdW	Ligands	HB	vdW	Ligands	HB	vdW
	Leu55(A)			Cys56(A)	Phe123(B)			Glu47(A)
	Ala57(A)			Cys89(A)				Cys56(A)
	Glu58(A)			Zn				Cys89(A)
								Phe123(B)
1JTK			***1ZAB***			***2FR5***	Asn54(A)	
	Asn42(A)			Asn54(A)	Val38(A)		Tyr60©	Phe36(A)
	Tyr48(A)	Val26(D)		Tyr60(C)	Cys59(.C)		Cys65(A)	Val38(A)
	Ala54(D)	Cys53(D)		Ala66(A)	Glu67(A)		Ala66(A)	Cys59(.C)
	Glu55(D)	Phe125(B)		Cys65(A)	Phe137(B)		Glu67(A)	Phe137(B)
	Cys86(D)						Cys99(A)	
							Cys102(A)	
2FR6	Asn54(A)		***1MQ0***	Asn54(A)				
	Cys65(A)	Val38(A)		Tyr60(B)	Val38(A)			
	Ala66(A)	Phe137(B)		Cys65(A)	Cys59(B)			
	Glu67(A)			Ala66(A)				
	Ser97(A)			Glu67(A)				

Table 6 The main residues which are making interactions with each protein-ligand complexes during the simulation

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts	
Compound 1	***1ns***		Val22	Glu58
		Asn45	Phe27	Cys89
		Glu47	Val29	Gly124(B)
			Cys56	Leu125(B)
			Ala57	Tyr51(D)
	2ns		Val22	Phe123(B)
			Phe27	Gly124(B)
		Asn45	Val29	Leu125(B)
		Glu47	Cys56	Val49(D)
			Ala57	Ser50(B)
			Glu58	Tyr51(B)
			Cys89	
Compound 2	***1ns***		Phe27	
		Ser25	Val29	
		Asn45	Glu47	
		Thr54	Cys56	
			Tyr51(D)	
	2ns		Phe27	
		Ser25	Val29	
		Asn45	Cys56	
		Thr54	Leu86	
		Tyr51(D)	Asp126(B)	
Compound 3	***1ns***	Glu47	Phe27	Tyr51(D)
		Cys56	Pro88	
		Glu58	Cys92	
		Met87	Phe123(B)	
		Cys89	Ser50(D)	

Table 6 (continued)

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts			
	2ns		Ser25			
			Phe27			
		Glu58	Val29			
		Leu86	Cys56			
		Arg91	Pro88			
		Cys92	Cys89			
			Phe123(B)			
Compound 4	***1ns***		Phe27		Thr54	
			Val29		Cys56	
			Asn45		Asp124(B)	
			Val46		Leu125(B)	
			Glu47		Tyr51(D)	
	2ns		Ser25		Tyr51	
			Phe27		Thr54	
			Val29		Leu55	
			Val46		Leu125(B)	
			Glu47		Tyr51(D)	
Compound 5	***1ns***		Ser25		Leu125(B)	
		Asp124(B)	Arg26		Leu128(B)	
		Tyr51(D)	Phe27			
			Ala57			
			Phe123(B)			
	2ns		Ser25			
		Asp126(B)	Phe27			
		Tyr51(D)	Ala57			
			Leu125(B)			
			Leu128(B)			
Compound 6	***1ns***		Ser25	Glu58	Leu125(B)	
			Phe27	Leu86	Ser50(D)	
		Ala57	Val29	Met87	Tyr51(D)	
			Glu47		Cys89	
			Cys56		Phe123(B)	
	2ns		Val22	Ser25	Glu58	Leu125(B)
			Phe27	Leu86	Cys92	Ser50(D)
		Ala57	Val29	Met87	Tyr51(D)	Gly127(B)
			Glu47		Cys78	
			Cys56		Phe123(B)	
Compound 7	***1ns***		Phe27		Phe123(B)	
		Arg91	Val29		Gly127(B)	
		Val49(D)	Glu47		Leu128(B)	
			Thr54			
			Cys56			
	2ns	Ser25	Phe27		Gly127(B)	
		Cys56	Val29		Leu128(B)	
		Arg91	Glu47			
		Phe123(B)	Thr54			
		Val49(D)				
Compound 8	***1ns***		Phe27	Leu86	Cys89	Ser50(D)
		Ser25	Val29	Met87	Leu128(B)	

Table 6 (continued)

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts		
		Asn45	Ala57	Cys78	Pro88
		Glu58	Cys56		Phe123(B)
		Tyr51(D)	Leu125(B)		
	2ns	Ser25	Ala57	Cys78	Pro88
		Glu58	Phe27		Leu86
		Cys87	Val29		Met87
		Tyr51(D)	Cys56		Phe123(B)
			Leu125(B)		
Compound 9	***1ns***		Ser25		Phe27
		Glu58	Val29		Cys56
		Val49(D)	Ala57		Leu86
			Leu128(B)		
			Ser50(D)		
	2ns		Ser25		Phe27
		Ala57	Val29		Cys56
		Cys78	Glu58		
		Ser50(D)	Leu86		
			Tyr51(D)		
Compound 10	***1ns***		Val22		Phe27
		Cys56	Val29		Val79
		Cys78	Asp80		Ser84
		Leu86	Val85		
		Tyr51(D)	Phe123(B)		
	2ns		Val22		Phe27
		Cys56	Val29		Val79
		Cys78	Ala57		Ser84
		Asp80	Leu86		Phe123(B)
		Tyr51(D)	Leu124(B)		

the subunits in order to maintain the quaternary structure. Taking these into account, we observed that Arg68, which is described as an important residue in the catalytic event in the human *CDA* [30], compensating the negative charges of the cysteines, corresponds to a cysteine in the *MtCDA*. Nevertheless, analyzing the *MtCDA* structure, it could be suggested that the residue responsible for counterbalancing the charges of the cysteines is Arg91, which is not conserved in the human *CDA*, where it corresponds to Ala101.

Molecular docking simulation

MtCDA molecular docking simulation was carried out by AutoDock 4.0.2, and all the docking parameters selected are described above. Before the molecular docking process, the partial atomic charges of the ligands were assigned according to the Gasteiger-Hückel [29, 30] method implemented in AutoDock. In order to obtain molecules that have better binding affinities than the substrates and products, we also perform the molecular docking simulation with the substrates and products (cytidine, deoxycytidine, uridine, and deoxyuridine). In this view, only the ligands that presented the estimated free energy of binding $3 kcal \times mol^{-1}$ lower than substrates and products were considered as potential inhibitors. Besides analysis of the free energy, all 91 compounds were subjected to visual inspection, considering the binding mode and their similarity with the substrates and/or products. Taking into account these filters, ten molecules that may have a potential activity as inhibitors for *MtCDA* were selected. Table 4 presents these molecules, which were ordered by the estimated free energy with AutoDock. When analyzed the Lipinski's rule of five, all ten compounds were in agreement to the main

Table 7 The hit compounds free energy of binding calculated with the LIE equation

Inhibitors	Lennard-Jones^{vdW}	Coulomb^{el}	ΔG_{bind} (Kcal)
Compound 1	-52.71	19.10	-6.27
Compound 2	-52.70	-36.99	-9.67
Compound 3	-40.79	-53.39	-8.99
Compound 4	-55.28	-36.37	-10.00
Compound 5	-43.77	-21.49	-7.48
Compound 6	-59.80	-62.09	-12.20
Compound 7	-38.12	-90.82	-10.88
Compound 8	-48.86	-153.55	-16.20
Compound 9	-45.73	9.27	-5.89
Compound 10	-78.79	-107.11	-17.61

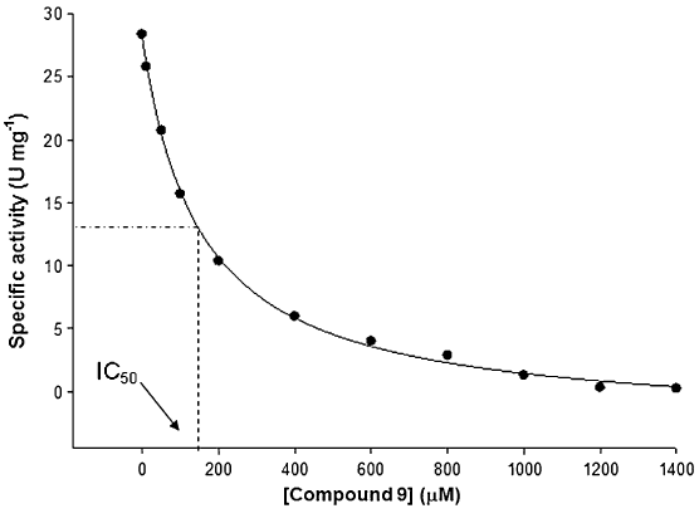


Fig. 5 Inhibitory effect of compound 9

criteria, such as hydrogen bonds, molecular weight, and ClogP. Figure 4 presents the four best molecules encountered by LGA. Interestingly, all ten best molecules, plus the substrates and/or products, presented five main residues making important interactions (hydrophobic or electrostatic): Glu47(A), Cys56(A), Glu58(A), Phe123(B), and Tyr51(D). In addition, the majority of the compounds were coordinated by the zinc ion, similarly to substrates and/or products bound to the binding pocket. By analyzing the distribution of the hydrogen bonds among the hit compounds, we observed that the major contacts are associated in the stability of the ribose moiety. On the other hand, the pyrimidine moiety also presents considerable hydrogen bond interactions, so its importance should not be neglected. Therewithal, the core contribution to the stability of the ligands into the binding cavity is driven by van der Walls interactions, which mostly involve the residues of the adjacent subunits, such as Phe123, Leu125, Asp126, Gly127, and Leu128 of the monomer B, and Val49, Ser50, and Tyr51 of the monomer D. Table 5 shows the main residues which have interactions with each protein-ligand complexes during the simulation.

The presence of the residues from another monomer that can be observed in other tetrameric *CDA*s highlights the importance of the right packaging in order to analyze potential inhibitors. These finding was corroborated when we analyzed other five crystal structures from *Bacillus Subtillis* (1), *Mus Musculus* (3), and human (1) (PDB access codes: 1JTK, 1ZAB, 2FR5, 2FR6, and 1MQ0, respectively) [9, 31], which also presented the residues from other monomers participating in the ligand stability into the active site. Table 6 shows the residues that are making interactions with each of the ten molecules, and the other five crystal structures interacting with their respective ligands.

Free energy of binding estimation

In this section, we analyzed the estimated free energy of binding using the LIE method. As described above, 40 simulations were performed (four for each protein-ligand complex) to obtain the Lennard-Jones (a) and Coulomb (b) potentials in order to allow the ΔG of binding calculation. The coefficients were calibrated taking into account five human *CDA* inhibitors with inhibition constants previously described. Prior to calibrating the LIE equation terms, we investigated which kind of interaction was prevalent between the protein-ligand complexes. Finally, the parameters adopted were α=0.59, β=0.43, and γ=0, as previously mentioned. Table 7 presents the hit compounds free energy of binding. The lowest value was achieved by compound 10 ($-17.61 kcal \times mol^{-1}$) followed by compounds 8 ($-16.2 kcal \times mol^{-1}$), 6 ($-12.2 kcal \times mol^{-1}$), 7 ($-10.88 kcal \times mol^{-1}$), and 4 ($10 kcal \times mol^{-1}$). Four of these five compounds (10, 8, 6, and 7) present sulfur group replacing amino or hydroxyl groups, commonly found in the substrate or products, respectively. The zinc ion contribution could be highlighted, mainly by Coulomb interactions, to the inhibitor accommodation into the binding cavity.

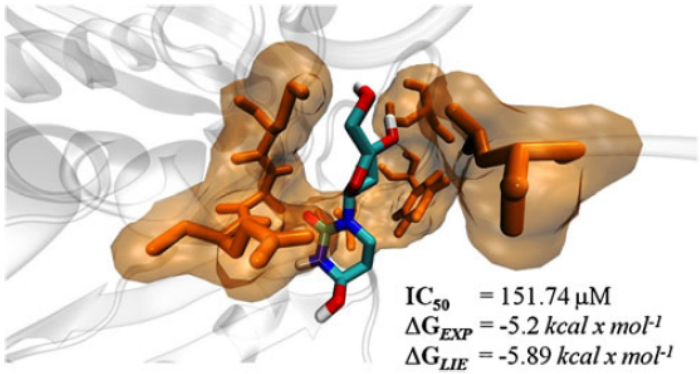


Fig. 6 Compound 9 binding cavity schematic drawing

Compound 4 shows only a hydrogen atom replacing the amino or hydroxyl groups in the pyrimidine moiety, thereby presenting an additional hydroxyl group, which could assist the inhibitor binding mode. On the other hand, the IC_{50} determined for compound 9 ($-5.89 kcal \times mol^{-1}$), a well-established human *CDA* inhibitor, was 151.74 μM (~ -5.2 $kcal \times mol^{-1}$) diverging only $0.7 kcal \times mol^{-1}$ from the value obtained using the LIE equation, demonstrating that the parameters applied into the molecular docking process are capable of finding a potential inhibitor for *MtCDA*.

IC_{50} determination

An IC_{50} value of 151.74 μM was determined for compound 9, a hit compound obtained among several by the molecular docking process (Fig. 5). Although this value might stand at an unacceptable concentration for an ideal inhibitor, it demonstrates a certain level of inhibition, which was not observed for other compounds tested (data not shown) for the same enzyme. Thereby, compound 9 might have the potency improved by chemical derivatization guided by the findings here presented. At any rate, compound 9 represents a hit that, hopefully, could lead to a lead compound with anti-TB activity.

Conclusions

The present study provides structural insight of the *MtCDA* quaternary structure. Analysis between the region involved in the catalytic event in human *CDA* ($_{65}$CAERTA$_{70}$) and *MtCDA* ($_{56}$CAECAV$_{61}$) revealed that the arginine residue found in human *CDA*, important to compensate the charges in the binding cavity, corresponds to a cysteine in *MtCDA*. Even so, it could be suggested that *MtCDA* Arg91, corresponding to Ala101 in human *CDA*, is responsible for counterbalancing the cysteine charges. The molecular docking protocol was capable of identify hit compounds, and suggests that most of them involve interactions among adjacent subunits making clear the importance of van der Walls interactions for ligand binding. Glu47(A), Cys56(A), Glu58(A), Phe123(B), and Tyr51(D) were identified as a core domain to suitably accommodate the ligands into the binding cavity. The IC_{50} value determined for compound 9 (151.74 μM) (Fig. 6) was in agreement with the approach used in the molecular docking process and with the coefficients used in order to estimate the free energy of binding. Furthermore, the coefficients (α, β, and γ) used in view to estimate the free energy of binding through the MD simulations gives promising results, diverging only 0.7 $kcal \times mol^{-1}$ from compound 9 IC_{50} value. Summarizing, our findings open up the possibility to extend this protocol (combining docking and molecular dynamics simulations) to different databases in order to find new potential inhibitors for promising targets based on a rational drug design process.

Acknowledgments This work was supported by National Institute of Science and Technology on Tuberculosis (Decit/SCTIE/MS-MCT-CNPq-FNDCTCAPES). D.S.S. (304051/1975-06), L.A.B. (520182/99-5), and W.F.A. Jr. (300851/98-7) are research career awardees of the National Council for Scientific and Technological Development of Brazil (CNPq). Z.A.S.Q. acknowledges a scholarship awarded by CNPq. R.G.D. is a postdoctoral fellow of CNPq. L.F.S.M.T. acknowledges a scholarship awarded by CAPES. L.F.S.M.T. is grateful to Dr. Ernesto Raúl Caffarena for helpful discussions regarding free energy of binding.

References

1. Donald PR, van Helden PD (2009) The global burden of tuberculosis–combating drug resistance in difficult times. N Engl J Med 360:2393–2395
2. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C (2003) Tuberculosis. Lancet 36:887–899
3. Jassal M, Bishai WR (2009) Extensively drug-resistant tuberculosis. Lancet Infect Dis 9:19–30
4. World Health Organization. Communicable Disease Cluster, 1999
5. Ducati RG, Ruffino-Netto A, Basso LA, Santos DS (2006) The resumption of consumption - a review on tuberculosis. Mem Inst Oswaldo Cruz 101:697–714
6. Shambaugh GE (1979) Pyrimidine biosynthesis. Am J Clin Nutr 32:1290–1297
7. Ireton GC, Black ME, Stoddard BL (2003) The 1.14 A crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. Structure 11:961–972
8. Hyde JE (2007) Targeting purine and pyrimidine metabolism in human apicomplexan parasites. Curr Drug Targets 8:31–47
9. Johansson E, Mejlhede N, Neuhard J, Larsen S (2002) Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 A resolution. Biochemistry 41:2563–2570
10. Wheeler PR (1990) Biosynthesis and scavenging of pyrimidines by pathogenic mycobacteria. J Gen Microbiol 136:189–201
11. Sánchez-Quitian ZA, Schneider CZ, Ducati RG, de Azevedo WF, Jr JCB, Basso LA, Santos DS (2010) Structural and functional analyses of Mycobacterium tuberculosis Rv3315c-encoded metal-dependent homotetrameric cytidine deaminase. J Struct Biol 169:413–423
12. Chung SJ, Fromme JC, Verdine GL (2005) Structure of human cytidine deaminase bound to a potent inhibitor. J Med Chem 48:658–660
13. Vincenzetti S, Mariani PL, Cammertoni N, Polzonetti V, Natalini P, Quadrini B, Volpini R, Vita A (2004) Isoenzymatic forms of human cytidine deaminase. Protein Eng Des Sel 17:871–877
14. Lemaire M, Momparler LF, Raynal NJM, Bernstein ML, Momparler RL (2009) Inhibition of cytidine deaminase by zebularine enhances the antineoplastic action of 5-aza-2'-deoxycytidine. Cancer Chemother Pharmacol 63:411–416
15. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36:D901–D906
16. Irwin JJ, Shoichet BK (2005) ZINC – a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182

17. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
18. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading
19. Solis FJ, Wets RJB (1981) Minimization by Random Search Techniques. Math Oper Res 6:19–30
20. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718
21. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
22. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular Forces. Reidel, Dordrecht, the Netherlands
23. van Aalten DMF, Bywater B, Findlay JBC, Hendlich M, Hooft RWW, Vriend GJ (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J Comput-Aided Mol Des 10:255–262
24. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KM, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03. Gaussian, Wallingford, CT
25. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425
26. Aqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. Protein Eng 7:385–391
27. Vincenzetti S, Quadrini B, Mariani P, De Sanctis G, Cammertoni N, Polzonetti V, Pucciarelli S, Natalini P, Vita A (2008) Modulation of human cytidine deaminase by specific aminoacids involved in the intersubunit interactions. Proteins 70: 144–156
28. Copeland RA (2005) Evaluation of Enzyme Inhibitors in Drug Discovery. Wiley, New York
29. Gasteiger J, Marsili M (1978) A new model for calculating atomic charges in molecules. Tetrahedron Lett 19:3181–3184
30. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital lectronegativity - a rapid access to atomic charges. Tetrahedron 36:3219–3228
31. Teh AH, Kimura M, Yamamoto M, Tanaka N, Yamaguchi I, Kumasaka T (2006) The 1.48 A resolution crystal structure of the homotetrameric cytidine deaminase from mouse. Biochemistry 45:7825–7833

6. CONSIDERAÇÕES FINAIS

A pesquisa e o desenvolvimento de compostos mais efetivos contra a TB representam uma urgente necessidade à saúde pública mundial. O ressurgimento de diferentes cepas de *M. tuberculosis* resistentes a pelo menos três drogas de segunda geração (XDR- TB) reflete a evidente necessidade de gerar novas formas de combater a doença para melhorar as condições do tratamento utilizado atualmente. Nesse contexto, as rotas metabólicas envolvidas em processos bioquímicos essenciais à viabilidade do bacilo compartilham inúmeros alvos potenciais para drogas. Dentre estas rotas, a possível participação da via de salvamento de pirimidinas, na viabilidade da bactéria, torna as enzimas desta via importantes alvos para o desenvolvimento de futuros fármacos e tratamentos capazes de evitar a progressão e a reativação da infecção. Os programas globais de descobrimento de novas drogas antituberculose têm prestado particular interesse na identificação de análogos de pirimidinas com atividade seletiva contra *M. tuberculosis*.

O presente trabalho forneceu detalhes estruturais sobre o empacotamento da estrutura quaternária da MtCDA. Analises comparativas entre a região envolvida no evento catalítico da CDA humana ($_{65}$CAE**R**TA$_{70}$) e *MtCD* ($_{56}$CAE**C**AV$_{61}$) revelou que a arginina encontrada na estrutura da CDA humana, a qual apresenta um importante papel na compensação de cargas na cavidade de ligação, corresponde a uma cisteína na estrutura da MtCDA. Entretanto, foi possível sugerir que a Arg91 da MtCDA, a qual na estrutura da CDA humana corresponde a Ala101, e a responsável por fazer o papel de compensação de cargas.

A triagem virtual de ligantes foi capaz de identificar possíveis potenciais inibidores para MtCDA. Foi possível identificar que a maioria das interações observadas entre os melhores compostos e a enzima são de caráter hidrofóbico envolvendo resíduos das subunidades adjacentes. Os principais resíduos envolvidos na associação dos inibidores com a enzima são Glu47(A), Cys56(A), Glu58(A), Phe123(B), eTyr51(D).

A abordagem utilizada na triagem virtual de ligantes e os coeficientes utilizados para estimar a energia livre de ligação estavam de acordo com o valor de IC50 determinado para o composto 9 (151 μM).

Os coeficientes (α, β e γ) utilizados para estimar a energia livre de ligação através de simulações por dinâmica molecular apresentou resultados interessantes, divergindo apenas 0,7 *kcal x mol*$^{-1}$ do valor experimental. Com isso podemos concluir que a partir dos resultados obtidos e possível utilizar desta abordagem para diferentes bancos de dados com a finalidade de encontrar potenciais inibidores para diferentes tipos de alvos moleculares utilizando como base o processo de desenho racional de drogas.

7. REFERÊNCIAS

Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D. (2002) Molecular Biology of the Cell. 4th edition. New York: Garland Science Publishing.

Alonso H., Bliznyuk A.A., Gready J.E. (2006) *Med. Res. Reviews*., 26, 531.

Betts L., Xiang S., Short S.A., Wolfenden R., Carter C.W. Jr. (1994) *J. Mol. Biol*., 235, 635.

Burman, W.J.; Gallicano, K., and Peloquin C. (1999) *Clin. Infect. Dis.,* 28, 419.

Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., Richardson D.C. (2010) *Acta Crystallogr. D Biol. Crystallogr.*, D66, 12.

Chung S.J., Fromme J.C., Verdine G.L. (2005) *J. Med. Chem.,* 48, 658.

DeLano W.L. (2002) *DeLano Scientific*, San Carlos, CA, USA, (2002) [http://www.pymol.org].

Donald, P.R., and van Helden, P.D. (2009) *N. Engl. J. Med*., 360(23), 2393.

Frieden T.R., Sterling T.R., Munsiff S.S., Watt C.J., Dye C. (2003) *Lancet,* 362, 887.

Goodsell D.S. and Olson A.J. (1990) *Prot. Struc. Func. Gen.*, 8, 195-202.

Goodsell D.S., Morris G.M., Olson, A.J. (1996). *J. Mol. Recog*., 9, 1-5.

Humphrey W., Dalke A., Schulten, K. (1996) *J. Mol. Graphics,* 14, 33.

Ho M.C, Shi W., Rinaldo-Matthis A., Tyler P.C., Evans G.B., Clinch K., Almo S.C., Schramm V.L.(2010) *Proc Natl Acad Sci U S A*. 107(11), 4805.

Islam M.R., Kim H., Kang S.W., Kim J.S., Jeong Y.M., Hwang H.J., Lee S.Y., Woo J.C., Kim S.G. (2007) *Plant. Mol. Biol*., 63(4), 465.

Jain A.N. (2003) *J. Med. Chem*., 46, 499.

Jassal M. and Bishai W.R. (2009) *Lancet Infect. Dis*., 9, 19.

Johansson E., Mejlhede N., Neuhard J., Larsen S. (2002) *Biochemistry*, 41, 2563.

Johnson R., Streicher E.M., Louw G.E., Warren R.M., van Helden P.D., Victor T.C. (2006) *Curr. Issues Mol. Biol*., 8, 97.

Jones G., Willett P., Glen R.C., Leach A.R., Taylor R. (1997) *J. Mol. Biol*., 267, 727.

Kinnings S.L., Liu N., Buchmeier N., Tonge P.J., Xie L., Bourne P.E. (2009) *PLoS Comput Biol.,* 5(7), e1000423.

McCammon J.A., Gelin B.R., Karplus M. (1977) *Nature*, 267(5612), 585.

Metropolis N. and Ulam S. (1949) *J. Am. Stat. Assoc*., 44(247), 335.

Mills N. (2006). *J. Am. Chem. Soc.* 128:(41), 13649.

Moffatt B.A. and Ashihara H. (2002) *The Arabidopsis Book*, doi: 10.1199/tab.0018

Morris G.M., Goodsell D.S., Huey R., Olson A.J. (1996) *J. Comp. Aided Mol. Des.*, 10, 293-304.

Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. (1998) *J. Comput. Chem.*, 19, 1639.

Rivers, E.C., and Mancera, R.L. (2008) *Drug Discov. Today,* 13(23/24), 1090.

Sánchez-Quitian Z.A., Schneider C.Z., Ducati R.G., de Azevedo W.F. Jr., Bloch C. Jr., Basso L.A., Santos D.S. (2010) *J Struct Biol*., 169(3), 413.

Shambaugh 3rd G.E. (1979) *The American J. Clin. Nutrit*., 32, 1290.

Totrov M. and Abagyan R. (1997) *Proteins*, *S1*, 215.

Trouiller P, Torreele E, Olliaro P, White N, Foster S, Wirth D, Pécoul B. (2001) *Trop. Med. Int. Health*, 6(11), 945.

van der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., Berendsen H.J.C. (2004) *J. Comput. Chem.* 26, 1701.

Voet D, Voet J. (2008) Bioquímica. Porto Alegre: Artes médicas; Ed. John Wiley & Sons.

World Health Organization. Communicable Disease Cluster, 1999.

Xiang S, Short SA, Wolfenden R, Carter CW Jr. (1996) Biochemistry, 35(5), 1335.

Xu Y., Johansson M., Karlsson A. (2008) *J. Biol. Chem*., 283(3), 1563.

Zoete V., Grosdidier A., Michielin O. (2009) *J. Cell. Mol. Med.,* 13, 238.

ANEXO 1

Archives of Biochemistry and Biophysics 509 (2011) 108–115

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

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

Crystal structure determination and dynamic studies of *Mycobacterium tuberculosis* Cytidine deaminase in complex with products

Zilpa A. Sánchez-Quitian [a,b], Luís F.S.M. Timmers [b,c], Rafael A. Caceres [c,d], Jacqueline G. Rehm [a], Claudia E. Thompson [c], Luiz A. Basso [a], Walter F. de Azevedo Jr. [b,c,*], Diógenes S. Santos [a,*]

[a] Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF), Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Av. Ipiranga 6681, Porto Alegre, RS 90619-900, Brazil
[b] Programa de Pós Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil
[c] Faculdade de Biociencias, Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Laboratorio de Bioquimica Estrutural (LaBioQuest), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Av. Ipiranga 6681, Porto Alegre, RS 90619-900, Brazil
[d] Programa de Pós Graduação em Medicina e Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 28 December 2010
and in revised form 27 January 2011
Available online 2 February 2011

Keywords:
Tuberculosis
Cytidine deaminase
Crystallography
Molecular dynamics

ABSTRACT

Cytidine deaminase (CDA) is a key enzyme in the pyrimidine salvage pathway. It is involved in the hydrolytic deamination of cytidine or 2′-deoxycytidine to uridine or 2′-deoxyuridine, respectively. Here we report the crystal structures of *Mycobacterium tuberculosis* CDA (MtCDA) in complex with uridine (2.4 Å resolution) and deoxyuridine (1.9 Å resolution). Molecular dynamics (MD) simulation was performed to analyze the physically relevant motions involved in the protein–ligand recognition process, showing that structural flexibility of some protein regions are important to product binding. In addition, MD simulations allowed the analysis of the stability of tetrameric MtCDA structure. These findings open-up the possibility to use MtCDA as a target in future studies aiming to the rational design of new inhibitor of MtCDA-catalyzed chemical reaction with potential anti-proliferative activity on cell growth of *M. tuberculosis*, the major causative agent of tuberculosis.

Introduction

Cytidine deaminase (CDA[1], EC 3.5.4.5, systematic name: cytidine aminohydrolase), which is also known as cytidine nucleoside deaminase, catalyzes the hydrolytic deamination of cytidine or 2′-deoxycytidine to, respectively, uridine or 2′-deoxyuridine (Fig. 1) [1,2]. The *cdd*-encoded CDA plays an important role in the pyrimidine salvage pathway. The latter serves two different functions: (1) scavenge pyrimidine compounds for nucleotide synthesis, and (2) degradation of pyrimidine compounds leading to generation of compounds that serve as carbon and nitrogen sources [3]. In addition, CDA is found in the metabolism of a number of analogues of cytosine nucleoside used as antitumoral and antiviral agents, leading to their pharmacological inactivation [3].

Biochemical and structural studies of CDAs from different organisms have led to the understanding of the enzyme reaction mechanism. CDA contains an active site zinc ion that is essential for catalysis, which is coordinated to a nucleophilic water/hydroxide [4]. Different oligomeric states were observed among CDAs from diverse species. Two groups can be identified, the homodimeric CDAs, such as in *Escherichia coli* [5], and the homotetrameric CDAs, such as in *Saccharomyces cerevisiae* [6], human [2] *Bacillus subtillis* [7] and *Mycobacterium tuberculosis* [8].

The three-dimensional structure of *M. tuberculosis* CDA has been previously solved at 2.0 Å resolution and shown to be composed of four identical subunits, each monomer comprising a single domain with five β-strands and five α-helices [8]. Secondary structural elements are arranged in a three-layer core α–β–α domain with mixed β-sheet of five strands, the last two strands of this structure form a β-wing flexible region.

Sequence alignment of CDAs from different organisms has shown relationships among residues that mediate substrate binding and are conserved in all sequences, in agreement with results obtained from crystal structure determination. For instance MtCDA contained one zinc atom per subunit coordinated by three cysteines (Cys56, Cys89, and Cys92), one glutamate (Glu58) and a water molecule [8]. The same pattern was observed in *B. subtillis* CDA [9].

As the first step toward structural and functional MtCDA characterization, we have previously determined the crystal structure of tetrameric *M. tuberculosis* CDA in the absence of substrate/product ligands [8]. Here we describe two refined crystal structures of

* Corresponding authors. Address: Av. Ipiranga 6681, TecnoPUC, Prédio 92A, Porto Alegre, RS 90619-900, Brazil. Fax: +55 51 33203629.
E-mail addresses: walter@azevedolab.net, wfdaj@uol.com.br (W.F. de Azevedo Jr), diogenes@pucrs.br (D.S. Santos).

[1] *Abbreviations used:* CDA, cytidine deaminase; URD, uridine; dURD, deoxyuridine; dCYT, deoxycytidine; PCA, principal component analysis; TB, tuberculosis; MTB, *Mycobacterium tuberculosis*.

0003-9861/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2011.01.022

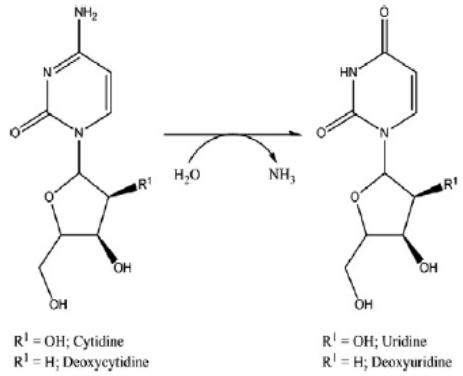


Fig. 1. Chemical reaction catalyzed by MtCDA.

MtCDA in complex with products. These data provide new structural information on the mechanism of active site closure upon ligand binding, and highlight the residues essential for catalysis. Furthermore, molecular dynamics (MD) simulation was performed to analyze the physically relevant molecular motions involved in protein–ligand binary complex formation.

Materials and methods

Crystallization and X-ray data collection

MtCDA was concentrated to 12 mg/mL in sample buffer (20 mM Tris–HCl pH 7.5). Crystals of MtCDA were obtained using the vapor-diffusion hanging-drop method at 298.15 K using 24-well tissue-culture plates. Hanging drops were prepared by mixing 2 μL protein solution and 2 μL reservoir solution and placed over 400 μL reservoir solution containing 0.1 M Hepes pH 7.5 and 4.3 M sodium chloride. The ligands uridine (URD) and deoxyuridine (dURD) were added by soaking method.

The crystals were flash frozen at 100 K and cryoprotected with 20% of glycerol. X-ray diffraction data were collected at wavelength 1.431 Å using Synchrotron Radiation Source (Laboratório Nacional de Luz Síncrotron, Campinas, SP, Brazil) and CCD detector (MARCCD) with an exposure time of 1.6 s per image at a crystal to detector distance of 140 mm. X-ray diffraction data were processed using MOSFLM and scaled using SCALA [10].

Structure resolution and refinement

The crystal structures of MtCDA:URD and MtCDA:dURD were solved by molecular replacement with AMoRe package using the atomic coordinates of the structure 3IJF [8] as the search model. All waters molecules were removed from the search model and one monomer was employed in the molecular replacement. For structural refinement, the models were refined using *REFMAC5* [10] to give the final R_{factor} = 0.223 and R_{free} = 0.2905 to MtCDA: URD and R_{factor} = 0.227 and R_{free} = 0.257 to MtCDA:dURD. The *MOLPROBITY* web server [11] was used to generate the Ramachandran plots, and the percentage of residues in favored, allowed, and disallowed for MtCDA:URD are, 90.2%, 97.5%, and 2.5%, and for MtCDA:dURD are 96.7%, 97.5%, and 2.5%, respectively. All main chain angles, for all structures, are found in the most favorable or generally allowed regions.

Molecular dynamics simulation setup

All molecular dynamics (MD) simulations were carried out using the GROMACS 4.0.5 package [12] with the 53a6 GROMOS force field. We performed MD simulations for the enzyme in free form (PDB access code: 3IJF), and with the products (URD and dURD – PDB access code: 3LQP and 3LQT, respectively), in order to investigate the plasticity of the active site, and the interactions between the protein–ligand complex, and water molecules during the MD simulation.

The MD simulations were carried out using particle mesh Ewald method [13] for the electrostatic interactions. The van der Walls and Coulomb cutoff were 14 Å, and 10 Å, respectively. The integration time step was 2.0 fs, with the neighbor list being updated every 10 steps by using the grid option and a cutoff distance of 12 Å. The simple point charge extended (SPC/E) [14] water model was used. Periodic boundary condition has been used with constant number of particles, pressure, and temperature (NPT) in the system. The V-rescale thermostat was applied using a coupling time of 0.1 ps to maintain the systems at a constant temperature of 298.15 K. The Berendsen barostat was used to maintain the systems at pressure of 1 bar, and values of the isothermal compressibility set to 4.5×10^{-5} bar^{-1} for water simulations. The temperature of the systems were increased from 50 to 300 K in five steps (50–100 K, 100–150 K, 150–200 K, 200–250 K, and 250–300 K), and the velocities at each step were reassigned according to the Maxwell–Boltzmann distribution at that temperature and equilibrated for 10 ps, except the last part of thermalization phase, which was of 40 ps. The systems were subjected to a steepest descent followed by conjugated gradient energy minimizations until a tolerance of 1000 kJ/mol. MD simulation with position restraints was carried for a period of 20 ps in order to allow the accommodation of the water molecules in the system. Finally, 10 ns MD simulations were performed to all systems. The topologies files and other force field parameters except the charges of ligands were generated using the PRODRG program [15]. The partial atomic charges to the ligands were calculated using the Gaussian 03W package [16], which were submitted to single-point *ab initio* calculations at DFT/B3LYP/6-311G (2d, p) level in order to obtain ESP charges. In order to clarify the importance of the tetrameric structure of the MtCDA, MD simulations were carried out with the monomeric state, using the same protocol previously described.

Principal component analysis (PCA)

Essential dynamics (ED), also known as principal component analysis (PCA), is a method for identification of the main conformational changes, which often have importance in biological process, for a protein during an MD simulation.

In the ED analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. Diagonalizing the matrix identified a set of eigenvectors and eigenvalues. The eigenvalues represented the amplitude of the eigenvectors along the multidimensional space, and the displacements of atoms along each eigenvector showed the concerted motions of protein along each direction. An assumption of ED analysis is that the correlated motions for the function of the protein are described by eigenvectors with large eigenvalues. The movements of protein in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis [17].

Results and discussion

Quaternary structure description

The final models (URD, dURD) were refined to, respectively, 2.4 and 1.9 Å resolution. The active site electron densities for complexes MtCDA:URD and MtCDA:dURD are presented in Fig. 2. Data collection and refinement statistics for all structures are given in

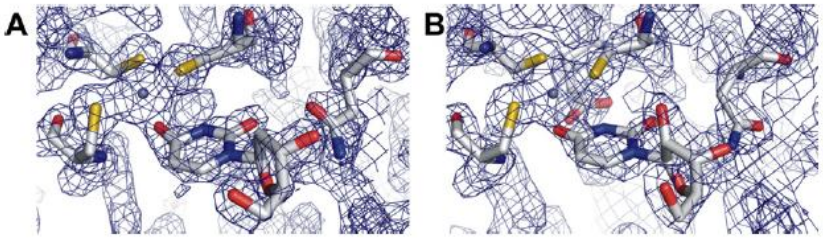


Fig. 2. Active site electron densities ($2F_{obs} - F_{calc}$) map for complexes MtCDA:dURD (A) and MtCDA:URD (B). The URD, dURD, and the residues are presented as sticks.

Table 1
Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell.

Data collection and refinement statistics*	MtCDA:UDR	MtCDA:dURD
X-ray wavelenght (Å)	1.427	1.427
Temperature (K)	100	100
Resolution range (Å)	39.16 (2.40)	37.216 (1.90)
Space group	C222	C222
Matthews coefficient (Å^3 Da^{-1})	2.60	2.63
Unit-cell parameters		
a (Å)	65.92	65.41
b (Å)	78.33	77.59
c (Å)	56.73	55.68
$\alpha = \beta = \gamma$ (°)	90.00	90.00
Highest resolution shell (Å)	2.40	1.90
Data completeness (%)	99.90	92.9
R_{merge} (%)[a]	7.3	7.7
Resolution range used in refinement (Å)	28.41–2.40	28.20–1.90
R_{factor} (%)[b]	22.33	22.70
R_{free} (%)[c]	29.05	25.70
Observed RMSD from the ideal geometry		
Bond lenghts (Å)	0.041	0.023
Bond angles (°)	3.548	2.435
B values (Å^2)[d]	24.065	25.524
Residues in most favored regions of the Ramachandran plot (%)	90.2	97
Residues in addicionally allowed regions of the Ramachandran plot (%)	7.3	2
Residues in generously allowed regions of the Ramachandran plot (%)	2.5	1
Residues in disallowed regions of the Ramachandran plot (%)	0	0
Number of ligands	1	1
Number of water molecules	93	50

[a] $R_{merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \sum_i I(h)_i$, where $I(h)$ is the intensity of reflection h, $\sum_h$ is the sum over all reflections and $\sum_i$ is the sum over i measurements.
[b] $R_{factor} = 100 \times \sum |F_{obs} - F_{calc}| / \sum (F_{obs})$, the sums being taken over all reflections with $F/\sigma(F) > 2$ cutoff.
[c] $R_{free} = R_{factor}$ for 10% of the data, which were not included during crystallographic refinement.
[d] *B* values = average *B* values for all non-hydrogen atoms.
* Values in parenthesis refer to the highest resolution shell.

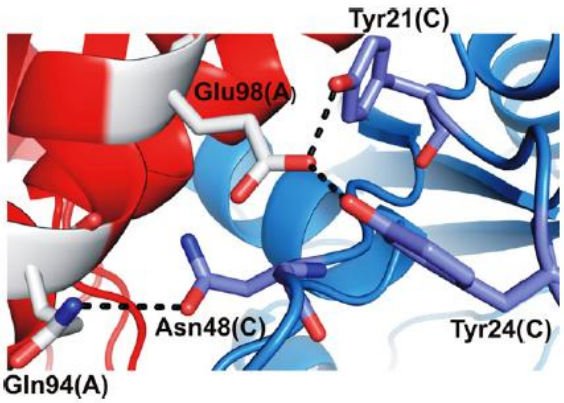


Fig. 3. Interface of the subunits A and C, showing the interactions between Glu98(A) with Tyr21(C), Tyr24(C), and Gln94(A) with Asn48(C). The protein is presented as cartoon and the residues as sticks.

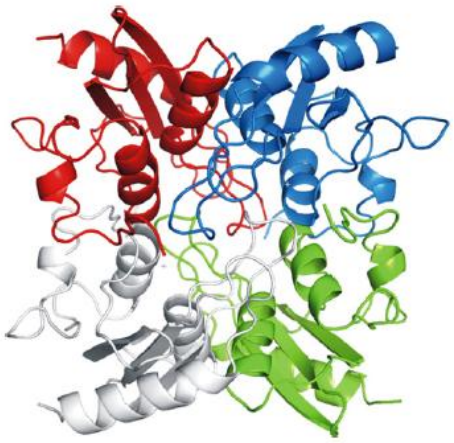


Fig. 4. Schematic drawing of tetrameric MtCDA. The protein is presented as cartoon and each monomer (A-red, B-blue, C-light grey, and D-green) is presented in different colors. This figure was generated using Chimera program [25]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1. The MtCDA monomer in complex with the products, follow the same arrangement α–β–α with a mixed β-sheet of five strands. These structures were crystallized in the space group C222. Examining the differences between MtCDA in free form and MtCDA complexed with products was observed that there are no major differences. The only changes are restricted to amino-acid side chains, which compose the active site. The RMSD value between apo structure and URD complex is 0.4 Å, and between apo structure and dUDR complex is 0.2 Å. Analysis of the interfaces reveals intermolecular contacts involving residues Tyr21–Glu98, Tyr24–Glu98, and Gln94–Asn48 of the adjacent subunits (A–C and B–D) (Fig. 3). In addition, analysis based on the web server PDBe PISA (Protein Interfaces, Surfaces and Assemblies) [18], reveals 23 hydrophobic interactions, and eight intermolecular hydrogen bonds between the adjacent subunits, and brought to light what seems to be a stable MtCDA quaternary structure (Fig. 4). Analyzing the contact areas between subunits of MtCDA:URD and MtCDA:dURD complexes, we suggest that the monomers can interact in three different ways. These different forms can be exemplified by the monomers A:B, A:C, and B:C, which are presented in the Fig. 5. Examining the interactions between A:B it was observed six hydrogen bonds and 17 hydrophobic contacts involving 23 residues, and the total area was 757.1 A^2. The total area for A:C was 1128.4 A^2, in addition 14 hydrogen bonds and

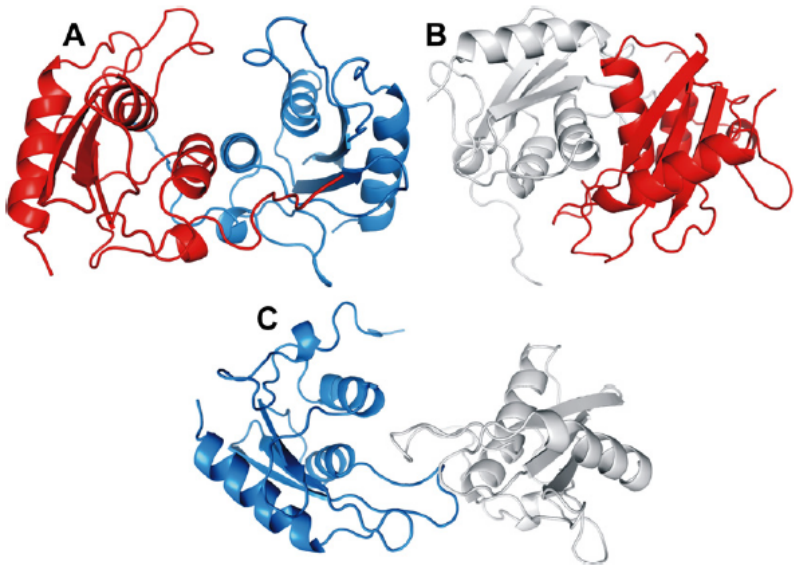


Fig. 5. The three interfaces contacts area, which are found to MtCDA:URD and MtCDA:dURD complexes. (A) Present A (red):B (blue); (B) A:C (light grey); and (C) B–C interfaces. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

17 hydrophobic contacts involving 31 residues were observed. Finally, the region between B:C presents a total area of 412.1 A^2, which is composed by two hydrogen bonds and 10 hydrophobic contacts. Furthermore, in order to verify if these conformations are in agreement with CDA of different organism previously published, we analyzed other 13 CDA structures (PDB access codes: 3IJF, 3MPZ, 1MQ0, 3DMO, 1UX1, 2D30, 1UX0, 1UWZ, 1JTK, 2FR5, 1ZAB, 2FR6, and 3B8F) [19,20], which revealed the same packing presented by MtCDA.

Characterization of the product-binding pocket

The hydrogen bonding networks in the active site were determined from interactions between MtCDA and URD (MtCDA:URD) and dURD (MtCDA:dURD) products. Analysis of hydrogen bonds between URD and MtCDA reveals seven hydrogen bonds: six hydrogen bonds between the pyrimidine moiety and Cys89 (2.9 and 3.1 Å), Glu58 (3.4 and 3.2 Å), Cys92 (3.2 Å) and Ala57 (3.3 Å) residues, and one between the ribose moiety and Asn45 (3.2 Å) residue (Fig. 6). Additionally the complexes show two non-ligand residues, Phe27 and Cys56 which could be important in stabilizing hydrophobic interactions between the ligand and the enzyme and thereby facilitate the catalytic process, as reported for human CDA [21]. All of these interactions are located in the subunit A.

In contrast, some differences were found in the active site between MtCDA:dURD binary complex and apo MtCDA, which involve subunits A and D. This complex reveals seven hydrogen bonds, three with pyrimidine Cys89 (2.9 Å), Glu58 (2.9 Å), Ala57 (3.3 Å) and four with ribose involving residue Asn45 (2.7 Å), Glu47 (2.8 Å), Tyr51(D) (3.1 Å), and Val49(D) (2.9 Å).

Three cysteine residues (Cys56, Cys89, and Cys92) are involved in zinc metal binding. However, only Cys56 is involved in hydrophobic contacts, as previously mentioned. The properties of the Cys56 (Cys65 human CDA numbering) suggested it is required for the enzymatic activity but not for the maintenance of zinc in the active site, whereas Cys65 (human CDA) plays a role in placing the zinc ion in the correct position and orientation within the active site [22]. Table S1 summarizes the interaction between *Mt*CDA:URD and dURD. Analyzing the three main regions, which are described to all tetrameric CDAs, we observed some differences that could be a guide to future rational drug design studies aiming

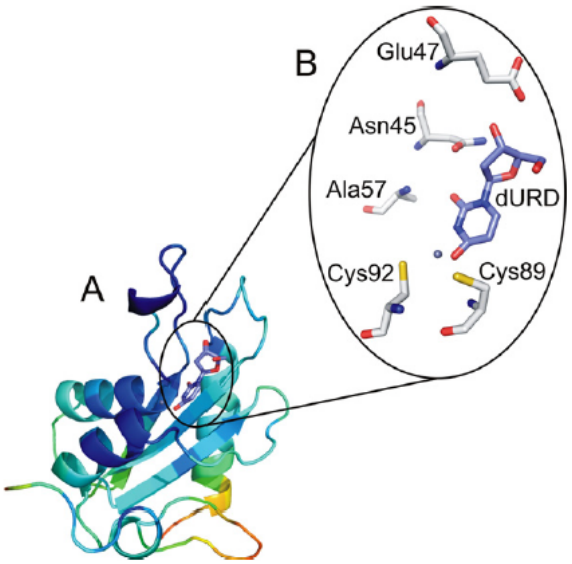


Fig. 6. Classical binding site of the MtCDA:dURD anchored by the residues Asn45, Glu47, Ala57, Cys89, and Cys92, which are important residues of the binding pocket. The protein is presented as cartoon and the binding site residues as sticks.

to find selective inhibitors to *Mt*CDA. In the first region, close to the active site, which comprises $_{32}$PYS**H**F$_{36}$ (human CDA) and $_{23}$PYS**R**F$_{27}$ (*Mt*CDA), the replacement of a histidine by an arginine does not implies physical chemistry changes. However, considering torsion angles, arginine presents increased entropy, which could influence ligand recognition. It is should be pointed that Tyr33, already characterized as a pivotal residue involved in the quaternary structure organization in human CDA [23], is conserved in both sequences, and could be playing the same role in *Mt*CDA. The second region comprises the residues $_{57}$N**A**C**Y**P$_{61}$ (human CDA) and $_{48}$N**V**S**Y**G$_{52}$ (*Mt*CDA). Three replacements are observed in this sequence. The first, an alanine residue was changed by a valine did not implied major structural difference, although valine presents one more torsion angle when compared to alanine. The replacement of a cysteine by a serine, involves of a thiol (Cys) and hydroxyl groups (Ser). However, both residues can participate

in hydrogen bonding. The last change in this sequence is the presence of a proline in human CDA and a glycine in *Mt*CDA. Finally, the last analyzed region, conserved in all T-CDAs, comprises the residues $_{65}$CAE**RTA**$_{70}$ in human CDA and $_{56}$CAE**CAV**$_{61}$ in *Mt*CDA. We observed that Arg68, described as an important residue in the catalytic event in the human CDA [24], compensating the negative charges of the cysteines, corresponds to a cysteine in *Mt*CDA. Nevertheless, analyzing the *Mt*CDA structure, it could be suggested that the residue responsible for counterbalancing the charges of the cysteines is Arg91, which is not conserved in the human CDA, where it corresponds to Ala101. Taking these differences in consideration, MtCDA presents peculiarities close to the active site, when compared to human CDA, which could be involved in the ligand recognition process and can provide some clues to rational drug design initiatives.

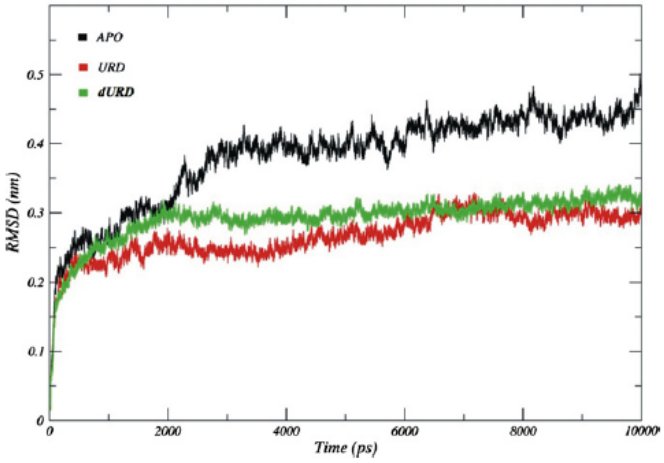


Fig. 7. Graphical representation of RMSD of each MtCDA structure from starting structure of tetramers as a function of time, over 10 ns of simulation.

Dynamics simulation and essential dynamics

In order to understand the stability of the MtCDA in solution during the simulation, the C–α root means square deviation (RMSD) was calculated (Fig. 7). The MtCDA in the apo form and in complex with either URD or dURD achieved a plateau within 2500 and 2000 ps, respectively, suggesting that ~10 ns is sufficient for stabilizing fully relaxed models. As shown in Fig. 7, the RMSD values in the MtCDA:URD and dURD remain less 3.0 Å, whereas in the apo form was less than 5.0 Å. These results suggest that the MtCDA when in complex with products presents a more stable system than the structure in absence of these ligands in the binding pocket. To analyze the physically relevant motions of the simulations, the principal components (PC) were calculated by diagonalizing the covariance matrix generated from timescale

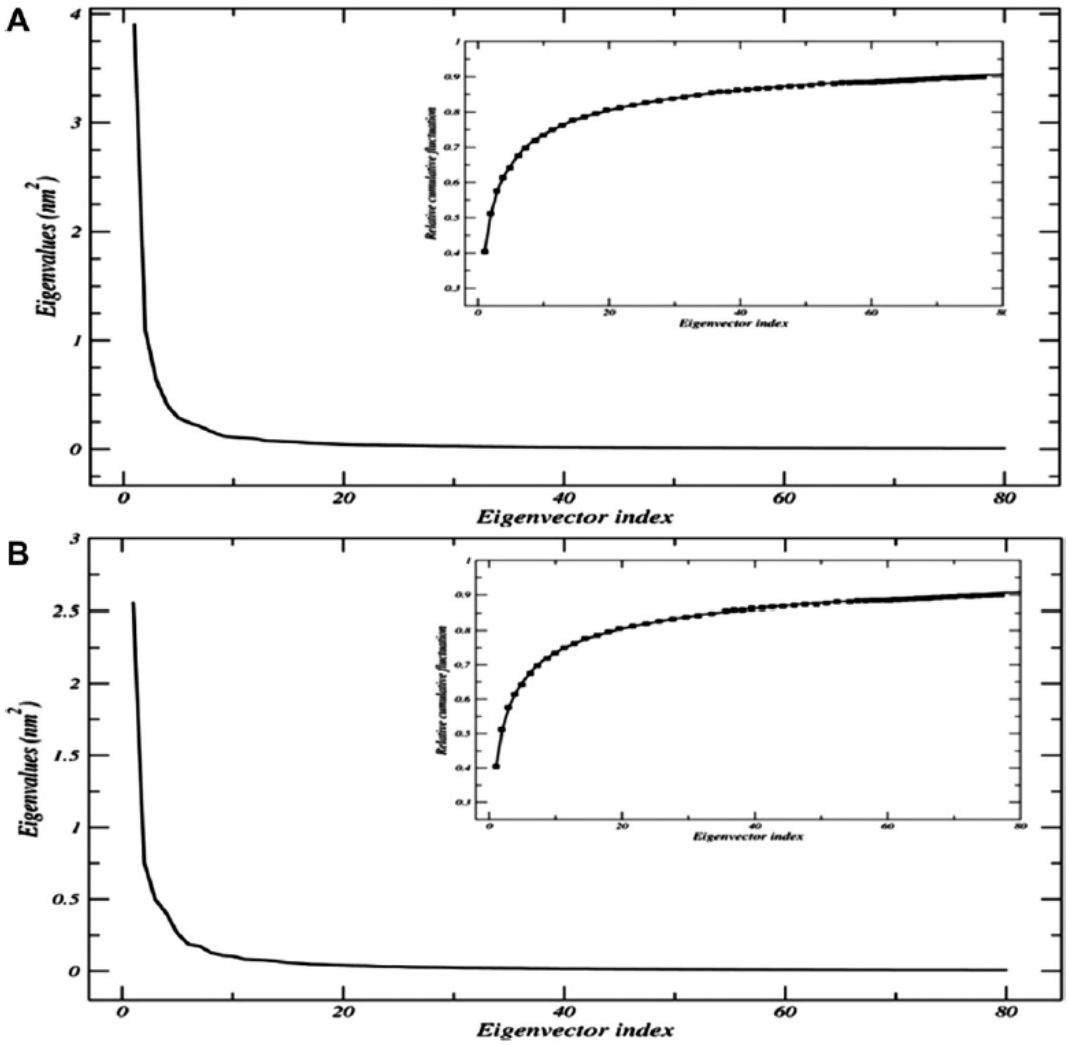


Fig. 8. Eigenvalues of the covariance matrix resulting from the simulations of the MtCDA:dURD (A) and MtCDA:URD (B) complexes.

(10 ns) explicit solvent MD simulation. The anharmonic and large-scale motions of the protein are isolated from the mostly harmonic and small-scale motion. Knowing that the large-scale anharmonic motions in the essential subspace are often correlated to the pivotal functions of the protein, we have only focused on these movements. The analysis of the first 50 eigenvectors describes 87.71% (dURD) and 85.33% (URD) of the fluctuations during the simulation, and the first eigenvector of each simulation correspond of ~40% of the total of 1488 eigenvectors (Fig. 8).

Flexibility of the product-binding pocket

In this section, we analyzed the main fluctuations of the MtCDA in complex with its products in comparison with the structure in apo form. Fig. 9 presents the most relevant fluctuations, obtained from the first eigenvector, of the monomer A during the simulation. Based on these results, it could be highlighted three principal regions that will be denominated R1, R2, and R3. The R1 region, that comprises residues Val22–Phe27, is located in the monomer interface, being crucial for the stabilization of the quaternary structure. Furthermore, it was observed the approximation of Tyr24 to Glu98 of the adjacent subunit, corroborating with the analysis previously published [9]. As an understanding of product binding to, and release from, the enzyme active site is pivotal to elucidate the MtCDA mode of action, the R2 region was analyzed. R2 comprises residues Thr42–Cys56. Analyzing the active site of the monomer A, one can observe three main interactions, made by Asn45, Glu47, and Cys56 contributing for the stability of the ligands into the binding cavity. In addition, it was also observed that Val49 and Tyr51 belong to the binding site of the monomer D. The R3 region is composed by Val110–Phe123 and is located close to the adjacent binding site (A–B and C–D). In this region we also detect π–π stacking interactions involving the pyrimidine moiety of the products and Phe123, contributing to ligand binding.

Tetramer vs monomer: the ligand stability based on dihedral analysis

In order to access the importance of the quaternary structure to the ligand stability into the binding pocket, we carried out MD simulations with the MtCDA in monomeric form. To analyze the stability of ligand in these two scenarios (tetramer and monomer) we monitored dihedral angle of the products (C6:N1:C1∗:O4∗) and the electrostatic interaction (*Coulomb – Short Range*) between protein–ligand complexes for the period of the simulation. Fig. S1 presents the dihedral angle that was evaluated, and Figs. S2 (URD) and S3 (dURD) show the dihedral angle and energy interaction during the simulations. The total number of transitions, considering the dihedral angle, was 97 for the monomeric structure (URD and dURD), 11 for tetramer complexed with URD and 39 when associated with dURD. In addition, the transitions of the dihedral angle in the tetramer have a direct influence in the electrostatic interactions, suggesting that, for a proper stabilization of the ligand into the binding site, the quaternary structure is required. When the average of the electrostatic interactions in the tetrameric and monomeric structures in complex with the products were compared, it was clear that the structure that enabled the most stable form of the ligand was the tetramer. Table S2 summarizes these findings.

The dynamics motions of the ligands (URD and dURD) were analyzed via characterization of the phase space behavior. Fig. 10 present the trajectory of the first two principal components (PC1 and PC2), taking into account the motions of the ligands in the phase space. As shown in Fig. 10A and B, we can identify three clusters of stable states for the ligands when in complex with monomeric structures. In contrast, Fig. 10C and D present only one cluster, defining only one stable state for the ligands. These results could be related with the stability of the ligand in the binding site, and justify the use of the quaternary structure for future studies of potential inhibitors to MtCDA. Furthermore, as described above, these results corroborate the importance of the adjacent subunits

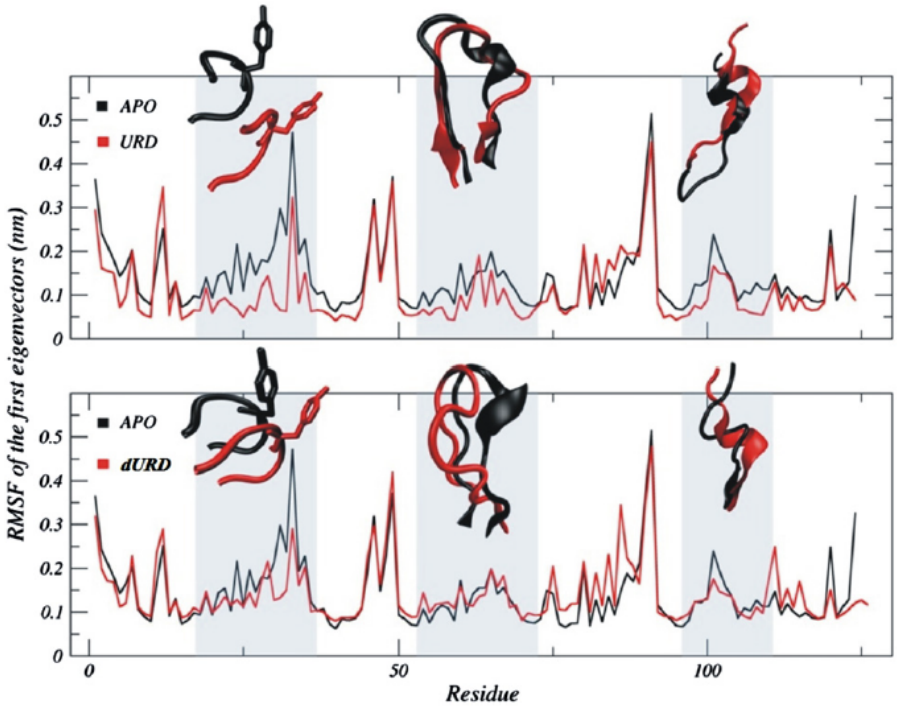


Fig. 9. Displacement of the components of the first eigenvectors for tetrameric structures of unbound MtCDA, MtCDA:URD and MtCDA:dURD complexes. The main fluctuations are presented for the regions R1 (Val22–Phe27), R2 (Thr42–Cys56) and R3 (Val110–Phe123). The main residues, which composed the regions described above, are presented as cartoon in order to facilitate the interpretation.

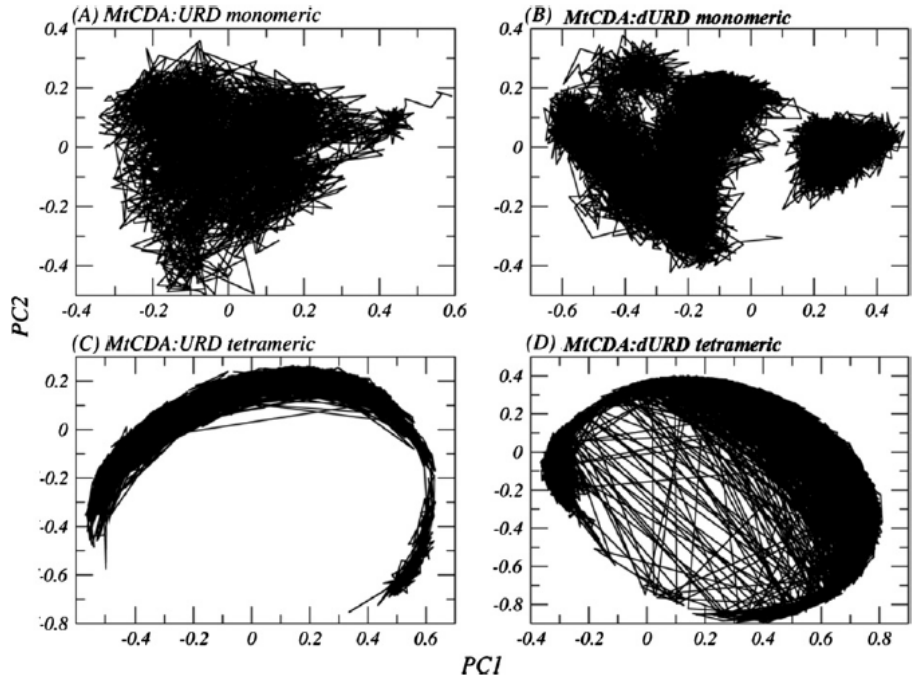


Fig. 10. Dynamics motions of the ligands (URD and dURD) were analyzed via characterization of the phase space behavior when in complex with either a tetrameric or a monomeric structure. (A and B) Trajectory of the first two principal components to URD and dURD, respectively, when associated with MtCDA monomeric form; (C and D) trajectory of the first two principal components to URD and dURD, respectively, when associated with MtCDA tetrameric form.

residues Val49 (monomer D), Tyr51 (monomer D), and Phe123 (monomer B) for stability of the ligand into the binding pocket, since in absence of these residues, the MtCDA reaction product does not remain in the active site during MD simulations.

Conclusions

The present study provided, for the first time, structural insight into the product binding patterns for MtCDA, elucidating the main residues involved in proper binding of products. In addition, our results highlight that the MtCDA quaternary structure follows the same packaging, which is observed in others CDAs from different organisms. Analysis of the active site revealed that the Asn45, Glu47, Ala57, Cys89, and Cys92 amino acid residues are pivotal to product recognition by MtCDA. The MD simulations corroborated with the proposed packing, showing that residues Tyr51, Val49, and Phe123 present an important role in the active site of the adjacent subunits (A–D and B–C). Simulations of the tetrameric and monomeric MtCDA structures suggest that the oligomeric state is essential for stabilization of the ligand into the enzyme-binding site, lending support to the importance of protein–protein interactions. In summary, these findings open-up the possibility to use MtCDA as a target in future studies aiming to the rational design of new inhibitor of MtCDA-catalyzed chemical reaction with potential anti-proliferative activity on cell growth of *M. tuberculosis*, the major causative agent of tuberculosis.

Deposit

The atomic coordinates have been deposited in the PDB (complexes MtCDA:URD and MtCDA:dURD – PDB access codes: 3LQP and 3LQT, respectively).

Acknowledgments

This work was supported by National Institute of Science and Technology on Tuberculosis (Decit/SCTIE/MS-MCT-CNPq-FNDCT-CAPES). D.S.S. (304051/1975-06), L.A.B. (520182/99-5), and W.F.A. Jr. (300851/98-7) are research career awardees of the National Council for Scientific and Technological Development of Brazil (CNPq). R.A.C. acknowledge a scholarship awarded by CNPq. L.F.S.M.T. and Z.A.S.Q. acknowledges a scholarship awarded by CAPES.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.abb.2011.01.022.

References

[1] R. Cohen, R. Wolfenden, J. Biol. Chem. 246 (1971) 7561–7565.
[2] D.F. Wentworth, R. Wolfenden, Metthods Enzymol. 51 (1978) 401–407.
[3] A. Somasekaram, A. Jarmuz, A. How, J. Scott, J. Biol. Chem. 274 (1999) 28405–28412.
[4] S.J. Chung, C. Fomme, G. Verdine, J. Med. Chem. 48 (2005) 658–660.
[5] E. Johansson, N. Mejlhede, S. Larsen, Biochemistry 41 (2002) 2563–2570.
[6] P. Itapa, G. Cercignani, E. Balestreri, Biochemistry 9 (1970) 3390–3395.
[7] L. Betts, S. Xiang, S.A. Short, R. Wolfenden, C.W. Carter, J. Mol. Biol. 235 (1994) 635–656.
[8] Z.A. Sánchez-Quitian, C.Z. Schneider, R.G. Ducati, W.F. de Azevdo Jr., C. Bloch, L.A. Basso, D.S. Santos, J. Struct. Biol. 169 (2010) 413–423.
[9] D. Carlow, C. Carter, N. Mejhede, R. Wolfenden, Biochemistry 38 (1999) 12258–12265.
[10] CCP4, Collaborative Computational Project Number 4, Acta Cryst. D. 50 (1994) 760–763.
[11] I.W. Davis, L.W. Murray, J.S. Richardson, D.C. Richardson, Nucleic Acids Res. 32 (2004) W615–W619.
[12] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, J. Comput. Chem. 26 (2005) 1701–1718.
[13] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089–10092.
[14] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), Intermolecular Forces, D. Reidel Publishing Company, Netherlands, 1981.
[15] D.M.F. van Aalten, B. Bywater, J.B.C. Findlay, M. Hendlich, R.W.W. Hooft, G.J. Vriend, J. Comput. Aided Mol. Des. 10 (1996) 255–262.
[16] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.J. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K.

Raghavachari, J.B. Foresman, J. Cioslowski, J.B. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 03, Gaussian, Wallingford, CT, 2003.
[17] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Proteins 17 (1993) 412–425.
[18] E. Krissinel, K. Henrick, J. Mol. Biol. 372 (2007) 774–797.
[19] E. Johansson, J. Neuhard, M. Willemoes, S. Larsen, Biochemistry 43 (2004) 6020–6029.
[20] A.-H. The, M. Kimura, M. Yamamoto, N. Tanaka, I. Yamaguchi, T. Kumasaka, Biochemistry 45 (2006) 7825–7833.
[21] S. Vincenzetti, G. De Sanctis, S. Costanzi, G. Cristalli, P. Mariani, G. Mei, J. Neuhard, P. Natalini, V. Polzonetti, A. Vita, Protein Eng. 16 (2003) 1055–1061.
[22] A. Cambi, S. Vincenzetti, J. Neuhard1, S. Costanzi, P. Natalini, A. Vita, Protein Eng. 11 (1998) 59–63.
[23] D. Micozzi, S. Pucciarelli, F.M. Carpi, S. Costanzi, G. De Sanctis, V. Polzonetti, P. Natalini, I.F. Santarelli, A. Vita, S. Vincenzetti, Int. J. Biol. Macromol. 47 (2010) 471–482.
[24] S. Vincenzetti, B. Quadrini, P. Mariani, G. De Sanctis, N. Cammertoni, V. Polzonetti, S. Pucciarelli, P. Natalini, A. Vita, Proteins 70 (2008) 144–156.
[25] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.