



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

Centro de Ciências Biológicas e da Saúde

Programa de Pós-Graduação em Biologia Molecular e Celular

Análises Estruturais *in silico* da Anidrase Carbônica Cd e Zn dependentes

Lelio Kallut Almeida Netto

Orientadora: Prof.^a Dr.^a Joelma Freire de Mesquita

Rio de Janeiro

2017

Lelio Kallut Almeida Netto

**Análises Estruturais *in silico* da Anidrase Carbônica
Cd e Zn dependentes**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Molecular e Celular da Universidade Federal do Estado do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular.

Orientadora: Prof.^a Dr.^a Joelma Freire de Mesquita

Rio de Janeiro

2017

A447

Almeida Netto, Lelio Kallut
Análises Estruturais in silico da Anidrase
Carbônica Cd e Zn dependentes / Lelio Kallut Almeida
Netto. -- Rio de Janeiro, 2017.
72 f. : il.

Orientadora: Joelma Freire de Mesquita.
Dissertação (Mestrado) - Universidade Federal do
Estado do Rio de Janeiro, Programa de Pós-Graduação
em Genética e Biologia Molecular, 2017.

1. Thalassiosira weissflogii. 2. anidrase
carbônica. 3. cádmio. 4. zinco. I. Mesquita, Joelma
Freire de , orient. II. Título.

Lelio Kallut Almeida Netto

Análises Estruturais *in silico* da Anidrase Carbônica Cd e Zn dependentes

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Molecular e Celular da Universidade Federal do Estado do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular.

Aprovado em: 07/07/2017

Banca Examinadora



Prof^ª. Dr^ª. Cleonice Alves de Melo Bento (UNIRIO)



Prof^ª. Dr^ª. Patricia Cristina dos Santos Costa (UNIRIO)



Prof^ª. Dr^ª. Rosa Maria Tavares Haido (UNIRIO)



Orientadora: Prof. Dr^ª. Joelma Freire de Mesquita (UNIRIO)

*Aos meus pais, Lelio e Isabel.
Às minhas irmãs, Viviane, Giselle e Ana Carolina.
À minha linda sobrinha Giovanna.
E à minha avó Joselita,
Por todo apoio e compreensão.*

AGRADECIMENTOS

Agradeço a Deus por me conceder a vida, por me oferecer a chance de estudar a vida, por me fazer feliz nesta vida e por me conceder a vida eterna.

Aos meus pais, Lelio e Isabel, por todo o apoio e pelas batalhas que tiveram que enfrentar para que os meus sonhos fossem realizados. Mesmo com altos e baixos, sempre estivemos juntos e tudo o que tenho é graças ao esforço de vocês. Educaram-me e me ensinaram a agir com respeito a todas as pessoas, sem exceção. Tudo o que tenho e tudo o que sou é graças a vocês dois. Obrigado pela minha vida. Amo muito vocês!

Às minhas irmãs, Viviane, Giselle e Ana Carolina, por me aturarem e compreenderem que foi preciso muita dedicação para a realização dos meus sonhos. Vou estar sempre aqui para ajuda-las no que for preciso, assim como vocês sempre estiveram dispostas a me ajudar. Amo muito todas vocês!

À minha sobrinha Giovanna, que trouxe mais alegria a nossa família, por sentir tudo que acontece dentro de casa e confortar os nossos corações com um abraço bem apertado. O tio ama muito você!

À minha avó Joselita, por ser a avó mais complexa que uma família pode ter. Por simplesmente me apoiar e incentivar nos estudos para que eu seja “um doutor cientista”. Em todos os momentos em que eu mais precisei, ela esteve presente e sei que, mesmo não estando mais comigo, continua torcendo por mim como sempre fez. Vou Te amar para sempre, vó!

Aos meus tios e tias, por toda a ajuda, força e boa vontade de me receber em uma visita para conversar e me motivar a não desistir dos meus planos.

À professora Dr.^a Joelma Freire de Mesquita, por me receber de braços abertos no momento em que eu mais precisava. Além de ser uma ótima orientadora, é uma grande amiga, psicóloga, me ajudando a resolver todos os problemas que surgiram com muitas gargalhadas. Obrigado por me acolher e por todo carinho. Tudo que me ensinou, aprendi e levarei para o resto da minha vida, seguindo seus passos de como ser um excelente profissional.

Aos professores da banca examinadora, todo o meu agradecimento e carinho por aceitarem fazer parte da minha última etapa para a conclusão do meu tão sonhado grau de mestre. Obrigado!

À Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e a todo o corpo docente do PPGBMC pelos conhecimentos adquiridos.

Aos meus queridos amigos por me aceitarem e me apoiarem nas minhas escolhas. Por me mostrar que mesmo distante, podemos contar com a ajuda dos outros a qualquer momento. Cada um de vocês teve um papel fundamental no meu crescimento. Agradeço muito por toda a ajuda, apoio, compreensão. Vocês me ensinaram a ser mais humano. Agradeço pelos nossos encontros loucos, que nos fazem relaxar e esquecer por um tempo os grandes obstáculos que surgem em nossas vidas. Amo vocês!

E por fim, agradeço a todos que direta ou indiretamente, me ajudaram bastante para a realização deste sonho.

Resumo

A anidrase carbônica é uma enzima presente em todos os organismos que fazem parte da árvore evolutiva, de archaea a mamíferos. Esta enzima contém zinco e catalisa a hidratação reversível de CO₂. Uma nova anidrase carbônica (CA), denominada CDCA1, foi descoberta na diatomácea marinha *Thalassiosira weissflogii*. A CDCA1 possui ação reversível uma vez que naturalmente usa Cd²⁺ como o íon de metal catalítico, mas se necessário pode espontaneamente trocar Cd²⁺ pelo Zn²⁺. A existência de uma enzima de cádmio nos fitoplânctons marinhos pode indicar que existe uma única seleção de pressão para metaloenzimas no ambiente marinho, e essa descoberta fornece uma explicação há muito tempo aguardada para o comportamento do cádmio como nutriente nos oceanos. Neste estudo, a CDCA1 foi submetida à análise *in silico* utilizando diferentes algoritmos para obter o modelo estrutural mais compatível, a fim de determinar suas características estruturais, possibilitando uma relação filogenética com outras espécies que absorvem cádmio. A sequência da proteína foi recuperada do banco de dados UniProt (UniProt ID: Q50EL4). Modelos estruturais teóricos foram construídos, por modelagem comparativa e *ab initio*. Os modelos preditos foram avaliados e comparados com as estruturas da CDCA1 presentes no Protein Data Bank. Os modelos foram otimizados e validados por diversos algoritmos, de modo a obter a melhor estrutura possível para a dinâmica molecular. A dinâmica foi realizada com sucesso, até a estabilização da estrutura, para a obtenção de dados significativos. Os dados da dinâmica apontaram variações na flexibilidade da estrutura da proteína. Esse resultado pode ser relacionado com a função de mobilidade correlacionada com o seu papel enzimático. Por ser uma proteína específica para a situação de deficiência de Zn, a análise de conservação filogenética mostrou que, apesar de muitos resíduos não serem analisados, devido à insuficiência de dados comparativos, as partes analisadas mantiveram considerável conservação. Compreender a relação da anidrase carbônica dependente de cádmio na *Thalassiosira weissflogii* abre o caminho para estudos filogenéticos a fim de trazer novas pistas sobre a evolução do metabolismo do cádmio em outras espécies que absorvem cádmio, mas não o utilizam em seu metabolismo.

Palavras chave: *Thalassiosira weissflogii*; anidrase carbônica; cádmio; zinco

Abstract

Carbonic anhydrase is an enzyme present in all organisms that are part of the evolutionary tree of archaea to mammals. This enzyme contains zinc and catalyzes the reversible hydration of CO₂. A novel carbonic anhydrase (CA), named CDCA1, was discovered in the marine diatom *Thalassiosira weissflogii*. CDCA1 is a cambialistic enzyme since it naturally uses Cd²⁺ as catalytic metal ion, but if it's necessary it can spontaneously exchange Cd²⁺ to Zn²⁺. The existence of a cadmium enzyme in marine phytoplankton may indicate that there is a unique selection of pressure for metalloenzymes in the marine environment, and our discovery provides a long-awaited explanation for the nutrient-like behaviour of cadmium in the oceans. In this study, CDCA1 was exposed by *in silico* analysis using different algorithms to obtain the most consistent structural model to determine their structural characteristics, enabling a phylogenetic relationship with other species that absorb cadmium. The protein sequence was retrieved from UniProt database (UniProt ID: Q50EL4). Structural theoretical models were created for the variants through comparative and *ab initio* modeling. The predicted models were evaluated using TM-align and compared with CDCA1 structure in the Protein Data Bank. The models were optimized and validated by many algorithms, obtaining the better possible structure to the molecular dynamics. The dynamics was successfully made, until the structure stabilization, in order to obtain significative results. The dynamics data pointed to variations in the flexibility of the protein structure. This result may be related to the mobility function correlated with its enzymatic role. Because it is a specific protein for the Zn deficiency situation, the phylogenetic conservation analysis showed that although many residues were not analyzed due to insufficient comparative data, the analyzed parts maintained considerable conservation. Understanding the relationship of Cadmium-Containing Carbonic Anhydrase in *Thalassiosira weissflogii* paves the way for phylogenetic studies in order to bring new insights for the evolution of cadmium metabolism in other species that absorb cadmium, but do not use it on its own metabolism.

Keywords: *Thalassiosira weissflogii*; carbonic anhydrase; cadmium; zinc

Lista de Abreviaturas

(GS) ₂ Cd	Bis glutionato de Cádmio
BER	Do inglês, <i>base excision repair</i>
B-factor	Fator Termodinâmico
CA	Anidrase Carbônica
Cad2	Transportador de Cádmio
CDCA1	Anidrase Carbônica dependente de Cádmio
C-score	Escore de Confiança
DFG	Do alemão, <i>Deutsche Forschungsgemeinschaft</i>
GSH	Tripeptídeo glutationa
HCO ₃ ⁻	Bicarbonato
IARC	Do inglês, <i>International Agency for Research on Cancer</i>
MD	Do inglês, Molecular Dynamics
MMM	Do inglês, Multiple Method Mapping
MMR	Do inglês, <i>mismatch repair</i>
MT	Metalotioneína
<i>NCE103</i>	Gene da Levedura <i>Saccharomyces cerevisiae</i>
NER	Do inglês, <i>nucleotide excision repair</i>
NTP	Do inglês, <i>National Toxicology Program</i>
pCO ₂	Pressão Parcial de Dióxido de Carbono
ProSA-web	Do inglês, Protein Structure Analysis
RMSD	Do inglês, Root-Mean-Square Deviation
RMSF	Do inglês, <i>Root Mean Square Fluctuation</i>
ROS	Do inglês, <i>Reactive Oxygen Species</i>
SASA	Do inglês, <i>Solvent Accessible Surface Area</i>
SH	Grupamento sulfidríla
SNPs	Do inglês, Single Nucleotide Polymorphisms
TM-score	Do inglês, Template Modeling score
TWCA1	Anidrase carbônica da diatomácea <i>Thalassiosira weissflogii</i>
USGS	Do inglês, United States of Geological Survey
<i>Δnce103</i>	Mutante do Gene da Levedura <i>Saccharomyces cerevisiae</i>

Lista de Figuras

Figura 1 - Modelos moleculares da CDCA1 obtidos por modelagem comparativa.	40
Figura 2 - Modelos moleculares da CDCA1 obtidos por modelagem ab initio.	41
Figura 3 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor (PS) ² .	43
Figura 4 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor M4T.	44
Figura 5 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor SwissModel.	44
Figura 6 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor I-Tasser.	45
Figura 7 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor Rosetta.	46
Figura 8 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor RaptorX.	47
Figura 9 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor Phyre2.	47
Figura 10 - Modelos moleculares da CDCA1 refinados pelo servidor GalaxyWeb, construídos a partir do modelo 1 do Rosetta.	49
Figura 11 - Modelo molecular da CDCA1 refinado pelo servidor KoBaMIN, construído a partir do modelo 1 do Rosetta.	49
Figura 12 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos pelo refinamento no GalaxyWeb do modelo do Rosetta.	50
Figura 13 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos pelo refinamento no KoBaMIN do modelo do Rosetta.	51
Figura 14 - Validação da estrutura da CDCA1 pelo servidor ProSA-Web.	53
Figura 15 - Validação da estrutura da CDCA1 pelo PROCHECK no gráfico de Ramachandran.	54
Figura 16 - Validação da estrutura da CDCA1 pelo QMEAN.	55
Figura 17 - Validação da estrutura da CDCA1 pelo Verify 3D.	55

Figura 18 – Análise filogenética da proteína CDCA1 mediante submissão do modelo 1 do Rosetta no servidor Consurf. _____	56
Figura 19 – RMSD da proteína CDCA1 ao longo da simulação de dinâmica molecular. _____	58
Figura 20 – RMSF da proteína CDCA1 ao longo da simulação de dinâmica molecular. _____	59
Figura 21 – Representação visual do fator térmico da estrutura nativa da CDCA1.	60
Figura 22 – Raio de giro ao longo da simulação de dinâmica molecular. _____	61
Figura 23 – Pontes de hidrogênio Intermoleculares formadas ao longo da simulação de dinâmica molecular. _____	62
Figura 24 – Superfície acessível ao solvente ao longo da simulação de dinâmica molecular. _____	63

Lista de Tabelas

Tabela 1 - Sequência e estruturas da CDCA1.	39
Tabela 2 - Compilação dos valores de RMSD e TM-Score do(s) modelo(s) gerado(s) por cada algoritmo.	42
Tabela 3 - Comparação dos valores dos servidores de validação estrutural dos modelos gerados por cada algoritmo.	48
Tabela 4 - Comparação dos valores dos servidores de validação estrutural dos modelos gerados após o refinamento estrutural.	51

Sumário

1. Introdução	15
1.1. ANIDRASE CARBÔNICA.....	15
1.1.1. ANIDRASE CARBÔNICA LIGANTE DE CÁDMIO.....	17
1.2. METAIS ESSENCIAIS E NÃO ESSENCIAIS.....	19
1.2.2 ZINCO E SAÚDE.....	20
1.2.3 TOXICIDADE DO ZINCO.....	21
1.3. CÁDMIO.....	21
1.3.1. CÁDMIO E SAÚDE.....	22
1.3.2. CÁDMIO E BIOTECNOLOGIA AMBIENTAL.....	25
1.3.3. CÁDMIO EM <i>SACCHAROMYCES CEREVISIAE</i>	25
1.4. BIOINFORMÁTICA.....	27
2. Objetivos	31
3. Metas	31
4. Material e Métodos	31
4.1. OBTENÇÃO DA SEQUÊNCIA E ESTRUTURA DA CDCA1.....	31
4.2. MODELAGEM MOLECULAR COMPUTACIONAL.....	31
4.3. ALINHAMENTO ESTRUTURAL E VALIDAÇÃO DOS MODELOS.....	34
4.4. VALIDAÇÃO DE ESTRUTURAS.....	35
4.5. REFINAMENTO ESTRUTURAL.....	36
4.6. ANÁLISE FILOGENÉTICA ESTRUTURAL.....	37
4.7. DINÂMICA MOLECULAR.....	37
5. Resultados e Discussões	38
5.1. OBTENÇÃO DA SEQUÊNCIA E ESTRUTURA DA CDCA1.....	38
5.2. MODELAGEM MOLECULAR COMPUTACIONAL.....	39
5.3. ALINHAMENTO ESTRUTURAL E VALIDAÇÃO DOS MODELOS.....	41
5.4. VALIDAÇÃO E REFINAMENTO DE ESTRUTURAS.....	48
5.5. ANÁLISE FILOGENÉTICA ESTRUTURAL.....	56
5.6. DINÂMICA MOLECULAR.....	57
5.6.1. <i>RMSD</i>	57
5.6.2. <i>RMSF</i>	58
5.6.3. <i>Fator B</i>	59
5.6.4. <i>Raio de Giro</i>	60
5.6.5. <i>Pontes de hidrogênio intermoleculares</i>	61
5.6.6. <i>Superfície acessível ao solvente</i>	62
5.6.7. <i>Análise da dinâmica molecular</i>	63
6. Conclusão	64
7. Referências Bibliográficas	65

1. Introdução

1.1. Anidrase Carbônica

A anidrase carbônica (CA) é uma metaloenzima, que é composta de um único polipeptídeo sem pontes dissulfeto e contém um íon Zn^{+2} necessário para a atividade catalítica (LINDSKOG, 1997). Esta enzima catalisa a hidratação reversível do CO_2 em bicarbonato com alta eficiência: $CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$ (SHEKH et al., 2012). Por ser produzida por uma variedade de tecidos, a anidrase carbônica desempenha um papel importante em muitas funções fisiológicas que envolvem reações de carboxilação e descarboxilação, incluindo a fotossíntese, respiração (HENRY, 1996), equilíbrio ácido-básico, transporte de íons e de dióxido de carbono, reabsorção óssea, ureogênese, gluconeogênese, lipogênese, geração de fluidos corporais (BAYRAM et al., 2008), calcificação (SMITH; FERRY, 2000), formação de humor aquoso (CENTOFANTI et al., 1997) e também participa do transporte de carbono inorgânico (C_i) (HENRY, 1996).

A anidrase carbônica foi primeiramente descoberta em eritrócitos bovinos em 1933 (MELDRUM; ROUGHTON, 1933; STADIE; ALLISON, 1928), mas desde então tem sido encontrada em muitos organismos, incluindo animais, plantas, arqueobactérias, algas e eubactérias (HEWETT-EMMETT; TASHIAN, 1996).

A importância industrial da anidrase carbônica encontra-se na ação como alvo para produção de fármacos diuréticos, que atuam no rim, aumentando o volume e diminuindo a concentração da urina (inibidores da anidrase carbônica) (DURÃO, 1979). Esta inibição permite reduzir a secreção de H^+ para o túbulo proximal do rim, evitando a reabsorção de sódio. Este fato justifica-se pela existência de um transporte ativo que ocorre entre o H^+ (em direção ao túbulo) e o Na^+ (em direção à célula). O sódio vai ser mais excretado e, uma vez que a reabsorção de água está diretamente dependente da de sódio, aumenta também a excreção de água (DURÃO, 1979). Outra importância industrial da anidrase carbônica (CA) é na aplicação em sistemas de captura enzimática de CO_2 atmosférico (BOND et al., 2001; DAVY, 2009; DILMORE et al., 2009; FAVRE; CHRIST; PIERRE, 2009; MIRJAFARI; ASGHARI; MAHINPEY, 2007; VINOBA et al., 2012), poluente

considerado como maior contribuinte para o efeito estufa (HOUGHTON JT et al., 2001). O aumento do aquecimento global tem estimulado cada vez mais o desenvolvimento de tecnologias eficazes para mitigação do CO₂. Dentro deste contexto, os sistemas biológicos, utilizando a enzima anidrase carbônica, vêm ganhando cada vez mais atenção. Estes sistemas são baseados em reações que ocorrem naturalmente nos organismos vivos e são uma via potencial para a melhoria destas tecnologias (FIGUEROA et al., 2008).

Baseado na estrutura proteica, as anidrases carbônicas são divididas em quatro classes: α , β , γ e δ (SMITH et al., 1999). Embora universal, a classe α é mais bem caracterizada em mamíferos, visto que isoenzimas da classe β foram descritas em muitas plantas, algas, fungos, archaea e bactérias. Membros conhecidos da classe γ são somente os procariontes e a classe δ tem somente um membro, a anidrase carbônica da alga *Thalassiosira weissflogii* (AGUILERA et al., 2005).

Considerando que as classes α e γ usam três resíduos de histidina para coordenar o átomo de zinco, e as classes β e δ de anidrases carbônicas usa dois resíduos de cisteína e um de histidina (STROP et al., 2001; TRIPP; SMITH; FERRY, 2001). As diferenças entre as estruturas terciárias e quaternárias entre as quatro classes constituem um caso espetacular de convergência evolutiva, na qual a mesma atividade catalítica evoluiu de diversas maneiras independentes evolutivamente muito antigas. As anidrases carbônicas de mamíferos (incluindo as 10 isoformas humanas ativas) junto com as duas enzimas periplasmáticas da alga verde unicelular *Chlamydomonas reinhardtii*, pertencem à classe (α). A classe (β) é composta de enzimas do cloroplasto das plantas monocotiledôneas e dicotiledôneas. Dentro do domínio bacteriano, as enzimas purificadas de *Neisseria gonorrhoeae* e *Escherichia coli* também pertencem às classes (α) e (β), respectivamente. Estudos indicaram que as anidrases carbônicas não somente são amplamente distribuídas em procariontes, mas que as classes (β) e (γ) são predominantes (SMITH et al., 1999). Isso sugere que a interconversão de CO₂ e bicarbonato, catalisada pela anidrase carbônica é importante para todas as formas de vida.

No que diz respeito às suas origens independentes, partilham a propriedade comum de ligação a um átomo de zinco em seus sítios ativos que é

importante para a atividade catalítica. Desse modo, os organismos empregam um mecanismo enzimático similar composto de duas etapas. A primeira etapa compreende o ataque nucleofílico de um íon hidróxido ligado ao zinco no CO₂. A segunda etapa é a regeneração do sítio ativo pela ionização da molécula de água ligada ao zinco e a remoção de um próton do sítio ativo. (AMOROSO et al., 2005). A propriedade comumente mais notável entre as anidrases carbônicas caracterizadas é a presença de um íon zinco, que é essencial para a atividade catalítica. De longe as mais estudadas e melhor compreendidas das anidrases carbônicas são as dos mamíferos (COX et al., 2000).

Embora a publicação de sequências genômicas de diversas espécies de fungos tem indicado a presença da anidrase carbônica como sequência, até agora anidrases carbônicas não funcionais foram caracterizadas em detalhes para esse grupo de organismos. Talvez o melhor exemplo para isso seja o gene *NCE103* da levedura *Saccharomyces cerevisiae* que codifica um polipeptídeo que mostra uma razoável homologia a membros da classe β (AMOROSO et al., 2005). Estudos mostraram que a transcrição do gene *NCE103* é inibida em concentrações elevadas de CO₂, independentemente das concentrações de O₂ presente nas culturas (AGUILERA et al., 2005). A inativação do gene *NCE103* resulta em um mutante (*Δnce103*) incapaz de crescer aerobicamente (AMOROSO et al., 2005), no entanto, estudos recentes mostraram que o crescimento aeróbico do *Δnce103* era possível quando o ar era enriquecido com CO₂ (AGUILERA et al., 2005).

1.1.1. Anidrase carbônica ligante de cádmio

Os perfis da concentração de cádmio nos oceanos mostram que este metal é utilizado na coluna d'água como um nutriente das algas: o cádmio se encontra em quantidades empobrecidas na superfície pela captação dos fitoplânctons e nas profundezas se encontra em maior quantidade devido à remineralização do "naufrágio" de matéria orgânica (BOYLE; SCLATER; EDMOND, 1976; BRULAND; KNAUER; MARTIN, 1978). Parte da explicação para este comportamento como nutriente é o uso do cádmio como um átomo de metal catalítico na anidrase carbônica em algas marinhas (LANE; MOREL, 2000; PRICE; MOREL, 1990). Estes organismos têm se adaptado à vida em um meio contendo

concentrações infimamente pequenas de metais essenciais. A expressão da anidrase carbônica ligante de cádmio (CDCA) é um notável exemplo desta evolução (XU et al., 2008).

A enzima anidrase carbônica na alga *Thalassiosira weissflogii* (TWCA1), constitui o principal uso do zinco celular, e a expressão de TWCA1 são também moduladas pela pCO₂ no meio, e está claro que a enzima desempenha um papel na aquisição de carbono inorgânico (LANE; MOREL, 2000).

Estudos com a alga *Thalassiosira weissflogii* sugerem fortemente que o cádmio pode desempenhar um papel direto na anidrase carbônica. A adição de cádmio em culturas zinco-limitadas não somente realça sua taxa de crescimento, mas resulta numa expressão aumentada de uma anidrase carbônica, que é claramente distinta da TWCA1 dependente de zinco (LANE; MOREL, 2000). A existência de uma enzima de cádmio nos fitoplânctons marinhos pode indicar que existe uma pressão de seleção única para metaloenzimas no ambiente marinho, e essa descoberta fornece uma explicação há muito aguardada para o comportamento do cádmio como nutriente nos oceanos (LANE et al., 2005).

Em algas marinhas, os metais cádmio, cobalto e zinco podem substituir funcionalmente um ao outro para manter ótimas taxas de crescimento. Esse efeito é, no mínimo, parcialmente devido à substituição de carbono inorgânico pela fotossíntese. Em adição a uma anidrase carbônica de zinco que pode substituir o cobalto no seu sítio ativo *in vivo*, a *T. weissflogii* tem uma já mencionada anidrase carbônica de cádmio que é também envolvida na aquisição de carbono inorgânico.

A proteína CDCA1 foi purificada da alga *Thalassiosira weissflogii* que tem atividade de anidrase carbônica e contém cádmio. A determinação de sua sequência foi complicada pela presença de uma tripla repetição: as três sequências de aminoácidos são virtualmente idênticas, mas existe uma maior variação no DNA. Esta enzima, presente em algumas algas, faz o intercâmbio de metais em seu centro catalítico e possui melhor atividade com o cádmio do que outras anidrases carbônicas. Estas propriedades presumivelmente evoluíram em resposta à baixa quantidade de metais nos ambientes oceânicos. Porém, dados de sua cinética mostram que a substituição do zinco pelo cádmio resulta, todavia, em uma diminuição da eficiência catalítica (XU et al., 2008). A proteína CDCA1 é

significativamente diferente de qualquer das principais classes de anidrases carbônicas conhecidas, e conseqüentemente, representa o primeiro membro de uma nova classe de anidrases carbônicas, a classe ζ (LANE et al., 2005)

A notável capacidade para fazer uso de um elemento previamente conhecido como sendo tóxico é, presumivelmente, uma significativa vantagem competitiva para as algas no ambiente pobre em metais dos oceanos. A anidrase carbônica é uma enzima-chave na maquinaria de captação de carbono destes organismos, que são responsáveis por uma grande fração da exportação de carbono proveniente da atmosfera para o fundo dos oceanos. O uso bioquímico do cádmio pode deste modo, ter desempenhado um papel na radiação global das algas durante a Era Cenozoica e, concomitantemente, ter diminuído o CO₂ atmosférico (XU et al., 2008).

O transporte de metais como cobre, manganês e zinco para o interior das células são justificados, pois muitas enzimas têm esses metais como cofatores. No entanto, o transporte de cádmio para o interior das células permanecia um paradoxo, pois o cádmio não era cofator de nenhuma enzima conhecida até o recente achado da nova classe ζ da anidrase carbônica da alga marinha *Thalassiosira weissflogii* que possui cádmio em vez de zinco em seu sítio ativo. Este achado abriu o caminho para estudos filogenéticos a fim de trazer novas pistas sobre a evolução do metabolismo do cádmio em outras espécies que absorvem cádmio, mas não o utilizam em seu metabolismo.

1.2. Metais essenciais e não essenciais

Os metais são uma parte integrante do meio ambiente e estão amplamente difundidos na natureza, e os organismos tornam-se expostos aos metais através de fontes naturais ou antropogênicas. Embora compreendam 75% dos elementos da tabela periódica, poucos são os metais essenciais à vida, sendo a maioria tóxicos para os organismos (BALLATORI, 2002). Os metais essenciais exercem uma série de funções importantes à fisiologia celular. Eles podem atuar como catalizadores de reações bioquímicas, estabilizadores de estruturas proteicas e paredes celulares, e ainda como reguladores do equilíbrio osmótico intracelular (CHANG; LEU, 2011). Assim são requeridos em pequenas quantidades e vitais ao

desenvolvimento dos organismos, porém acima de uma determinada concentração fisiológica tornam-se tóxicos e quando insuficientes causam deficiências nos processos biológicos. Assim, as concentrações intracelulares de metais essenciais devem ser mantidas dentro de um intervalo estreito. De maneira contrária, os metais não essenciais não possuem papel biológico e são tóxicos mesmo em pequenas concentrações (SHAKIR; SHAKEEL; QAZI, 2013). Em particular, o Cádmiio é um metal de transição altamente tóxico, conhecido como um poluente industrial e ambiental e com exposições relevantes à população em geral e à população ocupacional. Salvo seu uso como cofator da enzima anidrase carbônica, presente na diatomácea marinha *Thalassiosira weissflogii*, o cádmio não possui nenhum outro papel biológico conhecido, sendo considerado por isso, majoritariamente um metal não essencial (SERERO et al., 2008).

1.2.1 Zinco

O zinco é um dos elementos mais comuns da crosta terrestre. É encontrado no ar, solo e água e está presente em quase todos os alimentos. Tem ponto de fusão e ebulição, respectivamente, iguais a 419,5°C e 908°C, e sua densidade é de 7,14g/m³. O ar seco não o ataca, porém, na presença de umidade, forma uma capa superficial de óxido ou carbonato básico que isola o metal e o protege da corrosão. Reage com ácidos não oxidantes passando para o estado de oxidação +2 e liberando hidrogênio, e pode dissolver-se em bases e ácido acético (ATSDR, 2005).

1.2.2 Zinco e saúde

O zinco é um metal essencial, sendo que tanto a ausência quanto a exposição excessiva podem trazer efeitos nocivos (GOYER; CLARKSON, 2001). O zinco intervém no metabolismo de proteínas e ácidos nucleicos, estimula a atividade de mais de 100 enzimas, colabora no bom funcionamento do sistema imunológico e na síntese do DNA. O metal é encontrado na insulina, nos motivos de dedo de zinco (zinc finger) e em diversas enzimas como a anidrase carbônica (ATSDR, 2005).

Sua deficiência pode produzir retardamento no crescimento, perda de cabelo, diarreias, impotência sexual, lesões oculares e de pele, perda de apetite,

perda de peso, aumento do tempo de cicatrização de ferimentos e anomalias no sentido do olfato. As causas que podem provocar uma deficiência de zinco são a insuficiente quantidade na dieta alimentar e a dificuldade na absorção do mineral, que pode ocorrer em casos de etilismo, quando é eliminado pela urina ou, ainda, devido à excessiva eliminação por causa de desordens digestivas. O excesso de zinco tem sido associado a baixos níveis de cobre, alterações na função do ferro, diminuição da função imunológica e dos níveis de colesterol bom (ATSDR, 2005).

O zinco metálico não é considerado tóxico; porém, alguns de seus compostos, como o óxido e o sulfeto, são nocivos. Caso altas concentrações de zinco sejam ingeridas em um curto período de tempo, dores estomacais, náuseas e vômitos podem ocorrer. Ingerir altas concentrações de zinco por vários meses pode causar anemia, danos ao pâncreas e problemas sanguíneos. Não se sabe se altos níveis de zinco afetam a fertilidade ou causam efeitos no feto em humanos (ATSDR, 2005).

1.2.3 Toxicidade do Zinco

Grandes quantidades de zinco entram no meio ambiente como resultado de atividades antropogênicas, como mineração, purificação do zinco, chumbo, cádmio, produção de aço, queima de carvão e de lixo. Estes processos podem aumentar a concentração do metal na atmosfera. Lixos de indústrias químicas que utilizam zinco, esgoto doméstico e correntes de água de solos contendo zinco podem transportar o metal para as águas. No ar, o zinco está presente na forma de partículas muito finas. Estas partículas se distribuem sobre a terra e águas. O nível de zinco dissolvido na água pode aumentar à medida que a acidez da água aumenta. Alguns peixes podem bioacumular o zinco e outros animais que se alimentam destes ou de água e alimentos contaminados também podem acumular o metal em seus organismos (ATSDR, 2005).

1.3. Cádmio

De forma convencional na literatura, o termo “Metal Pesado” implica metais tóxicos de elevada densidade atômica, e dessa forma concorda com a literatura científica mais antiga. O Cádmio apresenta uma densidade de $8,6 \text{ g/cm}^3$,

além de um número atômico 48 e peso atômico 112,4 (IARC, 2012), e assim, de acordo com as exigências apresentadas acima, mesmo que mutáveis, o Cádmió nunca deixou de ser considerado um metal pesado.

O Cádmió (Cd) é um metal de transição que pertence ao grupo 12 da tabela periódica, foi descoberto na Alemanha em 1817 pelo químico F. Strohmeier, e ocorre naturalmente na crosta terrestre e nos oceanos, com uma abundância média em torno de 0,1-0,2 mg/Kg e <5-100 ng/L, respectivamente. É o 64º elemento mais abundante do planeta, e na natureza é encontrado principalmente em associação com minérios de zinco, chumbo e cobre. Além de sua ocorrência natural, o Cádmió pode ser emitido ao ambiente por atividades naturais ou antropogênicas. O intemperismo e erosão de rochas portadoras de cádmio são as principais emissões naturais. Outras fontes importantes são atividades vulcânicas e incêndios florestais. Em respeito às atividades antropogênicas, o Cádmió apresenta propriedades específicas interessantes às indústrias, e quando não necessário como matéria-prima, ainda pode estar presente na forma de impureza, não exercendo nenhum papel funcional durante o processo industrial (IARC, 2012; ICDA, 2011). Assim, o Cádmió está presente em uma ampla variedade de atividades industriais fundamentais, como mineração, metalurgia, galvanoplastia, ligas, estabilizadores, baterias, fertilizantes, têxtil, couro, combustíveis fósseis, pigmentos e tintas (BERTIN; AVERBECK, 2006; FILIPIC, 2012; MOHAMMADIAN FAZLI et al., 2015; PRADHAN et al., 2014). Segundo o Serviço Geológico dos Estados Unidos - USGS (do inglês, *United States of Geological Survey*), em 2016, mais de 80% do Cádmió industrial foi destinado para o uso em baterias, e os 20% restantes, em ordem decrescente, foram destinados à indústria de pigmentos, galvanização, estabilizadores, ligas e outras (OBER, 2016).

1.3.1. Cádmió e saúde

De acordo com a exposição ao Cádmió, a população humana pode ser dividida em dois grupos, exposição ocupacional e exposição à população em geral. A exposição ocupacional diz respeito aos locais fortemente contaminados como as áreas vizinhas às plantas industriais, e neste sentido, a contaminação humana ocorre principalmente pela água potável e principalmente pelo ar. Fora destes locais,

a exposição se dá pela ingestão de alimentos e pelo cigarro. Fumantes possuem uma concentração sanguínea de Cádmio em torno de 4-5 vezes maior do que os não fumantes, além de uma concentração renal aumentada em cerca de 2-3 vezes (SATARUG; MOORE, 2004; JÄRUP; ÅKESSON, 2009). A toxicidade do Cádmio foi revelada já em 1955 no Japão, com a doença de Itai-Itai, e pela primeira vez, a poluição por Cádmio mostrou ter consequências graves para a saúde humana. De fato, em 1968, o governo japonês reconheceu a importância da poluição ambiental para o desenvolvimento da doença. O Cádmio, liberado de uma mina nas montanhas, foi transportado pelo rio Jinzu para a planície, onde a água era utilizada para irrigação de campos de arroz. A planta absorveu o Cádmio a partir do solo e o consumo deste arroz contaminado intoxicou a população. Os efeitos mais importantes da doença são as lesões renais, deficiências imunológicas e lesões ósseas (BERTIN; AVERBECK, 2006; NORDBERG, 2009).

Está estabelecido na literatura que mesmo a população não ocupacional apresenta Cádmio em seus órgãos, sendo que o excesso desta exposição produz diversos efeitos adversos para a saúde dos seres humanos, tendo o rim como o alvo mais crítico (SATARUG et al., 2010). Além disso, o Cádmio possui uma meia-vida longa, de 30 anos, e uma taxa de excreção diária lenta, aproximadamente 0,001%, características que facilitam seu acúmulo no corpo humano e o aparecimento de lesões. As manifestações de nefrotoxicidade por Cd incluem proteinúria, calcinúria, aminoacidúria, glicosúria e necrose tubular (SATARUG; MOORE, 2004).

O Cádmio possui ainda uma elevada carcinogenicidade, reconhecida pela Agência Internacional de Pesquisa ao Câncer - IARC (do inglês, *International Agency for Research on Cancer*), Programa de Toxicologia Nacional dos Estados Unidos - NTP (do inglês, *National Toxicology Program*) e Comunidade Alemã de Pesquisa - DFG (do alemão, *Deutsche Forschungsgemeinschaft*)(DFG, 2006; IARC, 2012; NTP, 2000). Dados da literatura indicam diversos tipos de cânceres, como próstata, mama, útero, pulmão e pâncreas, além de outros tipos de lesões, como desordem neurológica, osteoporose, enfisema, anemia, eosinofilia, insônia, hipertensão, dano à placenta, fígado, estômago, sistema hematopoiético e bexiga (MICHAEL P, 2000; PESCH et al., 2000; SCHWARTZ; REIS, 2000; HU et al., 2002; GOYER; LIU; WAALKES, 2004; JÄRUP; ÅKESSON, 2009; NORDBERG, 2009). A

nível molecular, o efeito da intoxicação por Cádmio resulta de dois mecanismos, a sua alta afinidade pelo grupamento sulfidríla (SH) e a substituição de cofatores proteicos. Como consequência, o dano celular ocorre devido à inibição dos sistemas de reparo do DNA e aumento do estresse oxidativo (BERTIN; AVERBECK, 2006).

Sabendo que o Cádmio possui afinidade pelo grupamento sulfidríla, ele interfere em proteínas que contenham o motivo dedo de zinco, o qual participa da manutenção da estabilidade do genoma, do reparo e da sinalização de dano no DNA. Entre os sistemas de reparo, o Cádmio interfere em três vias, reparo de mal pareamento - MMR (do inglês, *mismatch repair*), reparo por excisão de base - BER (do inglês, *base excision repair*), e reparo por excisão de nucleotídeo - NER (do inglês, *nucleotide excision repair*). O NER é responsável por reparar uma variedade de lesões induzidas por produtos químicos genotóxicos. O BER remove a maioria das lesões resultantes de danos endógenos e o MMR repara o erros de mal pareamento gerados durante a replicação e recombinação do DNA. As proteínas dedo de zinco compreendem uma família de proteínas onde o zinco é complexado através de quatro resíduos conservados de cisteína ou histidina, formando um domínio dedo de zinco. Este domínio está envolvido principalmente na ligação ao DNA, mas também nas interações proteína-proteína (HARTWIG, 2010).

Quanto ao estresse oxidativo, o cádmio afeta o equilíbrio redox do grupamento sulfidríla por meio da diminuição do conteúdo intracelular de moléculas antioxidantes, como a glutatona, e a redução da atividade de enzimas antioxidantes celulares, como a superóxido dismutase, peroxidase e catalase, o que resulta no acúmulo intracelular de espécies reativas de oxigênio - ROS (do inglês, *Reactive Oxygen Species*). Outro mecanismo inclui o deslocamento de metais, como ferro e cobre, de diferentes proteínas citoplasmáticas ou de membrana, aumentando a concentração intracelular destes íons, que podem participar na produção de ROS por meio da reação de Fenton (FILIPIC, 2012). Uma vez desencadeado pelo Cádmio, ROS pode reagir com várias biomoléculas intracelulares e levar à mutações no DNA, alterações na função e estrutura de proteínas, mudanças na expressão gênica e apoptose (PRADHAN et al., 2014).

1.3.2. Cádmio e biotecnologia ambiental

Para regular a concentração de metais intracelulares, os organismos utilizam uma ampla variedade de mecanismos de homeostase e tolerância, que regulam a disponibilidade de metais essenciais e limitam os efeitos prejudiciais de elementos tóxicos. Como a poluição por metais pesados está piorando e se espalhando por todo o mundo, juntamente com o progresso industrial, em resposta, a maioria dos organismos adapta o seu metabolismo e desenvolve sistemas para limitar esta toxicidade e adquirir tolerância (ZHOU et al., 2013).

Neste sentido, destaca-se a biorremediação, uma técnica biológica considerada eficaz, ecológica e econômica, quando comparada às técnicas físico-químicas convencionais de remediação. Entende-se por biorremediação como um processo de restauração do meio ambiente que utiliza micro-organismos para neutralizar ou atenuar os contaminantes, e para isso baseia-se nas reações bioquímicas destes micro-organismos para degradar, modificar ou remediar as substâncias tóxicas (COUTINHO et al., 2015).

Estudar os mecanismos de acumulação do metal pesado Cd^{2+} e sua desintoxicação é um problema atual da pesquisa básica e aplicada, visto que o cádmio é um importante poluente ambiental que induz danos severos nos organismos vivos (BLAZHENKO et al., 2006).

1.3.3. Cádmio em *Saccharomyces cerevisiae*

A levedura *Saccharomyces cerevisiae* tem sido utilizada como organismo modelo para estudar os mecanismos de homeostase e resistência aos metais, e até o momento, quatro mecanismos distintos foram propostos. O primeiro deles é um mecanismo para diminuir o influxo de metais nas células, e depende da repressão do gene transportador do metal por íons metálicos intracelulares, ou por proteólise do transportador de membrana induzida por estes íons. Em segundo lugar, e mais amplamente conhecido, é o mecanismo de formação de complexos com os metais, como as metalotioneínas (MTs) e glutationas (GSH). Essas moléculas neutralizam o efeito tóxico dos íons metálicos livres por ligação com eles no ambiente intracelular. O terceiro mecanismo compreende a compartimentação em vacúolos destes íons, o

que limita sua concentração no citoplasma. Em quarto lugar está o efluxo destes metais pesados por meio de transportadores (SHIRAISHI et al., 2000).

Os metais não essenciais entram nas células com base no mimetismo molecular com os metais essenciais, através de permeases da membrana plasmática e proteínas canais, todos, necessários para a absorção de metais essenciais e outros nutrientes, como Ferro (Fe), Manganês (Mn), Zinco (Zn), fosfato, sulfato, glicerol e outros. Assim, o Cádmio entra nas células através destas proteínas transportadoras, tais como Zrt1p, Smf1p e Smf2p, Fet4p e Mid1p, que transportam, respectivamente, Zn, Mn, Fe e Cálcio (Ca). Em *Saccharomyces cerevisiae*, o transporte dos íons cádmio ocorre principalmente através da proteína Zrt1, e da transportadora de manganês, a Smf1 (GITAN et al., 2003). A principal molécula quelante de Cd²⁺ dentro da célula da levedura *S. cerevisiae* aparece sendo o tripeptídeo glutationa (GSH) (PENNINCKX, 2002).

Embora existam diferentes mecanismos de desintoxicação, aquele que proporciona o maior nível de tolerância, em microrganismos, é a remoção do metal através de vias de exportação (WYSOCKI; TAMÁS, 2010). No caso particular do Cádmio, quando este entra na levedura por meio de proteínas de membrana, rapidamente reage com duas moléculas de glutationa e gera o bis glutationato de Cádmio [(GS)₂Cd], e então a proteína Ycf1 medeia a acumulação vacuolar destes complexos, limitando assim as concentrações citoplasmáticas deste metal (GOMES et al., 2002; KERN et al., 2005). YCF1 é um transportador de membrana vacuolar, de *S. cerevisiae*, que catalisa o transporte de conjugados de glutationa (LI et al., 1997). Tipos mutantes que apresentam deficiência na proteína Ycf1 possuem altas taxas do complexo [(GS)₂Cd] no citoplasma devido a sua deficiência em transportar o complexo para o vacúolo, e apresentam taxas de mutação aumentadas (ADAMIS; PANEK; ELEUTHERIO, 2007). O controle do nível intracelular de cádmio em *S. cerevisiae* é mantido através da regulação hipotética do transporte de cádmio dependente de Zrt1 pelo complexo [(GS)₂Cd] (GOMES et al., 2002) e o efluxo extracelular de Cd²⁺ pelo transportador Cad2 (BLAZHENKO et al., 2006). As células que carregam o gene Cad2 exibem resistência ao cádmio. O gene Cad2 é uma mutação derivada da proteína PCA1, a qual exporta cobre por um transportador do tipo P-ATPase. Assim, o gene em questão controla os níveis intracelulares de

cádmio através de uma ativação no sistema de efluxo do cádmio (SHIRAIISHI et al., 2000). De acordo com a literatura, PCA1 é a principal rota de efluxo do Cádmio em *Saccharomyces cerevisiae* (ADLE et al., 2007), porém, outros transportadores de membrana parecem estar envolvidos na detoxificação de metal, mas pouco se sabe sobre a sua regulação ou modo de ação.

1.4. Bioinformática

A bioinformática é uma área de conhecimento multidisciplinar que emprega conceitos da biologia molecular, química, física e informática para o desenvolvimento de ferramentas computacionais aplicadas ao estudo de sistemas biológicos (VERLI, 2014). A forma de experimentação adotada pela bioinformática, embasada no uso de ferramentas computacionais (VERLI, 2014), é também conhecida como *in silico*, análogo aos termos *in vivo* e *in vitro* (EKINS; MESTRES; TESTA, 2007).

O emprego massivo de estratégias computacionais – e consequentemente, o desenvolvimento da bioinformática como área de conhecimento – acontece, sobretudo, a partir do projeto Genoma Humano, que, juntamente com a popularização de outros projetos genoma, criaram uma demanda crescente por armazenamento e análise do elevado número de sequências gênicas elucidadas (XIONG, 2006). Volume, este, que tende a aumentar ainda mais com os estudos de transcriptoma e metaboloma, exigindo cada vez mais avanços da bioinformática, seja em hardware, software ou em estratégias de análises de dados e construção de modelos (VERLI, 2014).

As ferramentas de simulações computacionais são imprescindíveis para obtenção e armazenamento de dados biológicos, incluindo, as sequências proteicas, que podem ser utilizadas para prever aspectos evolutivos, funcionais e estruturais de uma proteína recém descoberta (XIONG, 2006). Além disso, a partir da mesma sequência, é possível obter modelos *in silico* da estrutura tridimensional da proteína correspondente, que possui aplicações de grande impacto terapêutico e biotecnológico, como o desenho racional de fármacos e de novas proteínas com funções específicas (VERLI, 2014). Nesse sentido, é possível perceber que a

bioinformática possui inúmeras aplicações relevantes, que serão abordadas com maior nível de detalhamento nesta sessão.

A bioinformática se faz essencial para a gestão de dados biológicos, sobretudo mediante a grande quantidade de genomas eucarióticos e procarióticos completamente sequenciados e ainda por sequenciar. O armazenamento e organização de tais informações em banco de dados digitais constituem uma atribuição fundamental da bioinformática. Banco de dados biológicos de grande relevância, como NCBI (WHEELER et al., 2000), UniProt (PUNDIR et al., 2017) e Protein Data Bank (BERMAN et al., 2000) estão disponíveis para acesso livre de usuários, facilitando a busca por informações.

As sequências de proteínas e DNA são selecionadas ao longo de milhares de anos, acumulando mutações características, as quais constituem evidências evolutivas que permitem a identificação de ancestrais comuns. O alinhamento de sequências, por sua vez, permite a identificação de padrões de conservação que demonstram a relação evolutiva entre elas. Além disso, resíduos cruciais para estrutura e função da proteína tendem a ser preservados pela seleção natural – enquanto que resíduos de menor importância tendem a sofrer mutações com mais frequência – possibilitando, por exemplo, o reconhecimento de sítios ativos enzimáticos. A comparação entre sequências, nesse sentido, se faz essencial para a inferência de aspectos funcionais e evolutivos de novas proteínas a partir de proteínas já existentes nos bancos de dados (XIONG, 2006), viabilizando análises filogenéticas (ASHKENAZY et al., 2010) e predição funcional de SNPs (BROMBERG; ROST, 2007).

A bioinformática representa a solução para lidar com a defasagem entre a quantidade de sequências proteicas elucidadas e a estrutura tridimensional correspondente, tendo em vista que os métodos de sequenciamento estão muito mais avançados do que os métodos experimentais – como a Ressonância Magnética Nuclear e a Cristalografia de Raios-X – empregados na determinação de estruturas biomoleculares. Nesse sentido, ferramentas computacionais, se apresentam alternativas de menor custo, rápidas, além de, em alguns casos, significarem a única possibilidade de obtenção de modelos estruturais (VERLI, 2014).

Dentre os métodos computacionais empregados para a predição de estruturas, destacam-se a modelagem *ab initio* e a modelagem comparativa. A modelagem comparativa trata-se da técnica mais dependente de informações, necessitando de uma estrutura molde, resolvida experimentalmente, para gerar um modelo. Tal molde é selecionado de acordo com a taxa de identidade obtida entre a sequência alvo e os possíveis candidatos a molde disponíveis em banco de dados de sequência específicos. Já a modelagem *ab initio* independe de um molde, necessitando apenas da sequência de aminoácidos para predizer a estrutura. Esta técnica se baseia nas informações contidas na sequência de aminoácidos e o mínimo de energia livre global (VERLI, 2014).

A validação do modelo obtido ocorre a partir da verificação de sua qualidade, a qual é determinada por um conjunto de fatores físico-químicos, como comprimento e a energia de ligações, ângulos de torção e impedimento estérico. Além disso, o valor de Root Mean Square Deviaton (RMSD) – obtido a partir do alinhamento estrutural com o molde utilizado em sua confecção – deve ser levado em consideração para avaliar modelos gerados comparativamente. A qualidade do modelo pode ser verificada por meio de análises estereoquímicas, energéticas e funcionais. A análise estereoquímica avalia os aspectos tridimensionais da molécula, verificando sua estabilidade conformacional, por meio da detecção de regiões de tensão angular e torsional e impedimentos estéricos. Para tal, o gráfico de Ramachandran se apresenta como uma alternativa (VERLI, 2014). O mesmo avalia a correlação entre os ângulos phi e psi, permitindo a identificação de resíduos que não possuam arranjos espaciais factíveis (GOPALAKRISHNAN et al., 2007), os quais devem ser inspecionados criteriosamente. A análise de energia avalia quão próximo o modelo gerado está de proteínas que possuem o mesmo perfil molecular e de enovelamento, considerando parâmetros de estrutura secundária e relação sequência-estrutura (1D-3D). Por fim, as análises funcionais envolvem a comparação de um modelo obtido com aspectos funcionais determinados por métodos experimentais. Esses modelos tridimensionais de proteínas podem ser utilizados para o desenvolvimento de novos fármacos e dinâmica molecular (VERLI, 2014).

A bioinformática possui grande impacto no desenvolvimento de novos fármacos. Estudos computacionais da interação proteína-ligante provêm bases racionais para a identificação de candidatos a fármacos, uma vez que, o conhecimento da estrutura tridimensional de proteínas permite que moléculas sejam desenhadas com a capacidade de se ligar ao receptor no sítio alvo com grande afinidade e especificidade. A abordagem computacional reduz significativamente o custo e tempo necessários para o desenvolvimento de drogas com grande potencial, menos efeitos adversos e menos toxicidade quando comparada a abordagem tradicional de tentativa e erro (XIONG, 2006).

Além disso, ferramentas computacionais permitem o desenvolvimento de dinâmica molecular (MD – *Molecular Dynamics*), método que consiste na simulação de movimentos atômicos individuais ou moleculares conforme a segunda lei de Newton. Nesse sentido, tal procedimento descreve o comportamento molecular em função do tempo, viabilizando a caracterização de propriedades variáveis, como estrutura secundária, conformação tridimensional, orientação das cadeias laterais e energia de interação entre moléculas. A MD se aproxima em muito de métodos experimentais como a Ressonância Magnética Nuclear, por gerar médias representativas do comportamento de biomoléculas, tornando-as muito mais próximas da realidade biológica. Tal propriedade da dinâmica molecular faz com que esta seja usada em diversas metodologias de refinamento de estruturas tridimensionais. Além disso, a MD descreve com sucesso características como a flexibilidade molecular, que está associada à maioria dos fenômenos biológicos, como catálise de reações, modulação de canais iônicos e receptores acoplados a proteína G (VERLI, 2014).

A anidrase carbônica é alvo para drogas, com isso, ter um modelo de alta resolução e compreender a flexibilidade da proteína, pode ser importante para o desenho racional de fármacos. Nesse sentido, o presente trabalho aplica simulações computacionais ao estudo da proteína CDCA1, utilizando metodologias de modelagem, alinhamento e validação estrutural, dinâmica molecular e análise filogenética para caracterizar os aspectos funcionais e estruturais desta molécula.

2. Objetivos

Analisar a estrutura proteica da anidrase carbônica dependente de cádmio na *Thalassiosira weissflogii*, através de experimentos *in silico*, a fim de determinar suas características estruturais, possibilitando uma relação filogenética com outras espécies que absorvem cádmio, mas não o utilizam em seu metabolismo.

3. Metas

- I. Criar modelos teóricos da estrutura tridimensional (3D) da proteína, através de modelagem comparativa e *ab initio*;
- II. Analisar, através de alinhamentos estruturais, o melhor modelo teórico da proteína gerado e validar o modelo em questão;
- III. Determinar o nível de conservação evolutiva da CDCA1;
- IV. Realizar e analisar a Dinâmica Molecular do modelo obtido.

4. Material e Métodos

4.1. Obtenção da sequência e estrutura da CDCA1

A sequência e as estruturas experimentais da CDCA1 foram obtidas no banco de dados UniProt (PUNDIR et al., 2017). Todos os modelos experimentais obtidos por difração de raio X foram escolhidos (PDB ID 3BOB, 3BOC, 3BOE, 3BOH, 3BOJ, 3UK8).

4.2. Modelagem molecular computacional

A sequência da CDCA1 no formato fasta foi submetida aos algoritmos Rosetta (KIM; CHIVIAN; BAKER, 2004) e I-Tasser (YANG; ZHANG, 2015) para construção dos modelos 3D teóricos da CA pela técnica *ab initio*. Para geração de modelos por modelagem comparativa, a sequência no formato fasta foi submetida ao algoritmo (PS)² (HUANG et al., 2015), e aos algoritmos M4T (RYKUNOV et al., 2009), Phyre2 (KELLEY; STERNBERG, 2009), RaptorX (KÄLLBERG et al., 2012) e

Swiss-Model (BIASINI et al., 2014), que estão disponíveis no servidor Protein Model Portal (HAAS et al., 2013).

O servidor Robetta emprega métodos automatizados de predição estrutural para gerar modelos tridimensionais completos de proteínas a partir da sequência de aminoácidos, com ou sem o uso de estruturas moldes conhecidas. O servidor utiliza, inicialmente, um método chamado GinzU – que consiste na utilização do BLAST, PSI-BLAST, FFAS03 e 3D-Jury – para detectar, por meio do alinhamento múltiplo de sequências, regiões homólogas a estruturas determinadas experimentalmente. Tais regiões são levadas em consideração no processo de divisão da proteína em domínios, os quais são modelados comparativamente. As regiões não homólogas, e que, portanto, não compõem os domínios nem dispõem de *molde*, são modeladas *ab initio*, dando origem a fragmentos, unidos aos domínios em etapa final da construção do modelo teórico (KIM; CHIVIAN; BAKER, 2004).

O I-TASSER é um sistema automatizado para predição estrutural de proteínas. O servidor identifica moldes de estruturas depositadas no banco de dados PDB a partir da sequência de aminoácidos submetida, usando, para tal, o programa LOMETS, que consiste em um método integrado de algoritmos que realizam múltiplos alinhamentos de sequência. O I-TASSER emprega técnicas de modelagem comparativa para prever as regiões da proteína de interesse alinhadas continuamente aos *moldes* selecionados pelo servidor. As demais regiões, que não dispõem de estrutura molde no banco de dados PDB são construídas a partir de métodos *ab initio* de modelagem molecular. O algoritmo SPICKER é então empregado para identificar regiões modeladas de baixa energia livre, refinadas a seguir por simulações de dinâmica molecular nos algoritmos FG-MD e ModRefiner (YANG; ZHANG, 2015).

O servidor (PS)² prediz as estruturas tridimensionais de complexos proteicos com base na modelagem comparativa; Além disso, este servidor examina o acoplamento entre as subunidades da predição do complexo pela combinação de aspectos estruturais e evolutivos. As predições do complexo estrutural podem ser indicados e visualizados em gráficos 3D em Java e os perfis estruturais e evolutivos são apresentados e comparados cadeia-a-cadeia. Acredita-se que o servidor (PS)² seria uma ferramenta útil para quem está interessado, não só nas estruturas dos

complexos de proteína, mas também na ligação entre as subunidades dos complexos proteicos. (HUANG et al., 2015)

O M4T é um servidor de modelagem comparativa que realiza a combinação de vários moldes e otimização de alinhamentos alternativos por meio de uma abordagem de agrupamento (do inglês, clustering) interativo que leva em consideração as informações sobre a similaridade entre a sequência molde e o alvo, a similaridade de sequência entre os moldes, a integridade estrutural do domínio do molde, a sua qualidade experimental e a contribuição única de cada molde para o alvo (RYKUNOV et al., 2009). Esse servidor desenvolve alinhamentos e modelos precisos, minimizando os erros associados com as duas primeiras etapas da modelagem comparativa: o reconhecimento e alinhamento do molde. No primeiro passo, as estruturas das proteínas são pesquisadas, comparadas e analisadas, e um número de candidatos é selecionado para servirem como moldes. Em seguida, para reduzir os erros associados aos alinhamentos, o M4T utiliza o módulo MMM (do inglês, Multiple Method Mapping) que considera soluções de vários métodos de alinhamento e combina as partes melhor alinhadas em uma única solução, que é, em média, mais precisa do que qualquer entrada única de alinhamentos. Na etapa final um modelo comparativo da proteína é construído utilizando Modeller (FERNANDEZ-FUENTES et al., 2007).

O servidor Phyre2 realiza a predição de estrutura proteica através do reconhecimento da homologia/analogia entre proteínas. Essa ferramenta é amplamente utilizada pela comunidade biológica, com mais de 150 submissões por dia, os resultados são exibidos em uma interface simples e demora 30 minutos para predizer a estrutura de uma proteína de 250 resíduos de aminoácidos (KELLEY; STERNBERG, 2009).

O RaptorX é um servidor que promove a predição da estrutura proteica secundária e a modelagem da estrutura terciária baseada em um molde, sendo capaz de detectar sequências moldes mesmo remotamente relacionadas. Assim, é possível obter modelos estruturais de alta qualidade para muitas sequências de proteínas-alvo, quando apenas domínios de proteínas distantemente relacionadas têm estruturas experimentalmente resolvidas. Para terminar o processamento de

uma sequência de 200 resíduos de aminoácidos, o servidor leva aproximadamente 30 minutos (KÄLLBERG et al., 2014).

O SWISS-MODEL é um sistema automatizado para modelagem de estruturas tridimensionais de proteínas a partir da sequência de aminoácidos fornecida, empregando, para tal, técnicas de modelagem comparativa. O servidor procura moldes no banco de dados de *moldes* do SWISS-MODEL Template Library (SMTL) a partir do alinhamento múltiplo de sequências e da estratégia de busca *Hidden Markov Models* (BIASINI et al., 2014) – que constitui um método de aprendizado de máquina que avalia probabilisticamente a distribuição de sequências para escolher a mais adequada para a finalidade proposta (WEI, 2012). A estrutura selecionada por tal servidor é então utilizada como molde para modelagem comparativa (BIASINI et al., 2014).

4.3. Alinhamento Estrutural e Validação dos Modelos

Os modelos obtidos foram alinhados à estrutura experimental incompleta da CDCA1 (PDB ID 3BOB, 3BOC, 3BOE, 3BOH, 3BOJ, 3UK8) disponível no Protein Data Bank, utilizando, para tal, o servidor TM-Align. O TM-Align informa valores de RMSD e TM-score, que são medidas utilizadas para avaliar a similaridade estrutural das moléculas alinhadas no servidor (ZHANG; SKOLNICK, 2005). Define-se o RMSD (*Root Mean Square Deviation*) como as diferenças espaciais entre duas estruturas estáticas (KNAPP et al., 2011), não representando, portanto, os desvios locais de estrutura. Além de apresentar valores de RMSD, o TM-align ainda disponibiliza os valores de TM-score, calculados a partir de desvios de menores porções estruturais, proporcionando maior sensibilidade na avaliação de similaridade estrutural (ZHANG; SKOLNICK, 2005).

A utilidade biológica de modelos tridimensionais reside na confiabilidade da predição estrutural (ROY; KUCUKURAL; ZHANG, 2011). Nesse sentido, embora os valores de RMSD sejam proporcionais ao tamanho da estrutura alinhada, e, portanto, relativos (YANG; ZHANG, 2015), valores reduzidos de RMSD indicam maior similaridade estrutural entre o modelo obtido e o *molde* utilizado em sua modelagem, conferindo-o maior confiabilidade (ROY; KUCUKURAL; ZHANG, 2011). Os valores de TM-score, por sua vez, não são relativos, pois independem do

tamanho da estrutura analisada. O TM-score varia de zero a um, sendo os valores mais próximos de um, indicativos de maior similaridade estrutural. Estatisticamente, valores de TM-score menores que 0,17 indicam que o alinhamento estrutural ocorreu entre duas estruturas randomicamente relacionadas, ao passo que valores maiores que 0,5 indicam que o alinhamento ocorreu entre duas estruturas similares (YANG; ZHANG, 2015). Através dos escores inferidos pelo programa, o modelo mais semelhante à estrutura alvo foi selecionado.

4.4. Validação de estruturas

Todos os modelos obtidos tiveram sua qualidade verificada em diferentes âmbitos, utilizando quatro servidores de validação: ProSA (WIEDERSTEIN; SIPPL, 2007), PROCHECK (LASKOWSKI et al., 1993), QMEAN (BENKERT; KÜNZLI; SCHWEDE, 2009) e Verify3D (EISENBERG; LÜTHY; BOWIE, 1997).

O ProSa afere a qualidade do modelo à partir do cálculo dos potenciais de força e energia da estrutura submetida em seu servidor. A partir de tais cálculos, o servidor gera um valor referente à qualidade geral da estrutura, o Z-score. Este valor é disposto em um gráfico que contém os Z-scores de todas as estruturas experimentais depositadas no Protein Data Bank (WIEDERSTEIN; SIPPL, 2007).

O PROCHECK avalia a qualidade estereoquímica do modelo, analisando, para tal, a geometria geral e individual (resíduo por resíduo) da estrutura submetida baseada na disposição dos ângulos phi/psi. Os resultados obtidos são dispostos no gráfico de Ramachandran, no qual os resíduos são alocados em regiões favoráveis, permitidas e não permitidas, conforme seu arranjo espacial (LASKOWSKI et al., 1993).

O QMEAN estima a qualidade do modelo submetido em seu servidor com base em propriedades físico-químicas de sua estrutura. Tais propriedades são calculadas estatisticamente a partir de conformações estruturais conhecidas, por meio de um método consenso, no qual as conformações previstas com maior frequência para determinada sequência estão mais propensas a estarem corretas que os padrões estruturais ocorridos em apenas alguns modelos. A partir desses cálculos, o servidor gera um valor referente à qualidade geral da estrutura submetida, o QMEAN-score, que é disposto em um gráfico que contém o QMEAN-

score de 9766 estruturas experimentais de alta resolução (BENKERT; KÜNZLI; SCHWEDE, 2009).

O Verify3D determina a compatibilidade estrutura-sequência a partir do perfil tridimensional da proteína extraída de suas coordenadas atômicas. Este perfil é comparado à sequência de aminoácidos da proteína, e baseado nas combinações conhecidas de estruturas experimentais, é gerado um valor de compatibilidade, que é proporcional às combinações bem sucedidas e o indicativo da qualidade do modelo (EISENBERG; LÜTHY; BOWIE, 1997).

Selecionou-se o modelo que apresentou maior qualidade estrutural para a análise filogenética e dinâmica molecular.

4.5. Refinamento Estrutural

Foi realizada uma etapa de refinamento estrutural do modelo gerado por modelagem *ab initio* pelo servidor Robetta, a fim de otimizar a qualidade da estrutura, utilizando dois algoritmos de refinamento: GalaxyWeb (KO et al., 2012) e KoBaMIN (RODRIGUES; LEVITT; CHOPRA, 2012).

O servidor GalaxyWeb prediz a estrutura da proteína a partir da sequência de modelagem com base no modelo e refina regiões de loop ou regiões terminais por modelagem *ab initio*. Este servidor web é baseado no método testado no CASP9 (Nona Avaliação Crítica de Técnicas de Predição de Estrutura de Proteínas) sendo avaliado para estar entre os servidores de modelagem de melhor desempenho baseados em modelos. O método gera estruturas centrais de confiança de vários modelos e reconstrói loops ou terminais não confiáveis utilizando um método de refinamento baseado em otimização (KO et al., 2012).

O servidor web KoBaMIN fornece uma interface on-line simples, consistente e computacionalmente eficiente para um protocolo de refinamento da estrutura de proteínas baseado na minimização de um potencial de força média baseada no conhecimento. O servidor pode ser utilizado para refinar a estrutura de uma única proteína ou um conjunto de proteínas a partir de suas coordenadas não refinadas em formato PDB. O método de refinamento é particularmente rápido e preciso devido ao potencial do conhecimento subjacente derivado de estruturas depositadas no PDB. O servidor permite opções, mas são recomendados passos de

otimização estereoquímica usando o software MESHI. O servidor KoBaMIN também permite a comparação das estruturas refinadas com uma estrutura de referência fornecida para avaliar as mudanças trazidas pelo protocolo de refinamento (RODRIGUES; LEVITT; CHOPRA, 2012).

4.6. Análise filogenética estrutural

A análise filogenética da proteína CDCA1 foi realizada no servidor Consurf (ASHKENAZY et al., 2010) utilizando o modelo tridimensional validado como *input*. Para tal, selecionou-se os seguintes parâmetros: *Homolog search algorithm*: HMMER; *Number of iterations*: 1; *E-value cutoff*: 0.0001; *Proteins database*: UNIREF-90; *Maximum homologs to collect*: 150; *Maximal %ID between 11 sequences*: 95%; *Minimal %ID for homologs*: 35%; *Alignment method*: MAFFT-LINS-i; *Calculation method*: Bayesian; *Evolutionary substitution model*: JTT (default).

O Consurf é um servidor online desenvolvido para identificar regiões funcionalmente importantes em proteínas a partir da estimativa do grau de conservação dos aminoácidos de sua sequência. Tal estimativa é obtida a partir de uma estratégia integrada de algoritmos que realizam alinhamentos múltiplos de sequência e aplicam métodos Bayesianos de aprendizado de máquina para calcular a preservação evolutiva destes resíduos. O servidor atribui a cada resíduo um valor de conservação associado a uma escala colorimétrica, variando do azul, representando resíduos menos conservados, a bordô, representando resíduos mais conservados (ASHKENAZY et al., 2010).

4.7. Dinâmica Molecular

As simulações de dinâmica molecular da CDCA1 foram realizadas individualmente no pacote GROMACS 5.0.7 (VAN DER SPOEL et al., 2005). O GROMACS inclui mais de 100 programas, empregados não só na efetivação da dinâmica molecular, mas também na análise da trajetória gerada a partir desta, permitindo, assim, a caracterização de diversos parâmetros estruturais (VAN DER SPOEL et al., 2005).

Conforme metodologia estabelecida por KREBS; DE MESQUITA, 2016 – que utiliza o campo de força Amber 99SB-ILDN (LINDORFF-LARSEN et al., 2010)

para reger as interações interatômicas – solvatou-se inicialmente a estrutura com o modelo de água TIP3P em uma caixa cúbica com dimensões de 6,829nm x 7,257nm x 7,442nm. O sistema foi então neutralizado pela adição de íons Na⁺ e Cl⁻ em concentração de 1,5 M, e teve sua energia minimizada em 5000 etapas, utilizando, para tal, o método *steepest descent*.

Após a minimização, aplicou-se o conjunto NVT (*constant number, volume and temperature*) e NPT (*constant number, pressure and temperature*) para estabilizar a pressão e temperatura do sistema em torno de 300K e 1atm (MEATH et al., 1981). Esta etapa teve duração de 100 picosegundos para cada conjunto, resultando na equilibração do sistema.

O sistema foi então submetido à produção, etapa final da dinâmica molecular, utilizando-se o algoritmo LINCS (*Linear Constraint Solver*) para restringir as interações covalentes, e o método PME (*Particle Mesh Ewald*) para processar as interações eletrostáticas. A simulação foi realizada à 300K e teve duração de 40 nanossegundos, alcançando a estabilidade estrutural da proteína analisada.

As trajetórias obtidas a partir das simulações de dinâmica molecular foram analisadas pelos programas: gmxsasa, gmxbond, gmxrms, gmx gyrate e gmxrmsf, do pacote GROMACS. Tal análise viabilizou a caracterização de parâmetros estruturais e bioquímicos – como, desvio médio quadrático (*Root Mean Square Deviation* – RMSD), flutuação média quadrática (*Root Mean Square Fluctuation* – RMSF), fator termodinâmico (*B-factor*), raio de giro (*radius of gyration*), pontes de hidrogênio intermoleculares (Hydrogen bonds) e superfície de acesso ao solvente (*Solvent Accessible Surface Area* – SASA) que foram dispostos em gráficos e interpretados comparativamente.

5. Resultados e Discussão

5.1. Obtenção da sequência e estrutura da CDCA1

A sequência nativa da proteína CDCA1, em formato FASTA, contendo 616 resíduos de aminoácidos, foi obtida a partir do banco de dados UniProt (UniProt ID: Q50EL4) Os modelos experimentais escolhidos estão descritos na tabela 1.

Todos os modelos são incompletos, daí a importância de se realizar modelagem computacional para obter a estrutura completa da CDCA1. Os arquivos em formato PDB, contêm as coordenadas de cada átomo da estrutura.

Tabela 1 - Sequência e estruturas da CDCA1.

Sequência formato FASTA	Estrutura	Resolução (Å)	Posição	Método de Obtenção
>tr Q50EL4 Q50EL4_THAWE Cadmium-specific carbonic anhydrase (Fragment) OS=Thalassiosira weissflogii GN=cdca1 PE=1 SV=1 GWQAEIVTEFSLNEMVDVDPQGILKCV DGRGSDNTQ FCGPKMPGGIYIAHNRGVTTLEGLKQITKEVASKGH VPSVHGDHSSDMLGCGFFKLWVTGRFDDMGYP RPQFD ADQGAKAVENAGGVIEMHHGSHA EKVVYINLVENKTL EPDEDDQRFIVD GWAAGKFGLDV PKFLIAAAATVEML GGPKKAKIVIPSI SPAQIAEALQGRGWDAEIVTDASM AGQLVDVRPEGI LKCV DGRGSDNTRMG GPKMPGGIYA IAHNRGVTSIEGLKQITKEVASKGHLPSVHGDHSSDM LGC GFFKLWVTGRFDDMGYP RPQFDADQGANAVKDAG GIEMHHGSHTEKVVYINLLANKTLEPNENDQRFIVD GWAADKFGLDV PKFLIAAAATVEMLGGPKNAKIVVPS ITPPQIVSALRGRGWKASIVKASTMSSELKRVD PQGI LKCVDGRGSDNTQFGGPKMPGGIYIAHNRGVTTLEG LKDITREVASKGHVPSVHGDHSSDMLGCGFFKLWLTG RFDDMGYP RPEFDADQ GALAVRAAGGVIEMHHGSHEE KVVYINLVSGMTLEPNHDQRFIVD GWAASKFGLDV KFLVAAAATVEMLGGPKKAKIVIP	3BOB	1.45	197 - 406	Raio X
	3BOC	1.80	197 - 406	Raio X
	3BOE	1.40	197 - 406	Raio X
	3BOH	1.70	1 - 196	Raio X
	3BOJ	1.45	1 - 196	Raio X
	3UK8	2.30	407 - 616	Raio X

5.2. Modelagem molecular computacional

Na figura 1, observa-se o modelo teórico 3D da CDCA1 gerado pelos algoritmos de modelagem comparativa ((PS)², M4T, Phyre2, RaptorX, SwissModel). No modelo gerado pelo (PS)² (Figura 1 – A), com 616 resíduos de aminoácidos, uma região da proteína não foi modelada, o que pode ter sido falha desse algoritmo ou um indicativo de que esta região é uma região intrinsecamente desestruturada (IDR) (STANLEY; ESTEBAN-MARTÍN; DE FABRITIIS, 2015). O algoritmo M4T construiu dois modelos teóricos 3D da proteína, com 214 e 213 resíduos de aminoácidos, respectivamente. (Figura 1 – B e C). Os algoritmos Phyre2 e RaptorX criaram apenas um modelo, cada um com 616 aminoácidos (Figura 1 – D e E, respectivamente). O SwissModel gerou três modelos teóricos tridimensionais da proteína, com 210, 196 e 196 resíduos de aminoácidos, respectivamente (Figura 1 – F, G e H).

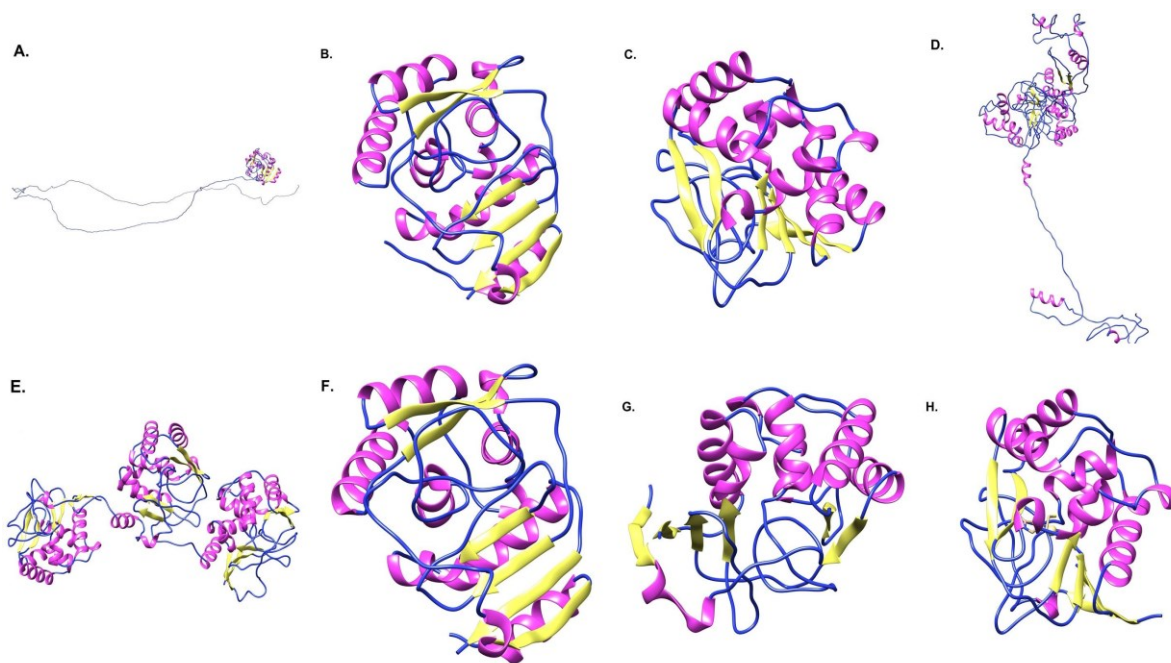


Figura 1 - Modelos moleculares da CDCA1 obtidos por modelagem comparativa.

Visualização dos modelos teóricos 3D construídos com base na sequência nativa da CDCA1 pelos algoritmos de modelagem comparativa. A) (PS)²; B) Modelo 1 M4T; C) Modelo 2 M4T; D) Phyre2; E) RaptorX; F) Modelo 1 SwissModel; G) Modelo 2 SwissModel e; H) Modelo 3 SwissModel. Os modelos foram visualizados e coloridos de acordo com sua estrutura secundária, através do programa UCSF Chimera. Os fragmentos em azul indicam os “loops”, as regiões em amarelo representam as folhas beta e as estruturas coloridas de rosa representam as alfa-hélices.

Os algoritmos de modelagem híbrida, Rosetta e I-Tasser, geraram cinco modelos teóricos tridimensionais da CDCA1. A figura 2 mostra, na respectiva ordem, os cinco modelos obtidos pelo Rosetta e pelo I-Tasser. A cada modelo do I-Tasser foram atribuídos, automaticamente pelo próprio algoritmo, um C-score, que estima a qualidade dos modelos e está tipicamente no intervalo [-5, 2], no qual um valor mais elevado reflete um modelo de melhor qualidade (ROY; KUCUKURAL; ZHANG, 2010; ROY; YANG; ZHANG, 2012; ZHANG, 2008). O melhor modelo gerado por esse algoritmo foi o Modelo 5, pois teve maior C-score (-1.87), quando comparado aos demais, que obtiveram C-Score de -3.09 , -3.52, -3.70 e -3.53, respectivamente. Apesar disso, todos os cinco modelos gerados pelo I-Tasser foram alinhados com as estruturas experimentais da CDCA1 na etapa de alinhamento estrutural e validação dos modelos.

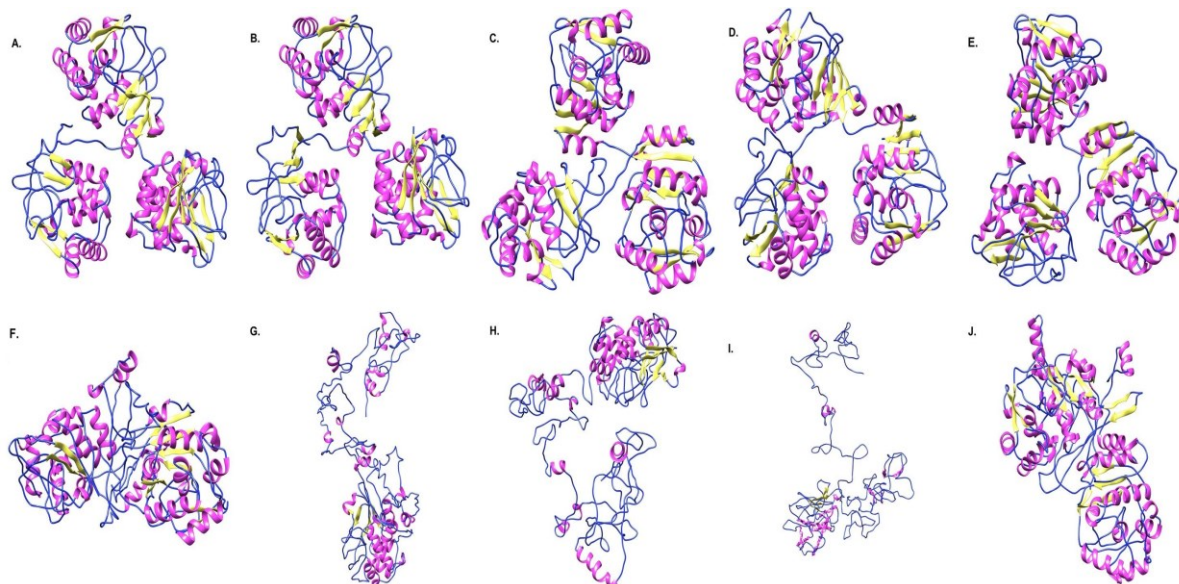


Figura 2 - Modelos moleculares da CDCA1 obtidos por modelagem *ab initio*.

Modelos teóricos 3D construídos por modelagem *ab initio* gerados pelos algoritmos Rosetta e I-Tasser. As letras “A, B, C, D e E” representam os cinco modelos gerados pelo Rosetta. As letras “F, G, H, I e J” indicam os cinco modelos gerados pelo I-Tasser. Os modelos foram visualizados e coloridos de acordo com sua estrutura secundária, através do programa UCSF Chimera. Os fragmentos em azul indicam os “loops”, as regiões em amarelo representam as folhas beta e as estruturas coloridas de rosa representam as alfa-hélices.

5.3. Alinhamento Estrutural e Validação dos Modelos

Avaliar a qualidade de modelos 3D, preditos computacionalmente, consiste em uma importante etapa de experimentos *in silico*, pois a confiabilidade atribuída aos modelos construídos determina a maneira com que estes poderão ser utilizados posteriormente (ROY; KUCUKURAL; ZHANG, 2010). Para avaliar essa qualidade dos modelos gerados, foi realizado o alinhamento entre cada estrutura proteica predita computacionalmente e as estruturas experimentais da CDCA1 (PDB ID 3BOB, 3BOC, 3BOE, 3BOH, 3BOJ, 3UK8).

Cada alinhamento gerou como resultado, uma imagem tridimensional das estruturas proteicas, experimental e teórica, sobrepostas, bem como um TM-Score e um valor de RMSD. O TM-score (do inglês, “template modeling score”) é definido para avaliar a similaridade topológica entre pares de estruturas. Seu valor varia entre 0 e 1, sendo que maiores valores indicam melhores alinhamentos estruturais e seu significado estatístico independe do tamanho da proteína. O valor de RMSD (do

inglês, “Root-Mean-Square Deviation”), por sua vez, indica a distância média, em angstroms, entre os átomos das cadeias principais de duas estruturas, podendo ser utilizado para avaliar a qualidade dos modelos (ROY; KUCUKURAL; ZHANG, 2010; ROY; YANG; ZHANG, 2012; ZHANG, 2008). O valor do RMSD deve ser menor que 2.0Å, sendo que valores mais próximos de 0.0Å representam maior similaridade topológica à estrutura experimental.

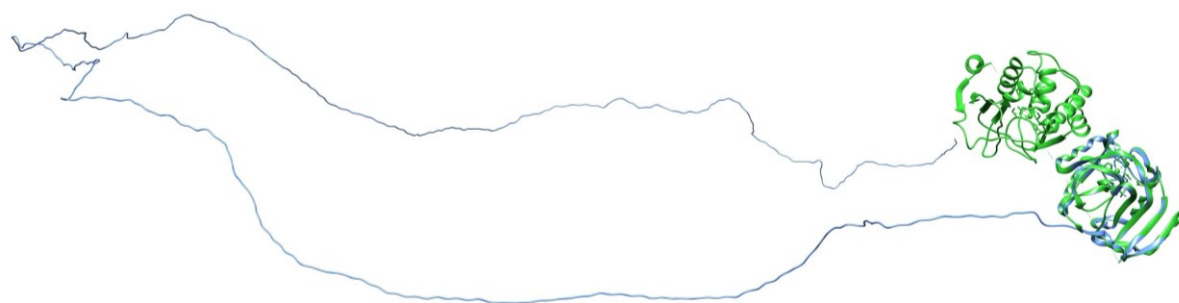
Conforme os resultados apresentados na tabela 2, a confiabilidade dos modelos construídos neste estudo, exceto o modelo 4 do I-Tasser, apresenta-se dentro dos padrões de alta qualidade, visto que os seus valores de RMSD encontraram-se abaixo de 2.0Å e, os de TM-score, foram iguais ou superiores a 0.92, indicando que os modelos criados possuem grande similaridade topológica à estrutura experimental da CDCA1.

Tabela 2 - Compilação dos valores de RMSD e TM-Score do(s) modelo(s) gerado(s) por cada algoritmo.

Algoritmo	Modelo	RMSD	TM-Score	Cadeia de aa
PS2	PS2 Modelo 1	0.16	0.99916	616
M4T	M4T Modelo 1	0.12	0.99954	214
	M4T Modelo 2	0.26	0.99774	210
Phyre2	Phyre Modelo 1	0.52	0.99124	616
RaptorX	RaptorX Modelo 1	0.46	0.99302	616
SwissModel	SwissModel Modelo 1	0.07	0.98116	210
	SwissModel Modelo 2	0.07	0.93766	196
	SwissModel Modelo 3	0.07	0.92003	196
Rosetta	Rosetta Modelo 1	0.39	0.99496	616
	Rosetta Modelo 2	0.42	0.99436	616
	Rosetta Modelo 3	0.42	0.99414	616
	Rosetta Modelo 4	0.44	0.99357	616
	Rosetta Modelo 5	0.44	0.99386	616
I-Tasser	I-Tasser Modelo 1	0.22	0.99834	616
	I-Tasser Modelo 2	0.24	0.99799	616
	I-Tasser Modelo 3	0.49	0.99345	616
	I-Tasser Modelo 4	6.32	0.31456	616
	I-Tasser Modelo 5	0.45	0.99605	616

Apesar de ter apresentado um baixo RMSD (0.16Å) e o TM-Score próximo de 1 (0.99916), o modelo gerado pelo (PS)² não foi escolhido, pois uma região da proteína não foi modelada, como pode ser observado na figura 3. Os modelos gerados pelo M4T (Figura 4) e pelo SwissModel (Figura 5), mesmo obtendo um baixo valor de RMSD, apresentaram uma modelagem incompleta, não

contemplando o total dos 616 resíduos de aminoácidos da proteína. Em comparação com os demais modelos completos, o modelo 1 gerado pelo algoritmo I-Tasser (Figura 6) foi o que apresentou o maior TM-Score (0.99834) e menor RMSD (0.22Å), seguido do modelo 1 do Rosetta com RMSD igual a 0.39Å e TM-Score (0.99496) (Figura 7), do modelo do RaptorX que apresentou um RMSD de 0.46Å e TM-Score (0.99302) (Figura 8) e por último, do modelo gerado pelo algoritmo Phyre2 o qual exibiu um valor de 0.52Å de RMSD e 0.99124 (Figura 9). Contudo, como as diferenças dos valores dos RMSD e TM-Scores não foram significativas, todos os modelos com os melhores valores e que contemplaram o total de resíduos da proteína foram analisados na validação estrutural.



RMSD: 0.16
TM-Score: 0.99916

Figura 3 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor (PS)².

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e o modelo 3D teórico (azul), construídos com base na sequência nativa da CDCA1. Modelo construído pelo (PS)² com RMSD (0.16) e TM-Score (0.99916). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

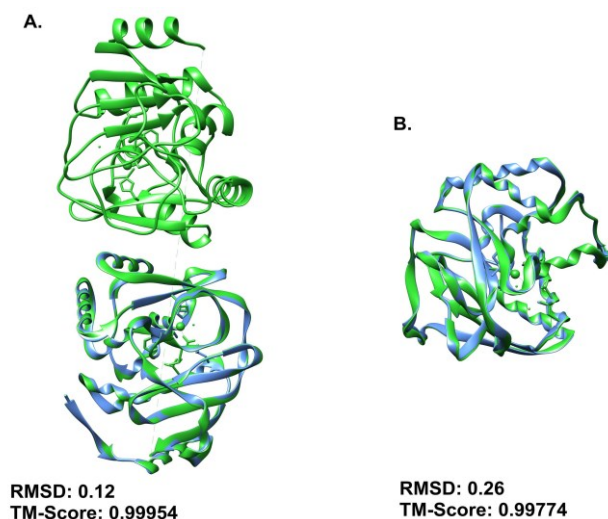


Figura 4 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor M4T.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D (azul), construídos com base na sequência nativa da CDCA1. A) Modelo 1 construído pelo M4T com RMSD (0.12) e TM-Score (0.99954) com um total de 214 resíduos de aminoácidos; B) Modelo 2 construído pelo M4T com RMSD (0.26) e TM-Score (0.99774) com um total de 213 resíduos de aminoácidos. Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

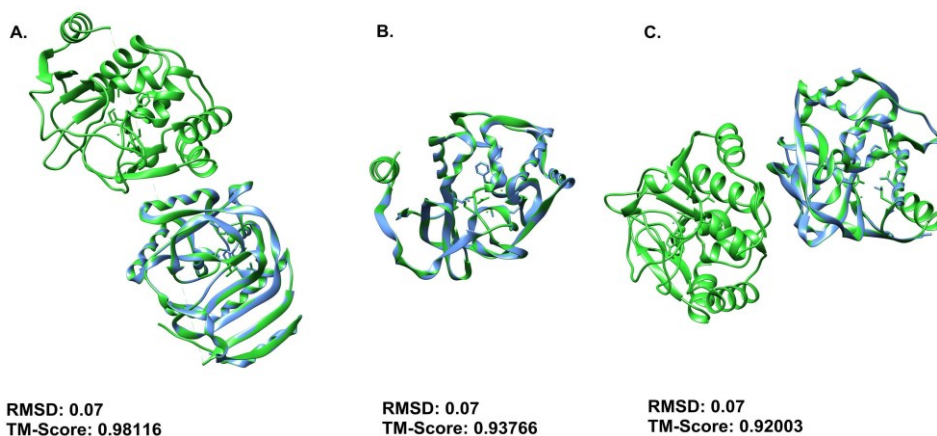


Figura 5 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor SwissModel.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D (azul), construídos com base na sequência nativa da CDCA1. A) Modelo 1 construído pelo SwissModel com RMSD (0.07) e TM-Score (0.98116) com um total de 210 resíduos de aminoácidos; B) Modelo 2 construído pelo SwissModel com RMSD (0.07) e TM-Score (0.93766) com um total de 196 resíduos de aminoácidos; C) Modelo 3 construído pelo SwissModel com RMSD (0.07) e TM-Score (0.92003) com um total de 196 resíduos de aminoácidos. Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

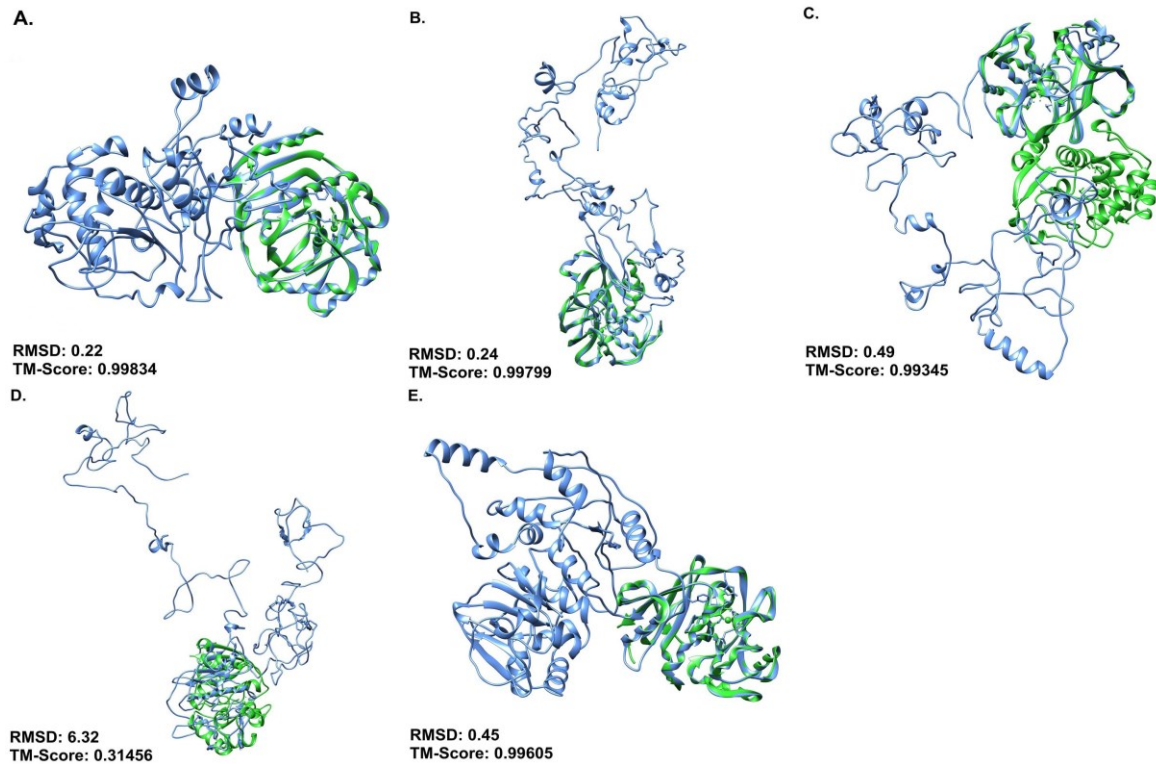


Figura 6 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor I-Tasser.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D (azul), construídos com base na sequência nativa da CDCA1. A) Modelo 1 construído pelo I-Tasser com RMSD (0.22) e TM-Score (0.99834); B) Modelo 2 construído pelo I-Tasser com RMSD (0.24) e TM-Score (0.99799); C) Modelo 3 construído pelo I-Tasser com RMSD (0.49) e TM-Score (0.99345); D) Modelo 4 construído pelo I-Tasser com RMSD (6.32) e TM-Score (0.31456); E) Modelo 5 construído pelo I-Tasser com RMSD (0.45) e TM-Score (0.99605). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

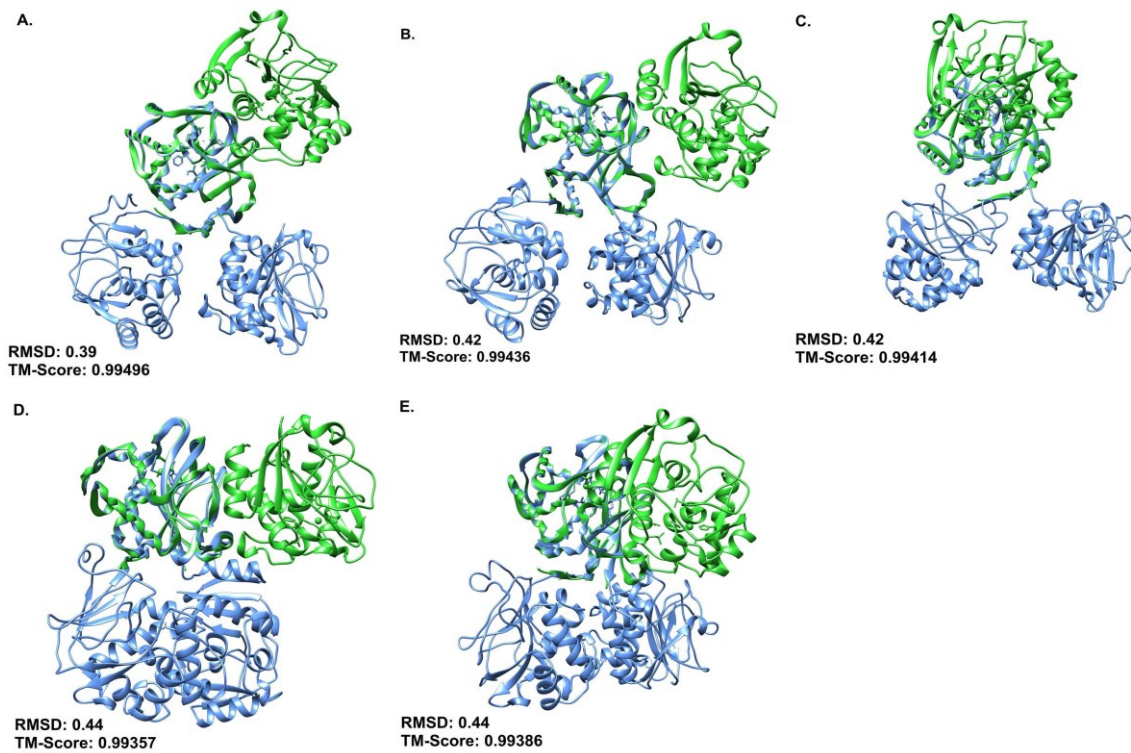


Figura 7 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos a partir do servidor Rosetta.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D (azul), construídos com base na sequência nativa da CDCA1. A) Modelo 1 construído pelo Rosetta com RMSD (0.39) e TM-Score (0.99496); B) Modelo 2 construído pelo Rosetta com RMSD (0.42) e TM-Score (0.99436); C) Modelo 3 construído pelo Rosetta com RMSD (0.42) e TM-Score (0.99414); D) Modelo 4 construído pelo Rosetta com RMSD (0.44) e TM-Score (0.99357); E) Modelo 5 construído pelo Rosetta com RMSD (0.44) e TM-Score (0.99386). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

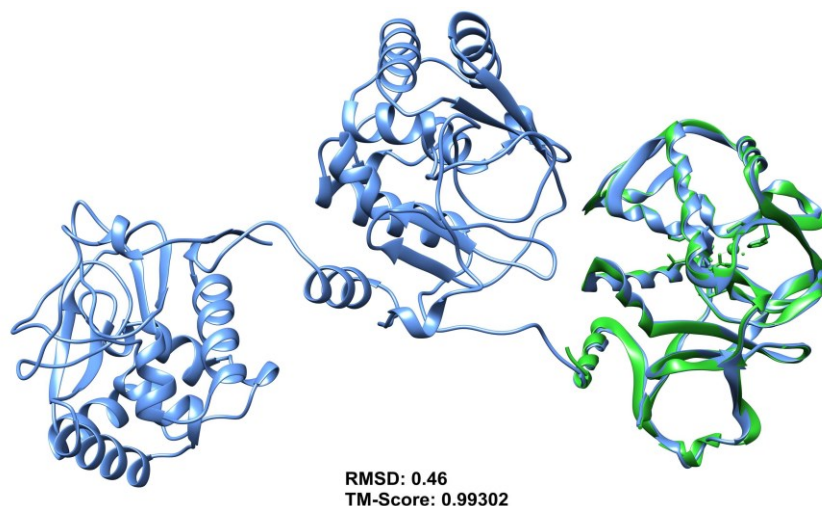


Figura 8 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor RaptorX.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e o modelo 3D teórico (azul), construídos com base na sequência nativa da CDCA1. Modelo gerado pelo RaptorX com RMSD (0.46) e TM-Score (0.99302). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

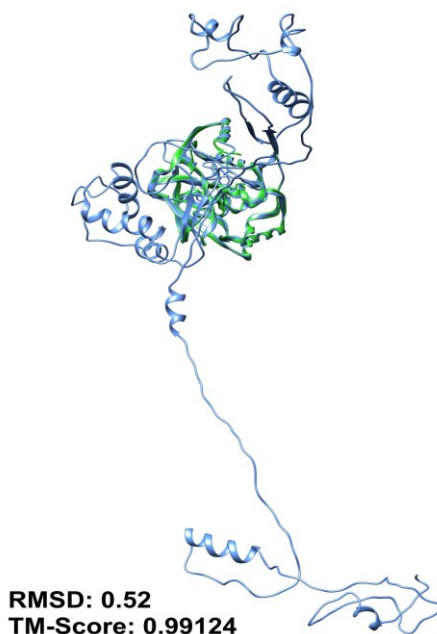


Figura 9 - Alinhamento entre a estrutura experimental da CDCA1 e o modelo obtido a partir do servidor Phyre2.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e o modelo 3D teórico (azul), construídos com base na sequência nativa da CDCA1. Modelo construído pelo Phyre2 com RMSD (0.52) e TM-Score (0.99124). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

5.4. Validação e Refinamento de Estruturas

A validação das estruturas proteicas obtidas tanto por metodologias experimentais quanto as computacionais é uma etapa fundamental para garantir a qualidade do dado obtido (TOUW; JOOSTEN; VRIEND, 2016).

Todos os modelos, com os melhores valores de RMSD e TM-Score obtidos pelo alinhamento estrutural, tiveram sua qualidade verificada em diferentes âmbitos, utilizando quatro servidores de validação conforme mostra a tabela 3.

Tabela 3 - Comparação dos valores dos servidores de validação estrutural dos modelos gerados por cada algoritmo.

Modelos		ProSA	Verify 3D (%)	Procheck (%)	QMEAN score	Z-Score QMEAN	RMSD	TM-Score
Phyre2	Modelo 1	-6,36	56.01	67,3	0,344	-4,63	0.52	0.99124
RaptorX	Modelo 1	-11,94	95.62	94,3	0,724	-0,43	0.46	0.99302
Rosetta	Modelo 1	-12,66	96.27	87.8	0,782	0,21	0.39	0.99496
I-Tasser	Modelo 1	-8,17	84.58	63,0	0,457	-3,39	0.22	0.99834

A seleção da estrutura foi feita com base nos valores produzidos pelos validadores. Tanto o modelo do Rosetta quanto o modelo obtido pelo RaptorX possuíram uma boa qualidade aceita pelos programas de validação. Devido o servidor Robetta ser amplamente utilizado, sendo melhor avaliado em relação ao servidor RaptorX, o modelo do Rosetta, passou por uma etapa de otimização, que busca o refinamento da estrutura, fornecendo assim um dado melhorado. O modelo do Rosetta foi submetido a dois servidores de refinamento e otimização estrutural. O servidor GalaxyWeb gerou como resultado cinco modelos estruturais teóricos 3D construídos pelo servidor Robetta que são mostrados na figura 10. O servidor KoBaMIN produziu apenas um modelo (Figura 11). Os resultados deste refinamento foram então alinhados (Figura 12) e validados, para serem comparados com os modelos do Rosetta e do RaptorX selecionados pelos validadores antes do refinamento.

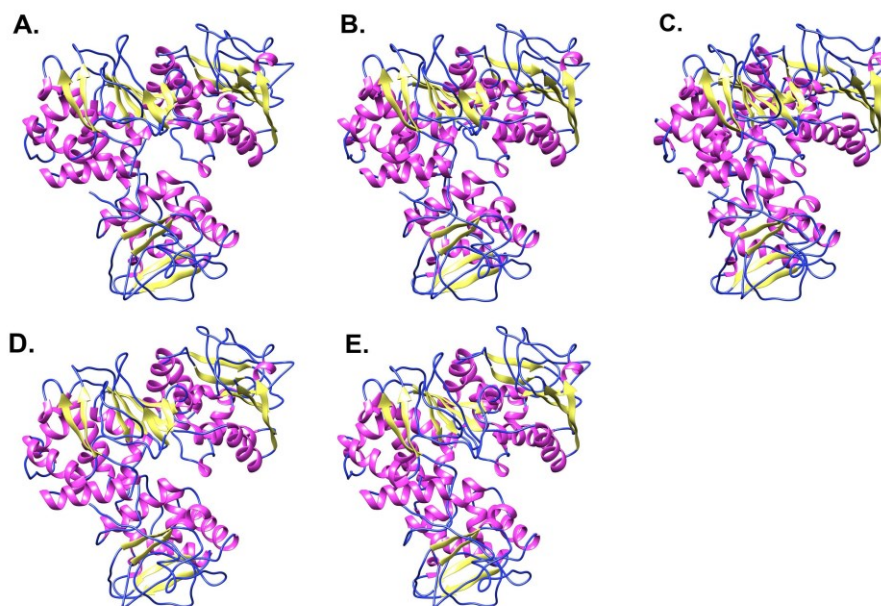


Figura 10 - Modelos moleculares da CDCA1 refinados pelo servidor GalaxyWeb, construídos a partir do modelo 1 do Rosetta.

Modelos teóricos 3D refinados pelo servidor GalaxyWeb construídos a partir do modelo 1 do Rosetta. As letras “A, B, C, D e E” representam os cinco modelos gerados. Os modelos foram visualizados e coloridos de acordo com sua estrutura secundária, através do programa UCSF Chimera. Os fragmentos em azul indicam os “loops”, as regiões em amarelo representam as folhas beta e as estruturas coloridas de rosa representam as alfa-hélices.

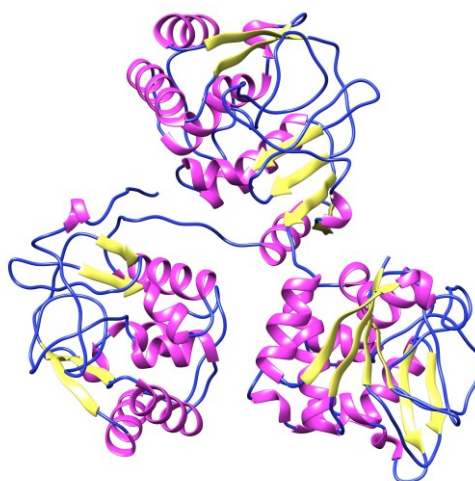


Figura 11 - Modelo molecular da CDCA1 refinado pelo servidor KoBaMIN, construído a partir do modelo 1 do Rosetta.

Modelo refinado pelo servidor KoBaMIN construído a partir do modelo 1 do Rosetta. Os modelos foram visualizados e coloridos de acordo com sua estrutura secundária, através do programa UCSF Chimera. Os fragmentos em azul indicam os “loops”, as regiões em amarelo representam as folhas beta e as estruturas coloridas de rosa representam as alfa-hélices.

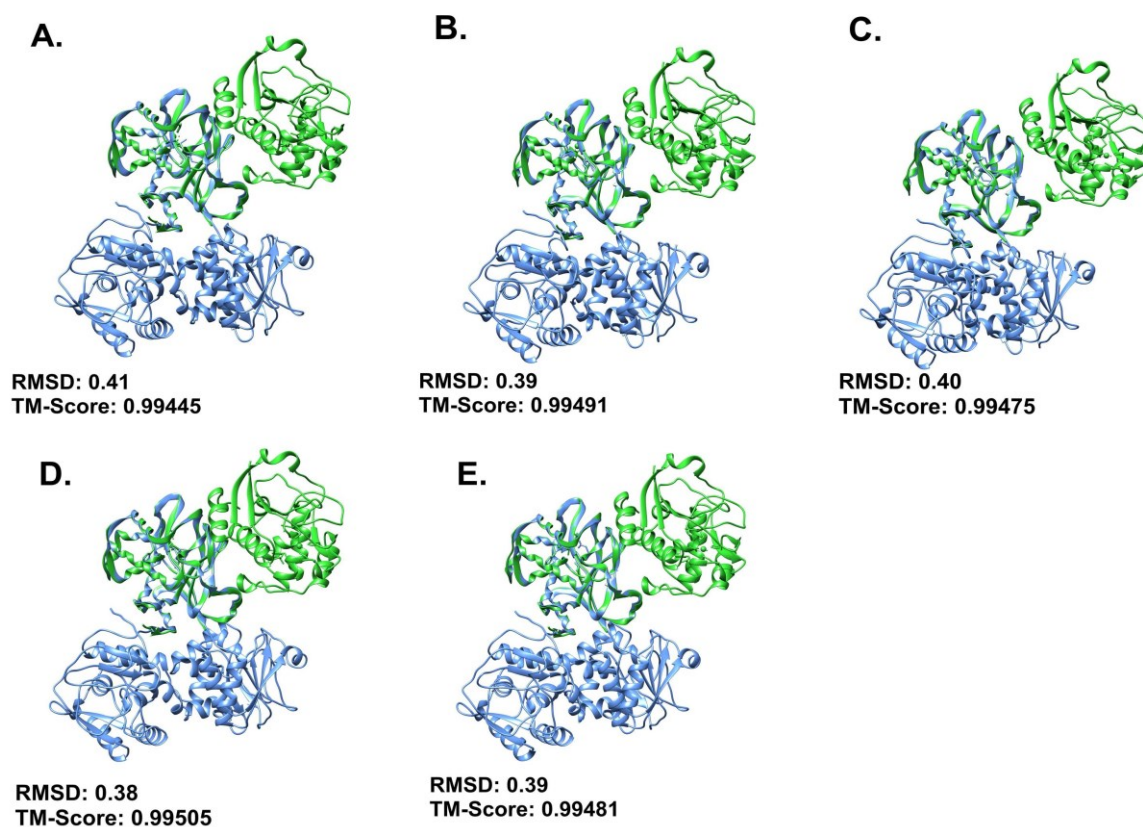
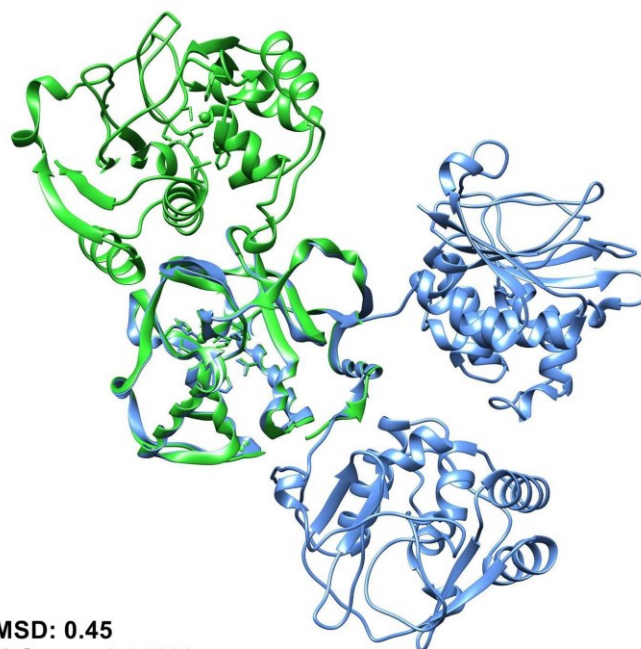


Figura 12 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos pelo refinamento no GalaxyWeb do modelo do Rosetta.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D refinados pelo GalaxyWeb (azul). A) Modelo 1 refinado do GalaxyWeb com RMSD (0.41) e TM-Score (0.99445); B) Modelo 2 refinado do GalaxyWeb com RMSD (0.39) e TM-Score (0.99491); C) Modelo 3 refinado do GalaxyWeb com RMSD (0.40) e TM-Score (0.99475); D) Modelo 4 refinado do GalaxyWeb com RMSD (0.38) e TM-Score (0.99505); E) Modelo 5 refinado do GalaxyWeb com RMSD (0.39) e TM-Score (0.99481). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.



RMSD: 0.45
TM-Score: 0.99328

Figura 13 - Alinhamento entre a estrutura experimental da CDCA1 e os modelos obtidos pelo refinamento no KoBaMIN do modelo do Rosetta.

Visualização do alinhamento entre a estrutura experimental da CDCA1 (verde) e os modelos teóricos 3D refinados pelo KoBaMIN (azul). Modelo refinado do KoBaMIN com RMSD (0.45) e TM-Score (0.99328). Os modelos foram visualizados e coloridos com auxílio do programa UCSF Chimera.

A tabela 4 mostra a comparação entre os valores obtidos pelos servidores de validação estrutural antes e após o refinamento.

Tabela 4 - Comparação dos valores dos servidores de validação estrutural dos modelos gerados após o refinamento estrutural.

Modelos		ProSA	Verify 3D (%)	Procheck (%)	QMEAN score	Z-Score QMEAN	RMSD	TM-Score
Antes Refinamento								
RaptorX	Modelo 1	-11,94	95.62	94,3	0,724	-0,43	0.46	0.99302
Rosetta	Modelo 1	-12,66	96.27	87.8	0,782	0,21	0.39	0.99496
Após Refinamento								
GalaxyWeb (Rosetta)	Modelo 1	-12,76	92,86	95,1	0,752	-0,13	0.41	0.99445
	Modelo 2	-12,88	93,67	94,3	0,763	-0,01	0.39	0.99491
	Modelo 3	-12,84	93,34	94,3	0,762	-0,01	0.40	0.99475
	Modelo 4	-12,74	93,67	94,3	0,755	-0,09	0.38	0.99505
	Modelo 5	-12,94	93,83	94,5	0,746	-0,19	0.39	0.99481
KoBaMIN (Rosetta)	Modelo 1	-12,78	94.97	93,9	0,741	-0,24	0.45	0.99328

Conforme os resultados mostrados na tabela 4, após a otimização da estrutura produzida pelo Rosetta, todos os modelos refinados foram aprovados pelos programas descritos na tabela. Devido o resultado do modelo 4, refinado pelo servidor GalaxyWeb e construído a partir do modelo 1 do algoritmo Rosetta (aqui chamado de Rosetta/GalaxyWeb4), apresentar um conjunto com os melhores valores gerados tanto pelo RMSD e TM-Score quanto pelos valores obtidos pelos programas de validação estrutural, este modelo foi o escolhido para as análises.

A estrutura da CDCA1 modelada pelo algoritmo Rosetta e refinada utilizando o servidor GalaxyWeb (Rosetta/GalaxyWeb4) apresentou os seguintes resultados para os programas de validação estrutural: de acordo com o ProSA-Web, a estrutura apresentada tem qualidade de cristalografia de raios-X, apontando um Z-score -12,74 (Figura 14); para PROCHECK, o modelo tem qualidade estereoquímica, uma vez que a geometria de ângulos Phi/Psi no gráfico de Ramachandran apresentou 94,3% de seus resíduos em regiões mais favorecidas, superando os 90% de corte estipulado para a plataforma (Figura 15); o servidor QMEAN atribuiu uma pontuação de 0,755 na qualidade global do modelo, qualificando-o como uma estrutura de alta resolução (Figura 16); finalmente o resultado do Verify 3D indica que 93,67% dos resíduos da estrutura tem compatibilidade 3D-1D, ao longo da passagem de 80% de corte estipulado (Figura 17).

Overall model quality

Z-Score: **-12.74**

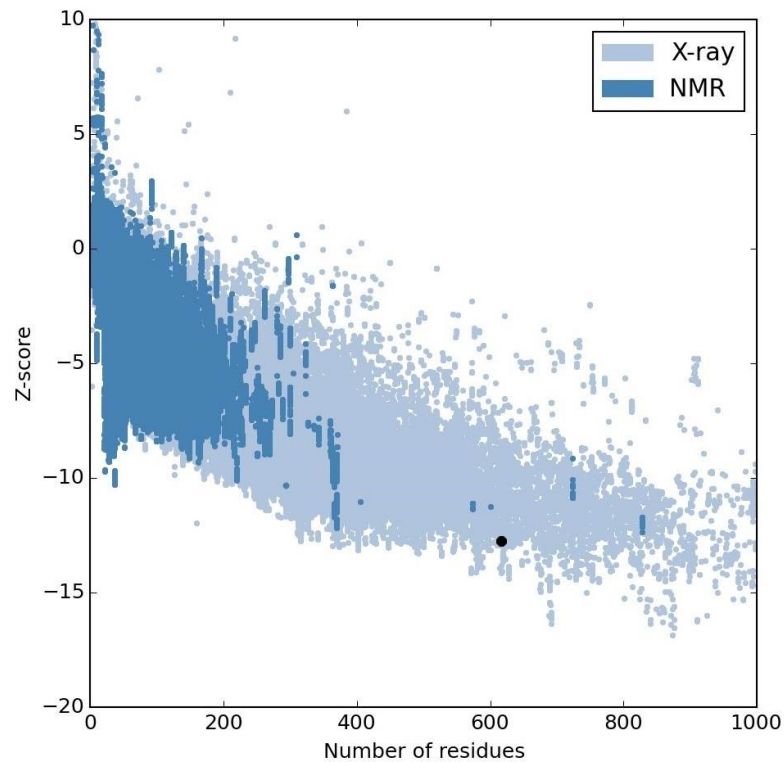
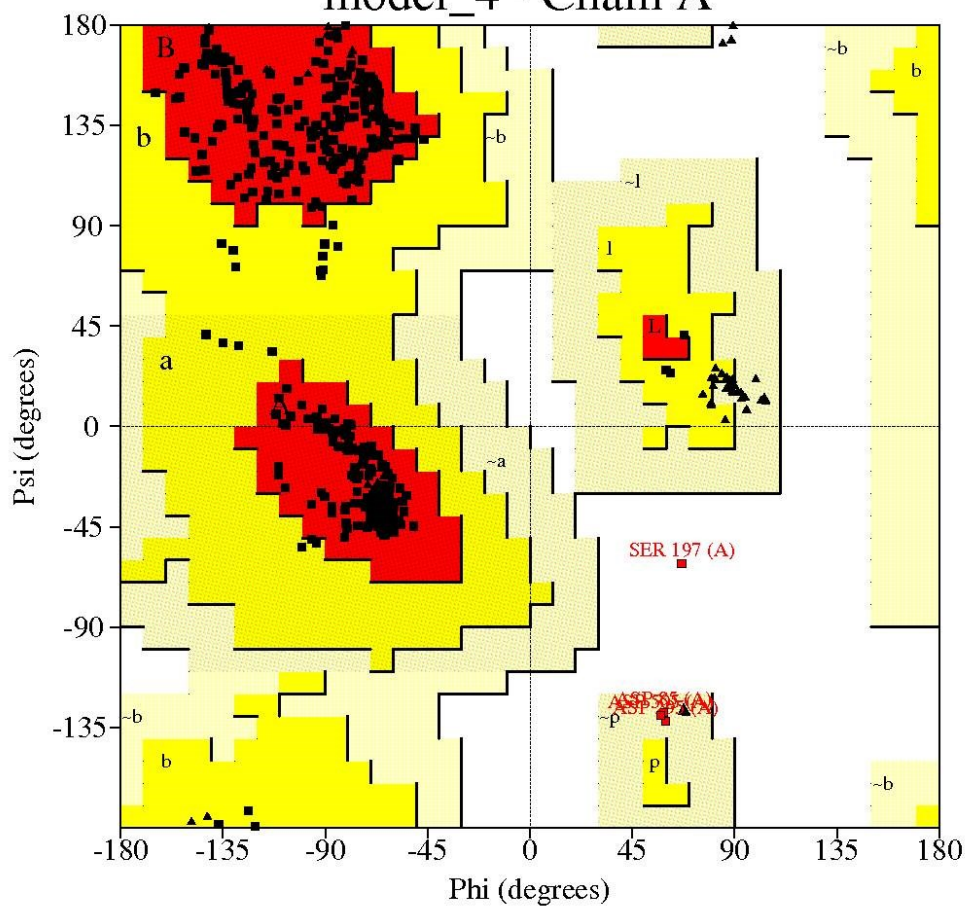


Figura 14 - Validação da estrutura da CDCA1 pelo servidor ProSA-Web.

O servidor Prosa-web gera Z-scores de estruturas PDB determinadas por cristalografia de raios-X em azul claro e por espectroscopia de RMN em azul escuro. De acordo com o gráfico, a estrutura da CDCA1 apresentada tem qualidade de cristalografia de raios-X e um Z-score -12,74 que está representado pelo ponto preto.

Ramachandran Plot model_4 - Chain A



Plot statistics

Residues in most favoured regions [A,B,L]	479	94.3%
Residues in additional allowed regions [a,b,l,p]	25	4.9%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.6%
Residues in disallowed regions	1	0.2%
<hr/>		
Number of non-glycine and non-proline residues	508	100.0%
Number of end-residues (excl. Gly and Pro)	0	
Number of glycine residues (shown as triangles)	76	
Number of proline residues	32	
<hr/>		
Total number of residues	616	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figura 15 - Validação da estrutura da CDCA1 pelo PROCHECK no gráfico de Ramachandran.

O Procheck mostrou que: 94,3% de resíduos de aminoácidos encontram-se nas regiões mais favorecidas [A, B e L]; 4,9% em regiões permitidas adicionais [a, b, l e p]; 0,6% nas regiões generosamente permitidas [~ a, ~ b, ~ l e ~ p] e; 0,2% nas regiões não permitidas.

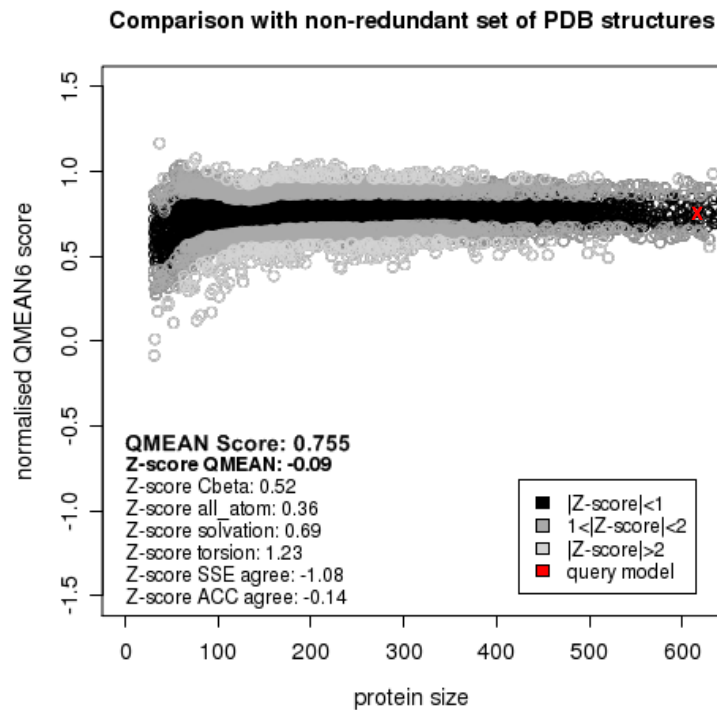


Figura 16 - Validação da estrutura da CDCA1 pelo QMEAN.

O QMEAN-score gera um gráfico de estruturas PDB determinada por cristalografia de raios-X em escala de cinza. O QMEAN-score do modelo da CDCA1 apresentou o valor de 0.755 e está representado pela cruz vermelha no gráfico.

Verify 3D Results plot

93,67% of the residues had an averaged 3D-1D score \geq 0,2

Pass

At least 80% of the amino acids have scored \geq 0,2 in the 3D/1D profile.

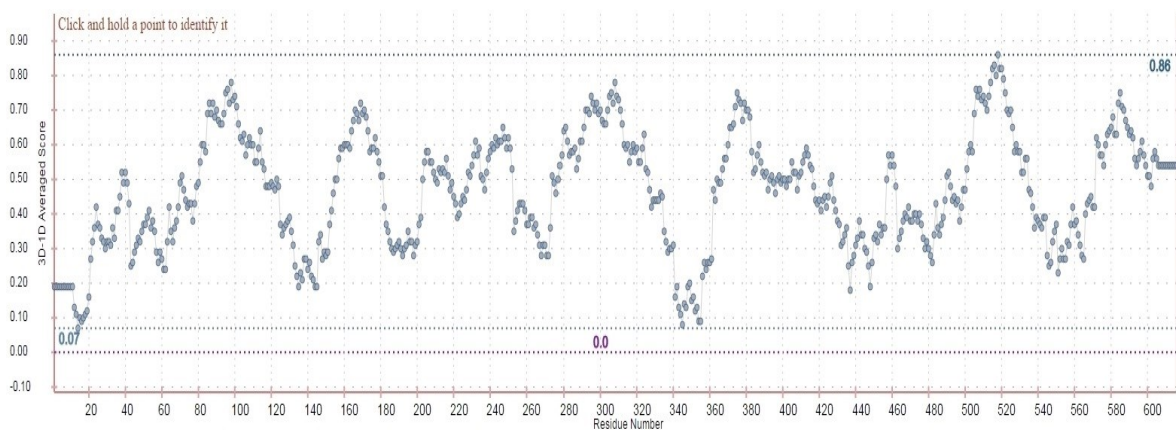


Figura 17 - Validação da estrutura da CDCA1 pelo Verify 3D.

Verify 3D mostrou que 93,67% dos resíduos da estrutura tem pontuação 3D-1D igual ou superior a 0.2.

5.5. Análise filogenética estrutural

O modelo selecionado foi submetido à análise filogenética no servidor Consurf. Este servidor permite identificar regiões de importância funcional em proteínas por meio da estimativa do grau de conservação dos aminoácidos de sua sequência (ASHKENAZY et al., 2010). Isso porque, resíduos cruciais para estrutura e função da proteína tendem a ser preservados pela seleção natural. Nesse contexto, a comparação entre sequências se faz essencial para inferir aspectos funcionais e evolutivos de novas proteínas a partir de proteínas depositadas em bancos de dados (XIONG, 2006). Para tal, o Consurf utiliza métodos de aprendizado de máquina e alinhamento múltiplos de sequências (ASHKENAZY et al., 2010).

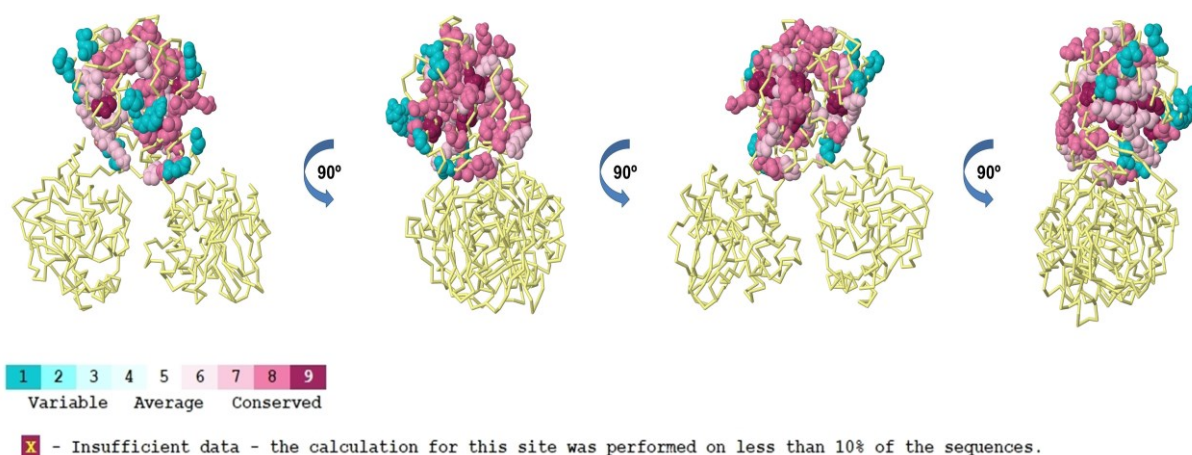


Figura 18 – Análise filogenética da proteína CDCA1 mediante submissão do modelo 1 do Rosetta no servidor Consurf.

Escala colorimétrica de conservação do Consurf aplicada à estrutura da proteína CDCA1. Tal escala varia de azul turquesa, para resíduos menos conservados, à bordô, para resíduos mais conservados. As partes amarelas mostra a insuficiência de dados.

A figura 18 mostra o resultado da análise filogenética da CDCA1. A conservação dos resíduos é relacionada à escala colorimétrica, na qual resíduos mais conservados são coloridos em bordô, enquanto resíduos menos conservados possuem uma coloração de azul turquesa. Resíduos que não apresentaram nenhuma conservação, devido à insuficiência de dados comparativos, não são analisados.

Por se tratar de uma proteína específica obtida evolutivamente para a situação de deficiência de zinco, não havendo dados comparativos, a análise filogenética estrutural da CDCA1 mostrou que a conservação talvez não seja tão indicativa, sendo normal ter uma baixa conservação em relação a todos os resíduos da proteína. Quando porções da proteína são analisadas separadamente, a região central da proteína apresenta uma alta conservação estrutural, enquanto na cadeia lateral, os resíduos não foram analisados, devido à insuficiência de dados comparativos, causando assim uma baixa conservação nessa região.

5.6. Dinâmica Molecular

O modelo validado da CDCA1 foi submetido à dinâmica molecular no pacote GROMACS 5.0.7 (VAN DER SPOEL et al., 2005). A etapa de produção durou 40 nanossegundos, e as trajetórias geradas a partir desta foram analisadas pelos parâmetros RMSD, RMSF, raio de giro, número de pontes de hidrogênio intermoleculares, SASA e B-Factor.

5.6.1. RMSD

O RMSD é definido como a diferença espacial entre duas estruturas estáticas. Tal parâmetro é calculado entre um ponto de partida definido da simulação e todos os quadros sucessivos desta, determinando a convergência e estabilidade da estrutura. Nesse sentido, os valores de RMSD são inversamente proporcionais à rigidez protéica. Tal parâmetro pode ainda ser utilizado para determinar o equilíbrio da simulação, uma vez que, a convergência do sistema – aferida mediante estabelecimento de um platô de valores de RMSD – indica a estabilização de seus elementos constituintes (KNAPP et al., 2011).

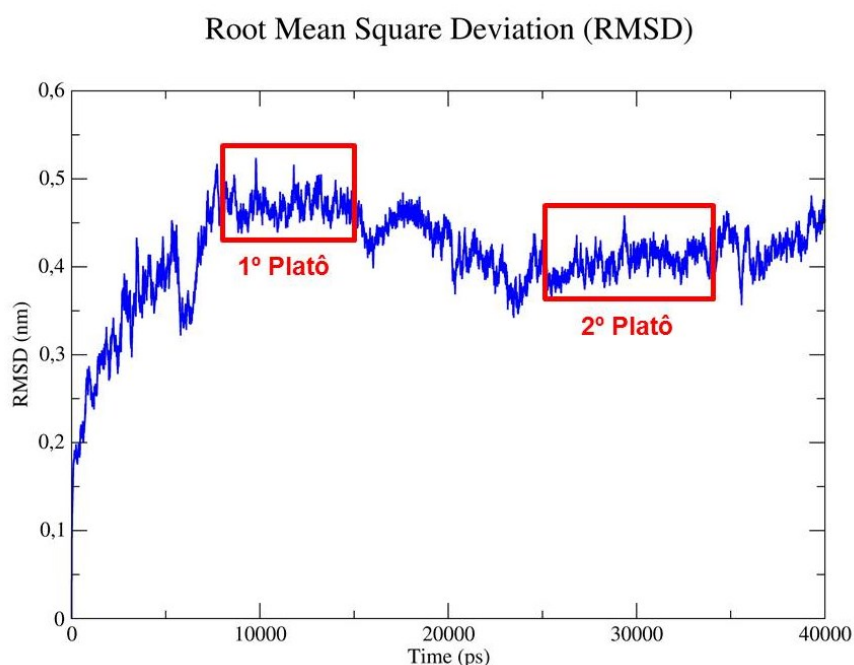


Figura 19 – RMSD da proteína CDCA1 ao longo da simulação de dinâmica molecular.

O valor de RMSD, em nanômetros, da estrutura nativa a 300K é mostrado em função do tempo, em picosegundos.

O equilíbrio foi alcançado na simulação realizada, uma vez que o RMSD da estrutura – disposto na Figura 19 – convergiu para um valor específico, estabelecendo dois platôs a 9-15ns e a 25-35ns do ponto de partida da dinâmica. A análise de RMSD pós-equilibração, apresentou a média dos valores de RMSD de cada platô, que correspondem, respectivamente, a 0,47nm e 0,41nm.

5.6.2. RMSF

O RMSF é uma medida empregada para mensurar o deslocamento de grupo de átomos em uma estrutura referenciada durante uma simulação. A utilização deste parâmetro só faz sentido se aplicado ao estudo de estruturas que flutuam em torno de uma conformação média estável, o que ocorre em simulações equilibradas. Dessa maneira, os valores de RMSF são proporcionais à movimentação de determinado resíduo em relação a sua posição média durante a dinâmica molecular, sendo utilizados para caracterizar localmente a estrutura quanto a sua flexibilidade (MARTINEZ, 2015). O resultado de RMSF é apresentado na Figura 20.

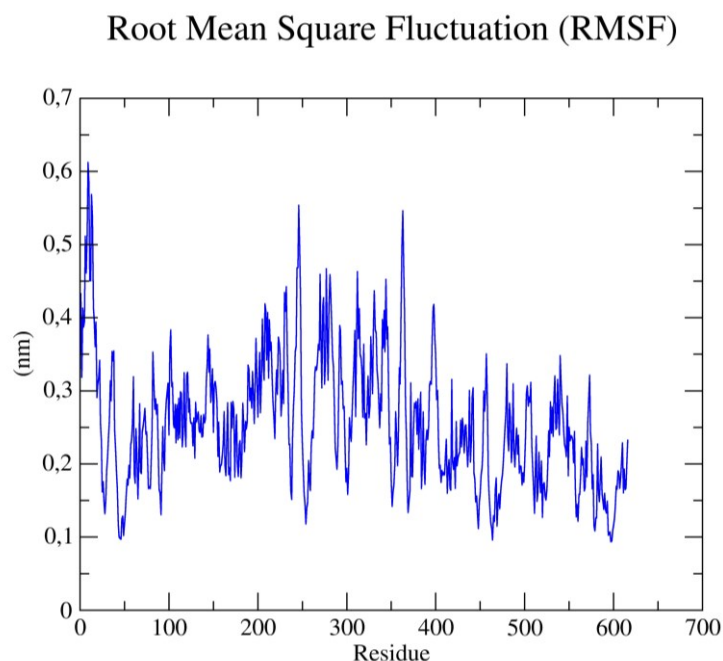


Figura 20 – RMSF da proteína CDCA1 ao longo da simulação de dinâmica molecular.
 O valor de RMSF, em nanômetros, da estrutura nativa a 300K por resíduo de aminoácido.

A proteína CDCA1 apresentou aumento comparativo de flexibilidade na região do domínio n-terminal – na posição 01 a 15 – bem como entre os resíduos 246 e 363 da proteína, incluindo o domínio de ligação ao cádmio nos resíduos Cys263, His315, e Cys325 na CDCA1 (TREVISOL, E. T. V.; DE MESQUITA, 2007). Com isso, a análise de RMSF sugere que a região N-terminal e a região central da proteína são mais flexíveis, em comparação a porção final da estrutura que se apresenta de forma mais rígida.

5.6.3. Fator B

A flexibilidade estrutural também pode ser analisada através do fator de Debye-Waller, também conhecido como fator B ou fator térmico. Este parâmetro reflete a mobilidade do átomo em função das vibrações térmicas e mensura a desordem estática de resíduos de aminoácidos e nucleotídeos, sendo utilizado na maioria dos métodos de predição de flexibilidade estrutural (CRAVEUR et al., 2015). O resultado da análise de fator-B foi apresentado na Figura 21.

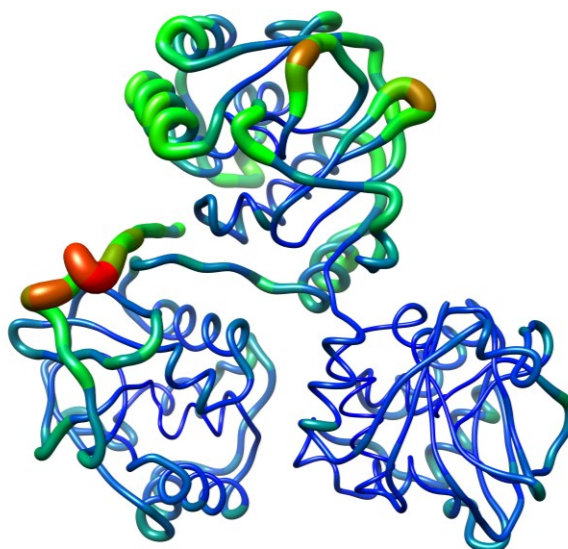


Figura 21 – Representação visual do fator térmico da estrutura nativa da CDCA1.

O fator térmico da CDCA1 nativa está representado conforme escala colorimétrica e de espessura. Resíduos delgados coloridos em azul possuem menor B-factor que resíduos espessos coloridos em vermelho.

A Figura 21 mostra as variações no fator B em diferentes regiões da proteína, de modo que regiões mais finas e azuis têm menor fator B e regiões mais grossas e vermelhas, maior. Era esperado que esse valor fosse elevado na região média da estrutura, local no qual existe uma função de mobilidade correlacionada com o seu papel enzimático. O B-factor pode ser correlacionado com os valores do RMSF.

5.6.4. Raio de Giro

O Raio de Giro é definido como a distância quadrática média ponderada em massa de um grupo de átomos a partir de seu centro de massa comum, o que dá uma medida acurada do volume e dimensões da proteína. Conseqüentemente, esta análise fornece as dimensões globais da estrutura, mensurando, portanto, sua compacidade (KUMAR et al., 2014). Além disso, proteínas enoveladas em configurações estáveis, provavelmente irão apresentar valores relativamente constantes de RG ao longo da simulação (KHAN et al., 2016). O resultado da análise do raio de giro foi apresentado na Figura 22.

Radius of gyration

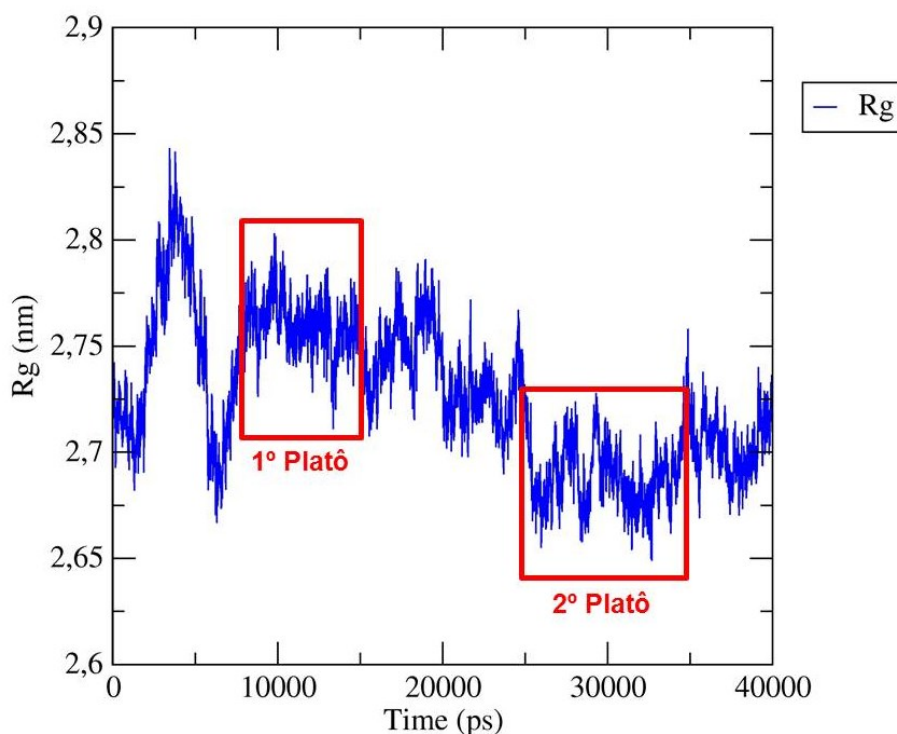


Figura 22 – Raio de giro ao longo da simulação de dinâmica molecular.

O Raio de Giração, em nanômetros, da estrutura nativa a 300K é mostrado como função do tempo, em picosegundos.

A proteína CDCA1 teve grande variação entre o primeiro e o segundo platô, mas entrou em estabilidade em cada período desses. O valor médio de RG pós-equilibração no primeiro platô foi de aproximadamente 2,76 nanômetros, enquanto no segundo platô a média foi de 2.69 nanômetros.

5.6.5. Pontes de hidrogênio intermoleculares

A estabilidade da proteína também é determinada pelo seu número de interações, tais quais interações hidrofóbicas e eletrostáticas, e ligações de hidrogênio (PIKKEMAAT et al., 2002). Nesse sentido, as ligações intermoleculares de hidrogênio, formadas durante a simulação de dinâmica molecular, possuem importante papel na estabilização da proteína, além de auxiliar a caracterização do perfil de interação proteína-ligante (VINAY KUMAR et al., 2014). O resultado da análise das pontes de hidrogênio intermoleculares foi apresentado na Figura 23.

Hydrogen Bonds

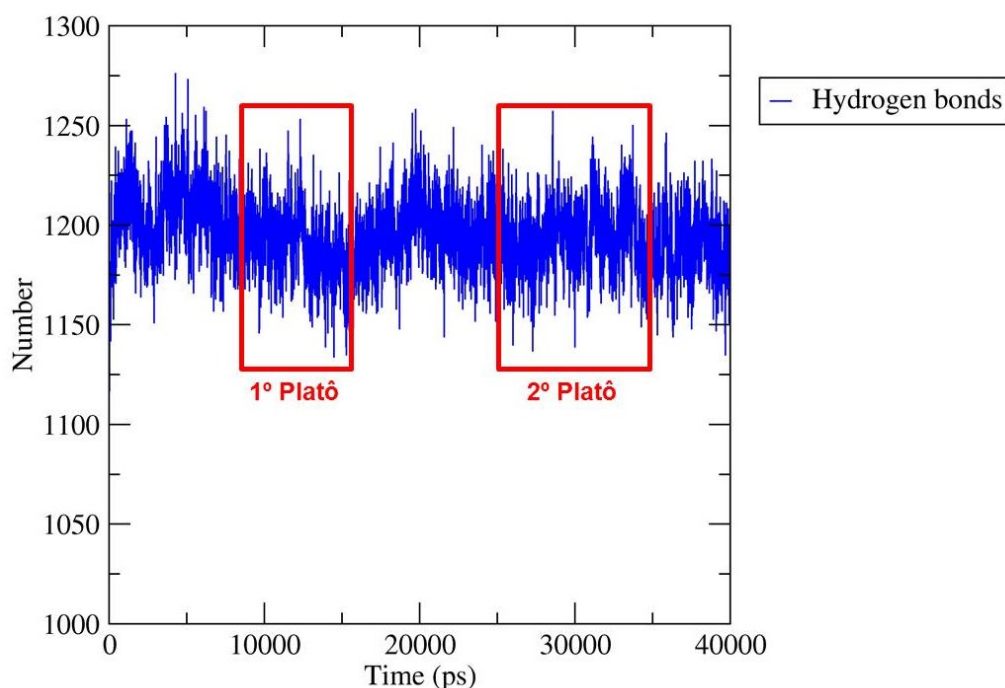


Figura 23 – Pontes de hidrogênio Intermoleculares formadas ao longo da simulação de dinâmica molecular.

O número de pontes de hidrogênio intermoleculares formadas são mostradas como função do tempo, em picosegundos, para a proteína nativa a 300K.

A proteína apresentou uma grande quantidade de pontes de hidrogênio formadas com o solvente durante a dinâmica molecular, dispondo de número médio de 1191 pontes no primeiro platô e 1194 pontes no segundo platô.

5.6.6. Superfície acessível ao solvente

Além da análise de pontes de hidrogênio intermoleculares, a superfície de acesso ao solvente também auxilia na caracterização do perfil de interação da proteína com o meio. A SASA é definida como a área de superfície disponível a moléculas de solvente (KHAN et al., 2016), podendo ser entendida como a superfície exposta de determinada proteína que possui capacidade de interagir com outras moléculas do meio (KUMAR et al., 2014), o que inclui, por definição, o solvente, mas que pode abranger também substratos enzimáticos e até mesmo

outras enzimas (RICHARDS, 1979; TAKAHASHI et al., 2000). O resultado da análise da superfície acessível ao solvente foi mostrado na Figura 24.

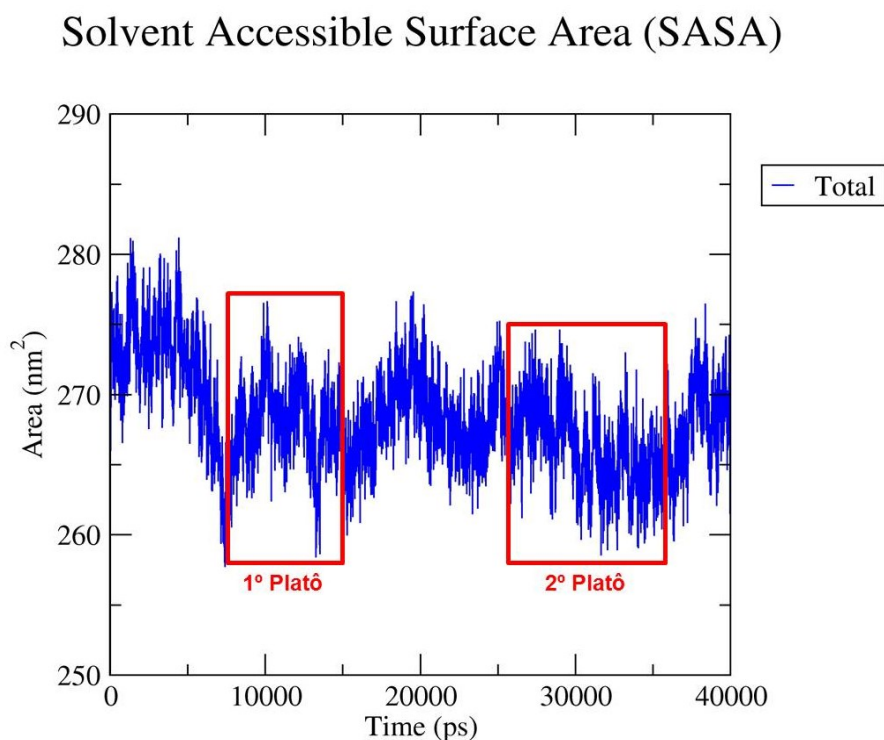


Figura 24 – Superfície acessível ao solvente ao longo da simulação de dinâmica molecular.

A área de superfície acessível ao solvente, em nanômetros quadrados, é mostrada em função do tempo, em picosegundos, para a proteína nativa a 300K.

A área média de superfície acessível ao solvente da proteína CDCA1 no primeiro platô, dispôs de 268nm², contra 267nm² de área do segundo platô. Nesse sentido, não há diferença nos valores da área entre os platôs indicados na figura 24.

5.6.7. Análise da dinâmica molecular

Nesta sessão, os resultados da dinâmica foram interrelacionados, viabilizando uma discussão mais completa e concisa, que favoreça maior compreensão das alterações estruturais e suas implicações funcionais para a proteína.

O aumento comparativo de flexibilidade no domínio N-terminal foi consenso nas análises de B-factor e RMSF. Ainda observou-se que tais alterações incidiam sobre a região média da estrutura – compreendida entre os resíduos 246 e

363 da proteína – local no qual existe uma função de mobilidade correlacionada com o seu papel enzimático. A flexibilidade de biomoléculas está associada à maioria dos fenômenos biológicos (VERLI, 2014), permitindo interações com ligantes e outras proteínas ou ácidos nucleicos, resultando na formação de estruturas complexas (CRAVEUR et al., 2015).

O segundo platô, de acordo com os resultados do RMSD, é mais rígido e, com base nos resultados do RG, é mais compacto. Isso pode estar associado com a entrada em um vale no qual o mínimo local está mais próximo do mínimo global de energia. Um mínimo local de energia é um ponto estacionário onde todos os deslocamentos implicam um aumento na energia. Um mínimo local também é chamado de mínimo global se corresponder ao isômero mais estável, ou seja, se não houver nenhum outro ponto estacionário com menor energia na superfície de energia potencial (LAVOR, 2001).

Comparando o RG e o SASA dos dois platôs, observa-se um aumento da razão superfície/volume, uma vez que há a compactação da proteína no segundo platô, sem mudar a superfície acessível. Tal característica pode ocorrer devido ao maior número de reentrâncias na molécula, o que elevaria sua área, sem elevar seu volume na mesma proporção (RICHARDS, 1979).

Durante a simulação, não foram observadas grandes diferenças quanto às interações com o meio, de modo a não ser esperado, portanto, diferenças substanciais de fenômenos envolvidos com interações proteína-ligante.

6. Conclusão

Este trabalho forneceu uma estrutura tridimensional confiável e completa da proteína CDCA1, obtida por modelagem comparativa e *ab initio*, e validada por algoritmos de validação e de alinhamento estrutural.

A análise filogenética estrutural indicou que, por ser uma proteína específica para a situação de deficiência de Zn, a conservação talvez não seja tão indicativa. Muitos resíduos não foram analisados, devido à insuficiência de dados comparativos, mas os resíduos que foram analisados na região central da proteína mantiveram considerável conservação estrutural.

As simulações de dinâmica molecular foram bem sucedidas, viabilizando a estabilização do sistema e a obtenção de dados significativos. A trajetória sugere alterações de flexibilidade em diferentes regiões da proteína devido ao seu papel enzimático.

7. Referências Bibliográficas

ADAMIS, P. D. B.; PANEK, A. D.; ELEUTHERIO, E. C. A. Vacuolar compartmentation of the cadmium-glutathione complex protects *Saccharomyces cerevisiae* from mutagenesis. **Toxicology letters**, v. 173, n. 1, p. 1–7, 30 ago. 2007.

ADLE, D. J. et al. A cadmium-transporting P1B-type ATPase in yeast *Saccharomyces cerevisiae*. **Journal of Biological Chemistry**, v. 282, n. 2, p. 947–955, 2007.

AGUILERA, J. et al. Carbonic anhydrase (Nce103p): an essential biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at atmospheric carbon dioxide pressure. **The Biochemical journal**, v. 391, n. Pt 2, p. 311–316, 2005.

AMOROSO, G. et al. The gene NCE103 (YNL036w) from *Saccharomyces cerevisiae* encodes a functional carbonic anhydrase and its transcription is regulated by the concentration of inorganic carbon in the medium. **Molecular Microbiology**, v. 56, n. 2, p. 549–558, 2005.

ASHKENAZY, H. et al. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. **Nucleic acids research**, v. 38, n. Web Server issue, p. W529-33, jul. 2010.

ATSDR. **Toxicological Profile for Zinc**. Disponível em: <<http://www.atsdr.cdc.gov/toxprofiles/tp60.pdf>>. Acesso em: 27 maio. 2016.

BALLATORI, N. Transport of toxic metals by molecular mimicry. **Environmental Health Perspectives**, v. 110, n. SUPPL. 5, p. 689–694, 2002.

BAYRAM, E. et al. In vitro inhibition of salicylic acid derivatives on human cytosolic carbonic anhydrase isozymes I and {II}. **Bioorganic & Medicinal Chemistry**, v. 16, n. 20, p. 9101–9105, 2008.

BENKERT, P.; KÜNZLI, M.; SCHWEDE, T. QMEAN server for protein model quality estimation. **Nucleic Acids Research**, v. 37, n. SUPPL. 2, p. 510–514, 2009.

BERMAN, H. M. et al. The Protein Data Bank. **Nucleic acids research**, v. 28, n. 1, p. 235–242, 2000.

BERTIN, G.; AVERBECK, D. Cadmium: cellular effects, modifications of biomolecules, modulation of DNA repair and genotoxic consequences (a review). **Biochimie**, v. 88, n. 11, p. 1549–1559, 2006.

- BIASINI, M. et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, n. W1, p. 252–258, 2014.
- BLAZHENKO, O. V. et al. Accumulation of cadmium ions in the methylotrophic yeast *Hansenula polymorpha*. **BioMetals**, v. 19, n. 6, p. 593–599, 2006.
- BOND, G. M. et al. Development of Integrated System for Biomimetic CO₂ Sequestration Using the Enzyme Carbonic Anhydrase. **Energy & Fuels**, v. 15, n. 2, p. 309–316, 2001.
- BOYLE, E. A.; SCLATER, F.; EDMOND, J. M. On the marine geochemistry of cadmium. **Nature**, v. 263, n. 5572, p. 42–44, 2 set. 1976.
- BROMBERG, Y.; ROST, B. SNAP: predict effect of non-synonymous polymorphisms on function. **Nucleic acids research**, v. 35, n. 11, p. 3823–35, jan. 2007.
- BRULAND, K. W.; KNAUER, G. A.; MARTIN, J. H. Cadmium in northeast Pacific waters. **Limnology and Oceanography**, v. 23, n. 4, p. 618–625, 1978.
- CENTOFANTI, M. et al. COMPARATIVE EFFECTS ON INTRAOCULAR PRESSURE BETWEEN SYSTEMIC AND TOPICAL CARBONIC ANHYDRASE INHIBITORS: A CLINICAL MASKED, CROSS-OVER STUDY. **Pharmacological Research**, v. 35, n. 5, p. 481–485, 1997.
- CHANG, S. L.; LEU, J. Y. A tradeoff drives the evolution of reduced metal resistance in natural populations of yeast. **PLoS Genetics**, v. 7, n. 3, p. 1–10, 2011.
- COUTINHO, P. W. R. et al. Alternative Soil Remediation: Bioremediation and Phytoremediation. **Nucleus**, v. 12, n. 1, p. 59–68, 2015.
- COX, E. H. et al. The active site structure of *Thalassiosira weissflogii* carbonic anhydrase 1. **Biochemistry**, v. 39, n. 40, p. 12128–12130, 2000.
- CRAVEUR, P. et al. Protein flexibility in the light of structural alphabets. **Frontiers in Molecular Biosciences**, v. 2, n. May, p. 1–20, 2015.
- DAVY, R. Development of catalysts for fast, energy efficient post combustion capture of CO₂ into water; an alternative to monoethanolamine (MEA) solvents. **Energy Procedia**, v. 1, n. 1, p. 885–892, 2009.
- DFG, D. F. Cadmium and its compounds (in the form of inhale dusts/aerosols). **The MAK collection for occupational health and safety**, v. 22, 2006.
- DILMORE, R. et al. Carbonic anhydrase-facilitated {CO₂} absorption with polyacrylamide buffering bead capture. **International Journal of Greenhouse Gas Control**, v. 3, n. 4, p. 401–410, 2009.
- DURÃO, V. Revisão terapêutica farmacologia renal dos diuréticos. **Acta Médica Portuguesa**, p. 495, 1979.

EISENBERG, D.; LÜTHY, R.; BOWIE, J. U. VERIFY3D: assessment of protein models with three-dimensional profiles. **Methods in enzymology**, v. 277, p. 396–404, 1997.

EKINS, S.; MESTRES, J.; TESTA, B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. **British Journal of Pharmacology**, v. 152, n. December 2006, p. 9–20, 2007.

FAVRE, N.; CHRIST, M. L.; PIERRE, A. C. Biocatalytic capture of {CO₂} with carbonic anhydrase and its transformation to solid carbonate. **Journal of Molecular Catalysis B: Enzymatic**, v. 60, n. 3–4, p. 163–170, 2009.

FERNANDEZ-FUENTES, N. et al. M4T: a comparative protein structure modeling server. **Nucleic acids research**, v. 35, n. Web Server issue, p. W363-8, jul. 2007.

FIGUEROA, J. D. et al. Advances in CO₂ capture technology-The U.S. Department of Energy's Carbon Sequestration Program. **International Journal of Greenhouse Gas Control**, v. 2, n. 1, p. 9–20, 2008.

FILIPIC, M. Mechanisms of cadmium induced genomic instability. **Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis**, v. 733, n. 1–2, p. 69–77, 2012.

GITAN, R. S. et al. A cytosolic domain of the yeast Zrt1 zinc transporter is required for its post-translational inactivation in response to zinc and cadmium. **Journal of Biological Chemistry**, v. 278, n. 41, p. 39558–39564, 2003.

GOMES, D. S. et al. Regulation of cadmium uptake by *Saccharomyces cerevisiae*. **Biochimica et Biophysica Acta - General Subjects**, v. 1573, n. 1, p. 21–25, 2002.

GOPALAKRISHNAN, K. et al. **Ramachandran plot on the web (2.0).Protein and peptide letters**, 2007.

GOYER, R. A.; LIU, J.; WAALKES, M. P. **Cadmium and cancer of prostate and testis**. *BioMetals*. **Anais...**2004

GOYER, R. A; CLARKSON, T. W. Toxic Effects of Metals. **Casarett and Doull's Toxicology, The Basic Science of Poison**, p. 811–867, 2001.

HAAS, J. et al. The protein model portal - A comprehensive resource for protein structure and model information. **Database**, v. 2013, n. 8, p. 1–8, 2013.

HARTWIG, A. Mechanisms in cadmium-induced carcinogenicity: Recent insights. **BioMetals**, v. 23, n. 5, p. 951–960, 2010.

HENRY, R. P. Multiple Roles of Carbonic Anhydrase in Cellular Transport and Metabolism. **Annual Review of Physiology**, v. 58, n. 1, p. 523–538, 1996.

HEWETT-EMMETT, D.; TASHIAN, R. E. Functional Diversity, Conservation, and Convergence in the Evolution of the α -, β -, and γ -Carbonic Anhydrase Gene Families. **Molecular Phylogenetics and Evolution**, v. 5, n. 1, p. 50–77, 1996.

HOUGHTON JT et al. Climate Change 2001: The Scientific Basis. **Climate Change 2001: The Scientific Basis**, p. 881, 2001.

HU, J. et al. Renal cell carcinoma and occupational exposure to chemicals in Canada. **Occupational Medicine**, v. 52, n. 3, p. 157–164, 2002.

HUANG, T.-T. et al. (PS)2: protein structure prediction server version 3.0. **Nucleic acids research**, v. 43, n. W1, p. W338-42, 1 jul. 2015.

IARC, International Agency for Research on Cancer. **Monograph summary**, v. 100C, 2012.

ICDA, ICdA, International Cadmium Association, 2011. Disponível em <http://www.cadmium.org>

JÄRUP, L.; ÅKESSON, A. Current status of cadmium as an environmental health problem. **Toxicology and Applied Pharmacology**, v. 238, n. 3, p. 201–208, 2009.

KÄLLBERG, M. et al. Template-based protein structure modeling using the RaptorX web server. **Nature protocols**, v. 7, n. 8, p. 1511–22, ago. 2012.

KÄLLBERG, M. et al. RaptorX server: a resource for template-based protein structure modeling. **Methods in molecular biology (Clifton, N.J.)**, v. 1137, p. 17–27, jan. 2014.

KELLEY, L. A.; STERNBERG, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. **Nature protocols**, v. 4, n. 3, p. 363–71, jan. 2009.

KERN, A. L. et al. The function of Alr1p of *Saccharomyces cerevisiae* in cadmium detoxification: Insights from phylogenetic studies and particle-induced X-ray emission. **BioMetals**, v. 18, n. 1, p. 31–41, 2005.

KHAN, F. I. et al. Current updates on computer aided protein modeling and designing. **International Journal of Biological Macromolecules**, v. 85, p. 48–62, 2016.

KIM, D. E.; CHIVIAN, D.; BAKER, D. Protein structure prediction and analysis using the Robetta server. **Nucleic Acids Research**, v. 32, n. WEB SERVER ISS., p. 526–531, 2004.

KNAPP, B. et al. Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? **Journal of computational biology : a journal of computational molecular cell biology**, v. 18, n. 8, p. 997–1005, 2011.

KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. W1, p. 294–297, 2012.

KREBS, B. B.; DE MESQUITA, J. F. Amyotrophic Lateral Sclerosis Type 20 - In Silico Analysis and Molecular Dynamics Simulation of hnRNPA1. **PloS one**, v. 11, n. 7, p. e0158939, jul. 2016.

LANE, T. W. et al. A cadmium enzyme from a marine diatom. **Nature**, v. 434, n. 7032, p. 455–456, 2005.

LANE, T. W.; MOREL, F. M. A biological function for cadmium in marine diatoms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 9, p. 4627–4631, 2000.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, abr. 1993.

LAVOR, C. C. **Uma abordagem determinística para a minimização global da energia potencial de molé- culas**. [s.l.] Universidade Federal do Rio de Janeiro, 2001.

LI, Z. S. et al. A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis(glutathionato)cadmium. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 1, p. 42–7, 1997.

LINDORFF-LARSEN, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins: Structure, Function and Bioinformatics**, v. 78, n. 8, p. 1950–1958, 2010.

LINDSKOG, S. Structure and mechanism of carbonic anhydrase. **Pharmacology & Therapeutics**, v. 74, n. 1, p. 1–20, 1997.

MARTINEZ, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. **PLOS ONE**, v. 10, n. 3, p. e0119264, mar. 2015.

MEATH, W. J. et al. Interaction Models for Water in Relation to Protein Hydration. **Nature**, n. May, p. 331–338, 1981.

MELDRUM, N. U.; ROUGHTON, F. J. Carbonic anhydrase. Its preparation and properties. **The Journal of physiology**, v. 80, n. 2, p. 113–142, 1933.

MICHAEL P, W. Cadmium carcinogenesis in review. **Journal of inorganic biochemistry**, v. 79, n. 1–4, p. 241–244, 2000.

MIRJAFARI, P.; ASGHARI, K.; MAHINPEY, N. Investigating the Application of Enzyme Carbonic Anhydrase for CO₂ Sequestration Purposes. **Industrial & Engineering Chemistry Research**, v. 46, n. 3, p. 921–926, 2007.

MOHAMMADIAN FAZLI, M. et al. Highly cadmium tolerant fungi: their tolerance and removal potential. **Journal of environmental health science & engineering**, v. 13, n. July, p. 19, 2015.

NORDBERG, G. F. Historical perspectives on cadmium toxicology. **Toxicology and Applied Pharmacology**, v. 238, n. 3, p. 192–200, 2009.

NTP, N. T. P. No Title. **Tenth Report on Carcinogens, Department of Health and Human Services**, v. III-42–III, n. Triangle Park, NC, 2000.

OBBER, J. A. Mineral commodity summaries 2016. **US Geological Survey**, 2016.

PENNINCKX, M. J. An overview on glutathione in *Saccharomyces* versus non-conventional yeasts. **FEMS Yeast Research**, v. 2, n. 3, p. 295–305, 2002.

PESCH, B. et al. Occupational risk factors for renal cell carcinoma: agent-specific results from a case-control study in Germany. **International Journal of Epidemiology**, v. 29, n. 6, p. 1014–1024, 2000.

PIKEMAAT, M. G. et al. Molecular dynamics simulations as a tool for improving protein stability. **Protein engineering**, v. 15, n. 3, p. 185–192, 2002.

PRADHAN, A. et al. Polyhydroxyfullerene binds cadmium ions and alleviates metal-induced oxidative stress in *Saccharomyces cerevisiae*. **Applied and Environmental Microbiology**, v. 80, n. 18, p. 5874–5881, 2014.

PRICE, N. M.; MOREL, F. M. M. **Cadmium and cobalt substitution for zinc in a marine diatom** *Nature*, 1990.

PUNDIR, S.; MARTIN, M. J.; O'DONOVAN, C. UniProt Protein Knowledgebase. In: WU, C. H.; ARIGHI, C. N.; ROSS, K. E. (Eds.). **Protein Bioinformatics: From Protein Modifications and Networks to Proteomics**. New York, NY: Springer New York, 2017. p. 41–55.

RICHARDS, F. M. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure. **Carlsberg Research Communications**, v. 44, n. 2, p. 47–63, 1979.

RODRIGUES, J. P. G. L. M.; LEVITT, M.; CHOPRA, G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. **Nucleic acids research**, v. 40, n. Web Server issue, p. W323-8, jul. 2012.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, v. 5, n. 4, p. 725–38, abr. 2010.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nat Protoc**, v. 5, n. 4, p. 725–738, 2011.

ROY, A.; YANG, J.; ZHANG, Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. **Nucleic acids research**, v. 40, n. Web Server issue, p. W471-7, jul. 2012.

RYKUNOV, D. et al. Improved scoring function for comparative modeling using the M4T method. **Journal of structural and functional genomics**, v. 10, n. 1, p. 95–9, mar. 2009.

SATARUG, S. et al. Cadmium, environmental exposure, and health outcomes. **Environmental Health Perspectives**, v. 118, n. 2, p. 182–190, 2010.

SATARUG, S.; MOORE, M. R. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. **Environmental Health Perspectives**, v. 112, n. 10, p. 1099–1103, 2004.

SCHWARTZ, G. G.; REIS, I. M. Is cadmium a cause of human pancreatic cancer? **Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology**, v. 9, n. 2, p. 139–45, 2000.

SERERO, A. et al. Yeast genes involved in cadmium tolerance: Identification of DNA replication as a target of cadmium toxicity. **DNA Repair**, v. 7, n. 8, p. 1262–1275, 2008.

SHAKIR, H. A.; SHAKEEL, S.; QAZI, J. I. Review Article Bioremediation of some frequently encountered non essential heavy metals: a review. **Punjab University Journal of Zoology**, v. 28, n. 1, p. 45–54, 2013.

SHEKH, A. Y. et al. Recent Advancements in Carbonic Anhydrase–Driven Processes for CO₂ Sequestration: Minireview. **Critical Reviews in Environmental Science and Technology**, v. 42, n. 14, p. 1419–1440, 2012.

SHIRAIISHI, E. et al. The cadmium-resistant gene, CAD2, which is a mutated putative copper-transporter gene (PCA1), controls the intracellular cadmium-level in the yeast *S. cerevisiae*. **Current genetics**, v. 37, n. 2, p. 79–86, 2000.

SMITH, K. S. et al. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 26, p. 15184–15189, 1999.

SMITH, K. S.; FERRY, J. G. Prokaryotic carbonic anhydrases. **FEMS Microbiology Reviews**, v. 24, n. 4, p. 335–366, 2000.

STADIE, C.; ALLISON, J. Enzyme from Red Blood Cells. **Journal of Biological Chemistry**, p. 521–529, 1928.

STANLEY, N.; ESTEBAN-MARTÍN, S.; DE FABRITIIS, G. Progress in studying intrinsically disordered proteins with atomistic simulations. **Progress in biophysics and molecular biology**, mar. 2015.

STROP, P. et al. Crystal Structure of the “cab”-type β Class Carbonic Anhydrase from the Archaeon *Methanobacterium thermoautotrophicum*. **Journal of Biological Chemistry**, v. 276, n. 13, p. 10299–10305, 2001.

TAKAHASHI, H. et al. Catalytic activity in organic solvents and stability of immobilized enzymes depend on the pore size and surface characteristics of mesoporous silica. **Chemistry of Materials**, v. 12, n. 11, p. 3301–3305, 2000.

TOUW, W. G.; JOOSTEN, R. P.; VRIEND, G. New biological insights from better structure models. **Journal of Molecular Biology**, 2016.

TREVISOL, E. T. V.; DE MESQUITA, J. F. Análise filogenética da anidrase carbônica. **Universidade Severino Sombra**, 2007.

- TRIPP, B. C.; SMITH, K.; FERRY, J. G. Carbonic Anhydrase: New Insights for an Ancient Enzyme. **Journal of Biological Chemistry**, v. 276, n. 52, p. 48615–48618, 2001.
- VAN DER SPOEL, D. et al. GROMACS: fast, flexible, and free. **Journal of computational chemistry**, v. 26, n. 16, p. 1701–18, dez. 2005.
- VERLI, H. Bioinformática da Biologia à Flexibilidade Molecular. **Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq**, 1 ed, p. 282, 2014.
- VINAY KUMAR, C. et al. Protein aggregation due to nsSNP resulting in P56S VABP protein is associated with amyotrophic lateral sclerosis. **Journal of Theoretical Biology**, v. 354, p. 72–80, 2014.
- VINOBA, M. et al. Carbonic anhydrase conjugated to nanosilver immobilized onto mesoporous SBA-15 for sequestration of {CO₂}. **Journal of Molecular Catalysis B: Enzymatic**, v. 75, p. 60–67, 2012.
- WEI, Z. Hidden markov models for controlling false discovery rate in genome-wide association analysis. **Methods in molecular biology (Clifton, N.J.)**, v. 802, n. Chapter 18, p. 337–44, 2012.
- WHEELER, D. L. et al. Database resources of the National Center for Biotechnology Information. **Nucleic acids research**, v. 28, n. 1, p. 10–4, 2000.
- WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, n. SUPPL.2, p. 407–410, 2007.
- WYSOCKI, R.; TAMÁS, M. J. How *Saccharomyces cerevisiae* copes with toxic metals and metalloids. **FEMS microbiology reviews**, v. 34, n. 6, p. 925–951, 2010.
- XIONG, J. Essential Bioinformatics. **Cambridge University Press**, p. 362, 2006.
- XU, Y. et al. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. **Nature**, v. 452, n. 7183, p. 56–61, 2008.
- YANG, J.; ZHANG, Y. I-TASSER server: new development for protein structure and function predictions. **Nucleic acids research**, v. 43, n. W1, p. W174-181, 2015.
- ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC bioinformatics**, v. 9, p. 40, jan. 2008.
- ZHANG, Y.; SKOLNICK, J. TM-align: A protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, v. 33, n. 7, p. 2302–2309, 2005.
- ZHOU, L. et al. Repression of class i transcription by cadmium is mediated by the protein phosphatase 2A. **Nucleic Acids Research**, v. 41, n. 12, p. 6087–6097, 2013.