



Universidade Federal do Estado do Rio de Janeiro  
Centro de Ciências Biológicas e da Saúde  
Instituto Biomédico  
Programa de Pós-Graduação em Biologia Molecular e Celular

**Desenvolvimento de Biorremediadores de Biossensores para  
Detecção e Avaliação da Toxicidade em Ambientes  
Contaminados por Metais Pesados Utilizando Biologia  
Computacional e de Sistemas**

**Rodolfo Galhardo Antunes de Figueiredo**

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Joelma Freire de Mesquita

Rio de Janeiro

2017

**Rodolfo Galhardo Antunes de Figueiredo**

**Desenvolvimento de Biorremediadores de Biossensores para  
Detecção e Avaliação da Toxicidade em Ambientes  
Contaminados por Metais Pesados Utilizando Biologia  
Computacional e de Sistemas**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular 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<sup>a</sup>. Dr<sup>a</sup>. Joelma Freire de Mesquita

Rio de Janeiro

2017

Figueiredo, Rodolfo Galhardo Antunes de  
Desenvolvimento de Biorremediadores e Biossensores  
para Detecção e Avaliação da Toxicidade em Ambientes  
Contaminados por Metais Pesados Utilizando Biologia  
Computacional e de Sistemas / Rodolfo Galhardo Antunes  
de Figueiredo - Rio de Janeiro, 2017.

Xf. : il

Dissertação (Programa de Pós-Graduação em Biologia  
Molecular e Celular)- Universidade Federal do Estado do  
Rio de Janeiro - UNIRIO

1. Biorremediação. 2. Cádmiu. 3. Alr1. 4. CorA  
I. Título

UNIRIO

**Rodolfo Galhardo Antunes de Figueiredo**

**Desenvolvimento de Biorremediadores de Biossensores para  
Detecção e Avaliação da Toxicidade em Ambientes  
Contaminados por Metais Pesados Utilizando Biologia  
Computacional e de Sistemas**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular 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.

**Banca examinadora:**

Dr<sup>a</sup> Cláudia Alessandra Fortes Aiub [Doutora em Biologia (Biociências Nucleares)] - UNIRIO

Dr<sup>a</sup>. Cristiane Barbosa Rocha (Doutora em Química Biológica) - UNIRIO

Dr<sup>a</sup>. Patrícia Cristina dos Santos Costa [Doutora em Ciências Biológicas (Fisiologia)] -UNIRIO

Dr. Ricardo Felipe Alves Moreira (Doutor em Bioquímica) - UNIRIO

## ÍNDICE

1.	INTRODUÇÃO.....	11
1.1.	Biorremediação e Biossensores.....	11
1.2.	Cádmio.....	14
1.3.	Proteína ALR1 e Superfamília CorA.....	17
2.	OBJETIVO.....	22
3.	METODOLOGIA.....	22
3.1.	Predição do Molde.....	22
3.2.	Predição da Estrutura Secundária.....	22
3.3.	Predição de Regiões Intrinsecamente Desordenadas .....	23
3.4.	Alinhamento de Sequência e Modelagem Comparativa.....	23
3.5.	Refinamento e Validação.....	24
3.6.	Predição da Estrutura Quaternária.....	24
3.7.	Análise Filogenética.....	24
3.8.	Análise das Interações Proteína-Proteína.....	25
3.9.	Análise do Poro.....	25
4.	RESULTADOS E DISCUSSÃO.....	26
4.1.	Busca pelo Molde e Modelagem Comparativa.....	26
4.2.	Predição de Desordem e Estrutura Secundária.....	27
4.3.	Refinamento e Validação.....	30
4.4.	Análise da Estutura.....	32
4.5.	Análise do Poro .....	35
4.6.	Análise da Região Citoplasmática.....	41
4.7.	Análise de Conservação.....	43
4.8.	Análise do Interatoma Físico.....	47
5.	CONCLUSÃO.....	49
6.	REFERÊNCIAS BIBLIOGRÁFICAS.....	50

## LISTA DE FIGURAS

Figura 1. Modelo proposto para a função da proteína ALR1 em um ambiente contaminado por Cádmi (Cd) (KERN et al., 2005).....	20
Figura 2. Representação da superfamília CorA (ESHAGHI, 2006).....	21
Figura 3. Predição de Regiões Intrinsicamente Desordenadas (IDRs) segundo o MobiDB 2.0 para a proteína ALR1 (ALR1p).....	28
Figura 4. Predição de Regiões Intrinsicamente Desordenadas (IDRs) segundo o D <sup>2</sup> P <sup>2</sup> para a proteína ALR1 (ALR1p).....	28
Figura 5. Validação do modelo in silico da proteína ALR1 segundo o PROCHECK, RAMPAGE e ERRAT.....	31
Figura 6. Validação do modelo in silico da proteínas ALR1 segundo o TM-align e o ProSA-web.....	32
Figura 7. Estrutura da Proteína ALR1.....	33
Figura 8. Estrutura dos homólogos TmCorA (NORDIN et al., 2013) e MjCorA (GUSKOV et al., 2012).....	33
Figura 9. Extremidade N-terminal dos monômeros da CorA de acordo com a estrutura secundária.....	35
Figura 10. Representação bidimensional do TM1 para proteínas CorA.....	36
Figura 11. Alinhamento estrutural entre os resíduos polares não carregados presentes no TM1 das proteínas CorA. ....	38
Figura 12. Poro da proteína ALR1 no sentido C-terminal para N-terminal.....	39
Figura 13. Poro dos homólogos CorA da proteína ALR1.....	39
Figura 14. Alinhamento estrutural entre os resíduos hidrofóbicos.....	41
Figura 15. Alinhamento estrutural entre ALR1 e TmCorA para sítio de ligação a metal.....	42
Figura 16. Alinhamento estrutural entre ALR1 e MjCorA para sítio de ligação a metal.....	42
Figura 17. Mecanismo de ação da TmCorA e alinhamento estrutural do resíduo N288.....	44
Figura 18. Análise filogenética para a ALR1p.....	45
Figura 19. Interatoma acerca das interações proteicas da ALR1p.....	47
Figura 20. Proteínas do interatoma da ALR1p segundo a frequência nos bancos de dados de interação.....	48

## LISTA DE TABELAS

Tabela 1 - Resultado da busca pelo molde para a modelagem comparativa.....	27
Tabela 2 - Características dos Aminoácidos Treonina e Asparagina.....	37
Tabela 3 - Grau de Conservação dos Aminoácidos Importantes da ALR1p.....	46

## ABREVIATURA

Å	Angstrom
ALR1p	Proteína ALR1
BER	do inglês, <i>base excision repair</i>
CONAMA	Conselho Nacional do Meio Ambiente
DFG	do alemão, <i>Deutsche Forschungsgemeinschaft</i>
GSH	Glutathiona Reduzida
IARC	do inglês, <i>International Agency for Research on Cancer</i>
IDRs	do inglês, <i>Intrinsic Protein Regions</i>
IUPAC	do inglês, <i>International Union of Pure and Applied Chemistry</i>
MTs	Metalotioneínas
MIT	do inglês, <i>Metal Ion Transporter</i>
MjCorA	<i>Methanocaldococcus jannaschii</i>
MMR	do inglês, <i>mismatch repair</i>
NER	do inglês, <i>nucleotide excision repair</i>
NTP	do inglês, <i>National Toxicology Program</i>
ONU	Organização das Nações Unidas
OPM	do inglês, <i>Orientations of Proteins in Membranes</i>
PDB	do inglês, <i>Protein Data Bank</i>
PNMA	Política Nacional do Meio Ambiente
PPM	do inglês, <i>Positioning of Proteins in Membranes</i>
RMSD	do inglês, <i>Root Mean Square Deviaton</i>
ROS	do inglês, <i>Reactive Oxigen Species</i>
SEMA	Secretaria Especial do Meio Ambiente
SH	Grupamento sulfidrila
TM1	Primeiro segmento transmembrana
TM2	Segundo segmento transmembrana
TmCorA	<i>Thermotoga marítima</i>
USGS	do inglês, <i>United States of Geological Survey</i>

## RESUMO

Cádmio é um metal tóxico não essencial e abundante que é liberado no meio ambiente por efeitos naturais e atividades antropogênicas. Tolerância e desintoxicação frequentemente envolvem a saída de íons tóxicos da célula e a redução de sua captação do meio ambiente. Não existe nenhum sistema de transporte de influxo específico para o Cádmio em leveduras e Pca1p constitui a principal rota de saída de Cádmio em *S. cerevisiae*. Outros transportadores de membrana plasmática estão envolvidos na desintoxicação de metal, porém pouco se sabe sobre sua regulação ou modo de ação. Neste contexto, a proteína ALR1, a qual pertence a superfamília CorA, é responsável pela entrada de magnésio na célula e pode contribuir para a tolerância ao Cádmio. O objetivo do presente estudo foi investigar a proteína ALR1 como potencial alvo para o desenvolvimento de biorremediadores e biossensores. Para esse estudo, primeiramente buscou-se identificar a estrutura secundária da proteína ALR1 e seu grau de desordem. A estrutura terciária do modelo ALR1 foi modelada pelo Modeller, refinada segundo a predição da estrutura secundária e validada pelos programas ERRAT, PROCHECK, ProSa-web, Rampage e TM-align, e por fim, sua oligomerização foi realizada pelo programa GalaxyGemini. O canal do modelo ALR1 foi analisado pelo programa Mole 2.0 e sua estrutura foi comparada, pelo TM-align, com as únicas duas estruturas conhecidas da superfamília CorA, provenientes de seus homólogos procariotos *Methanocaldococcus jannaschii* e *Thermotoga marítima*. Ao final do estudo foi realizada a análise do grau de conservação filogenética da proteína ALR1 pelo ConSurf, bem como a análise das suas interações proteína-proteína pelo programa CytoScape. De acordo com este trabalho conclui-se que a ALR1p é de fato um homólogo distante dos organismos *Methanocaldococcus jannaschii* e *Thermotoga marítima*, visto que manteve a estrutura característica da superfamília CorA (2 TMs + GMS), além de possuir resíduos de aminoácidos conservados na mesma posição de resíduos considerados essenciais para seus homólogos. O interatoma desta proteína demonstrou somente duas outras proteínas, SSA1 e SSE1, em conformidade com uma intoxicação aguda pelo Cádmio. Porém, para ter certeza sobre o transporte de Cd pela ALR1p são necessários estudos posteriores visto que, assim como seus homólogos, a ALR1p está em seu estado conformacional fechado.

## ABSTRACT

Cadmium is a ubiquitous and non-essential heavy metal released in the environment by natural and anthropogenic activities. Tolerance and detoxification often involve the release of toxic ions of the cell and the reduction of its uptake from the environment. There is no specific uptake transport system for cadmium in yeast and Pca1p is the main route of Cadmium efflux. Others plasma membrane transporters are involved in metal detoxification, but little is known about their regulation or mode of action. In this context, ALR1p, which belongs to the CorA superfamily, is responsible of the entry of magnesium into the cell and may contribute to cadmium tolerance. The objective of the present study was to investigate the ALR1 protein as a potential target for the development of bioremediation and biosensors. For this study, we first sought to identify the secondary structure of the ALR1 protein and its degree of disorder. The tertiary structure of the ALR1 model was modeled by Modeller, refined according to the prediction of the secondary structure and validated by the ERRAT, PROCHECK, ProSa-web, Rampage and TM-align programs, and finally, its oligomerization was performed by the GalaxyGemini program. The channel of the ALR1 model was analyzed by the Mole 2.0 program and its structure was compared by TM-align with the only two known structures of the CorA superfamily from their prokaryote homologues *Methanocaldococcus jannaschii* and *Thermotoga maritima*. After, the phylogenetic analyses for ARL1p was performed by ConSurf as well as the analyses of its protein-protein interactions by the CytoScape program. According to this work, ALR1p is in fact a distant homologue of the organisms *Methanocaldococcus jannaschii* and *Thermotoga maritima*, since it maintained the structure characteristic of the superfamily CorA (2 TMs + GMS), besides possessing conserved amino acid residues in the same place considered essential for their homologues. The interactome of this protein has demonstrated only two other proteins, SSA1 and SSE1, in accordance with an acute intoxication by Cadmium. However, in order to be certain about the transport of Cd by ALR1p, further studies are required since, as well as its homologues, the ALR1p is in its closed conformational state.

# 1) INTRODUÇÃO

## 1.1) BIORREMEDIAÇÃO E BIOSSENSORES

A partir do desenvolvimento industrial iniciado na Inglaterra, no século XVIII, busca-se uma constante modernização de máquinas e equipamentos que facilitem a vida humana, processo este, de evolução intrínseco ao progresso do capitalismo e necessário frente ao constante aumento da população mundial, que gera uma demanda crescente por alimento, espaço e condições de vida. Em contra partida, a grande quantidade de lixo lançada de forma indiscriminada no ambiente pela atividade industrial, aliada a falta de uma legislação ambiental, acentuaram a contaminação do meio ambiente e trouxeram a necessidade da conscientização acerca da poluição ambiental. Preocupações com a qualidade do ar e da água surgiram previamente às preocupações do solo, sendo este o último a ser tratado na política ambiental. Acreditava-se que o solo fosse um receptor ilimitado de poluentes devido a sua capacidade superestimada de autodepuração e poder tampão, além de sua contaminação não ser visualmente percebida. Dessa forma, em 1972 foi realizada a Conferência de Estocolmo, orientada pela Organização das Nações Unidas (ONU) e considerada marco inicial da preocupação com a poluição ambiental (SILVA et al., 2008; CARNEIRO; GARIGLIO, 2010;).

Durante a Conferência foi criado o Programa Internacional de Educação Ambiental, que reconhece a educação ambiental como elemento chave para o combate à crise ambiental no mundo. Na época, estas medidas foram negligenciadas pelo Brasil na tentativa de se obter o desenvolvimento industrial, ocorrido principalmente a partir da década de 70, com o argumento de que a redução da atividade industrial era uma tentativa dos países de primeiro mundo de impedirem o desenvolvimento dos demais países. Impulsionadas pela Conferência de Estocolmo surgiram diversas normas acerca do meio ambiente, como a criação da Secretaria Especial do Meio Ambiente (SEMA) em 1973 e da Política Nacional do Meio Ambiente (PNMA) em 1981, porém, foi somente a partir de 1986 que o gerenciamento ambiental de fato se iniciou no Brasil, com a criação do Conselho Nacional do Meio Ambiente (CONAMA), que definiu o conceito de alteração ambiental como “qualquer mudança física, química ou biológica provocada pelas atividades humanas, que afetam o meio biológico, a qualidade dos recursos naturais, as atividades sócio-econômicas e a saúde pública”. Dois anos mais tarde a Constituição de 1988 define o conceito de gestão ambiental (FONTENELLE, 2004).

Apesar da criação de uma política do meio ambiente, um dos problemas ambientais atuais é a contaminação do ar, da água e do solo, como resultado de mais de 200 anos de má gestão do lixo industrial descartados inadequadamente no meio ambiente, de forma proposital ou acidental. A forma proposital decorre de ações antrópicas necessárias ao progresso do desenvolvimento humano, como as atividades industriais propriamente ditas e o crescimento populacional, que cada vez mais geram resíduos industriais ou domésticos. A forma de poluição acidental ocorre quando, ocasionalmente, acidentes na cadeia de produção industrial acontecem (WEBER; SANTOS, 2013). Vale destacar o caso da companhia Ingá, indústria de zinco, situada a 85 km do Rio de Janeiro, considerada o maior passivo ambiental do Estado do Rio de Janeiro. No terreno da mineradora, que faliu em 1998, foram abandonados 390 mil m<sup>3</sup> de efluentes líquidos, formando uma bacia com 260 mil m<sup>2</sup>. Parte vazou, contaminando terrenos próximos, as águas da Baía de Sepetiba e a vegetação do mangue com zinco, cádmio, mercúrio e chumbo, afetando seriamente a vida da população.

Em decorrência desta realidade torna-se cada vez mais necessário desenvolver e aplicar técnicas de remediação, que visam reduzir as concentrações de contaminantes a níveis seguros para a saúde humana, além de impedir a dispersão dos mesmos no ambiente. 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 microorganismos para neutralizar ou atenuar os contaminantes, e para isso baseia-se nas reações bioquímicas destes microorganismos para degradar, modificar ou remediar as substâncias tóxicas (COUTINHO et al., 2015). Dessa forma, os princípios básicos desta técnica são: a presença de microorganismos com capacidade de metabolizar o contaminante; a presença do contaminante no ambiente; e condições ambientais favoráveis a ação do microorganismo. Neste sentido, vale a pena destacar a importância do sinergismo metabólico na biodegradação dos contaminantes. Os microorganismos envolvidos no processo de biorremediação podem ser divididos em primários e secundários. Os primários metabolizam o contaminante presente no ambiente, e os secundários, os metabólitos decorrentes dos organismos primários, e assim sucessivamente. Sabendo que a biorremediação é considerada efetiva quando libera no meio ambiente, água e gás carbônico, substâncias inofensivas à saúde, a estratégia de sinergismo metabólico potencializa as chances de atenuação do contaminante (GAYLARDE; BELLINASSO; MANFIO, 2005).

Além disso, a biorremediação pode ser classificada de diversas maneiras, a depender da origem do microorganismo utilizado, do local onde é realizada e da técnica utilizada. Dependendo da quantidade ou da capacidade dos microorganismos presentes no local contaminado, de degradar o contaminante, pode ser necessário adicionar organismos exógenos ao sítio de contaminação para que a biorremediação ocorra. Os microorganismos presentes no sítio contaminado são chamados de autóctones e os introduzidos, de alóctones. Organismos alóctones, embora possam ser geneticamente modificados e assim adaptados a uma determinada função, como a degradação de um contaminante específico, eles não se encontram adaptados ao meio ambiente, sendo portanto susceptíveis a toxinas ou predadores. De acordo com o local onde ocorre, a biorremediação é classificada com *in-situ* se realizada no local contaminado, ou *ex-situ*, se realizada em um local diferente, sendo para isso necessários a remoção e o transporte do ambiente contaminado. Embora a biorremediação *in-situ* seja considerada mais econômica, ecológica e menos contaminante, a biorremediação *ex-situ* possibilita um maior controle das condições ambientais. A julgar pela técnica utilizada, a biorremediação *in-situ* pode ser natural ou acelerada, sendo a diferença entre elas a adição de nutrientes ou microorganismos no local contaminado, caracterizando a técnica de Bioestimulação e Bioaugmentação, respectivamente. Os nutrientes são adicionados para estimular a atividade microbiana. Entre as técnicas mais utilizadas nos processos de biorremediação *ex-situ* encontram-se o Landfarming, Biorreatores e Compostagem (WEBER; SANTOS, 2013)

Neste contexto de Biorremediação, o desenvolvimento de biossensores é de extrema importância para monitorar os níveis de contaminantes no ambiente, visto que biossensor é um dispositivo analítico que combina um elemento biológico, responsável por captar o poluente no meio ambiente, com um elemento transdutor de sinal, cuja função é converter a resposta biológica a um sinal proporcional à concentração do analito. O primeiro biossensor foi desenvolvido em 1962, por Leyland C. Clark, e consistiu na análise da glicose, cujo princípio é utilizado até hoje para a detecção e monitoramento dos níveis de glicemia, prática essencial para pessoas com diabetes. Um biossensor pode ser formado por vários elementos biológicos, como enzimas, anticorpos, células e microorganismos. Atualmente, os biossensores enzimáticos são os mais utilizados, e suas vantagens incluem uma alta especificidade e sensibilidade, apesar de apresentarem elevados custos e tempo necessários para a purificação da enzima. Por sua vez, os microorganismos oferecem uma alternativa na fabricação de biossensores pois podem ser cultivados massivamente em laboratório, além de detectarem uma grande variedade de substâncias químicas e possuírem uma ampla faixa

operacional de pH e temperatura, características que fornecem uma maior flexibilidade no uso de biossensores. Em comparação com células de outros organismos, as culturas microbianas são mais fáceis de serem manipuladas visto que apresentam uma maior viabilidade e estabilidade no ambiente *in vitro*. As muitas enzimas existentes nos microorganismos ao mesmo tempo em que possibilitam a capacidade de consumir um grande número de produtos químicos e favoreça o sinergismo metabólico, também causam uma perda na seletividade do biossensor. Dessa forma, embora os microorganismos não sejam tão específicos quanto as enzimas, os avanços da biotecnologia e do sequenciamento permitem o desenvolvimento de biossensores microbianos altamente específicos por meio do bloqueio induzido de vias metabólicas (LEI; CHEN; MULCHANDANI, 2006).

## 1.2) CÁDMIO

A busca pela industrialização e desenvolvimento humano levaram à poluição ambiental e consequentemente deterioração da saúde humana, devido à contaminação do ambiente por metais pesados provenientes de água residuais das indústrias, visto que qualquer atividade industrial que utilize metais em sua linha de produção possui metais em seus resíduos industriais (MOHAMMADIAN FAZLI et al., 2015). Atualmente, a poluição ambiental causada por metais pesados dos efluentes industriais é um dos maiores problemas em muitas cidades de alta densidade demográfica em todo o mundo (BREIEROVÁ et al., 2002). Os metais pesados não são biodegradáveis e tendem a se acumular nos organismos por meio da cadeia alimentar, causando impactos no ecossistema e na saúde humana (LI et al., 2013), sendo as indústrias de mineração, moagem e galvanização as principais fontes de poluição que descarregam uma variedade de metais no ambiente, como Cádmio, Cobre, Níquel, Cobalto, Zinco e Chumbo (FEREIDOUNI; DANESHI; YOUNESI, 2009).

O termo “Metal Pesado” geralmente é usado para nomear grupos de metais e semimetais que estão associados com contaminação e potencial toxicidade ou ecotoxicidade, embora nunca tenha sido definido pela União Internacional de Química Pura e Aplicada - IUPAC (do inglês, *International Union of Pure and Applied Chemistry*). O uso científico mais antigo do termo, encontrado em inglês, data de 1936 e define “Metal Pesado” de acordo com seu valor de densidade atômica, e para isso estabelece um limite mínimo de  $7\text{g/cm}^3$ . Posteriormente, diversos outros valores foram atribuídos, 3,5; 4; 5; 6 ( $\text{g/cm}^3$ ). Na tentativa de classificar um “Metal Pesado” foi atribuído um valor mínimo para seu peso

atômico, e novamente, diferentes valores foram sugeridos, como um limite maior do que 23 ou 40. Uma outra ideia foi classificar por meio do número atômico, estabelecendo um valor mínimo de 11, que é referente ao sódio. Esta classificação, porém, caracterizaria metais essenciais à vida, como Magnésio e Potássio, como metais pesados, e dessa forma iria de encontro a conceitos já estabelecidos. Ainda, uma quarta classificação aborda os metais pesados como metais tóxicos (DUFFUS, 2002). 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 g/cm<sup>3</sup>, 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 Cádmio nunca deixou de ser considerado um metal pesado.

O Cádmio (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 Cádmio pode ser emitido ao ambiente por atividades naturais ou antropogênicas. O intemperismo e erosão de rochas portadoras de Cádmio é a principal emissão natural. Outras fontes importantes são atividades vulcânicas e incêndios florestais. Em respeito às atividades antropogênicas, o Cádmio 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 (ICdA, 2011; IARC, 2012). Assim, o Cádmio 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 Cádmio 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).

De acordo com a exposição ao Cádmio, 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*) (NTP 2000; DFG 2006; IARC, 2012). 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 hematopoético 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 sulfidrila (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) PROTEÍNA ALR1 E SUPERFAMÍLIA CorA

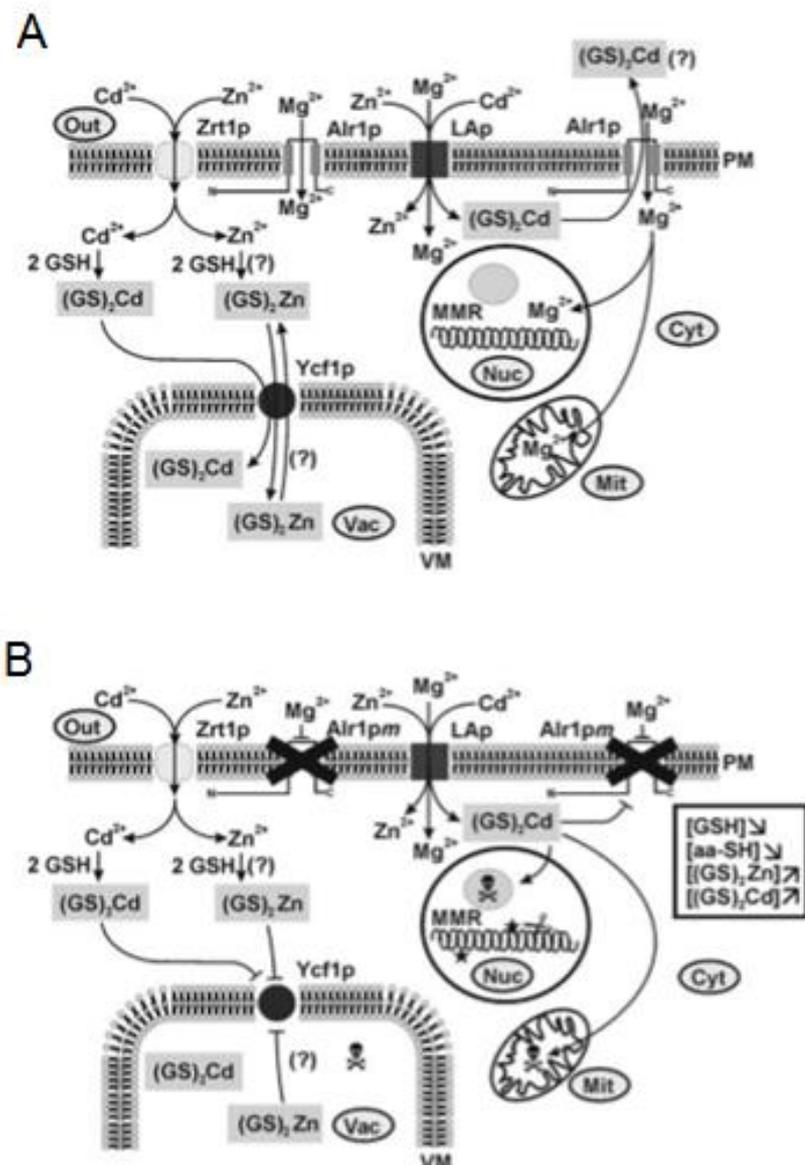
Metais são uma parte integrante do meio ambiente e estão amplamente difundidos na natureza, e os organismos tornam-se expostos aos metais através de fontes naturais ou antropogênicas. Embora compreendam 75% dos elementos da tabela periódica, poucos são os metais essenciais à vida, sendo a maioria tóxicos 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 nenhum papel biológico e são tóxicos mesmo em pequenas concentrações (SHAKIR; SHAKEEL; QAZI, 2013). Em particular, o Cádmio é um metal de transição altamente tóxico, conhecido como 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 *Thalassonia weissflogii*, o Cd não possui nenhum outro papel biológico conhecido, sendo considerado por isso, majoritariamente um metal não essencial (SERERO et al., 2008).

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, a levedura *Saccharomyces cerevisiae* tem sido utilizado 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 nos vacúolos destes íons, o que limita sua concentração no citoplasma. Em quarto lugar está a efluxo destes metais pesados por meio de transportadores (SHIRAIISHI 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,

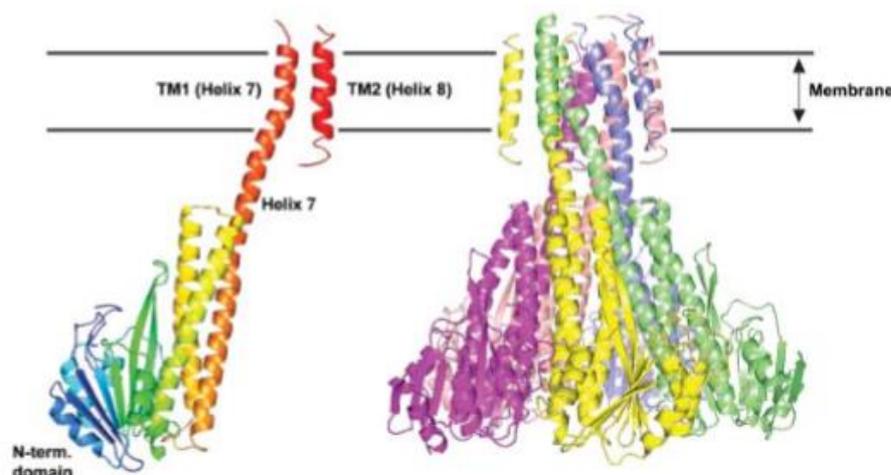
todos, necessários para a absorção de metais essenciais e outros nutrientes, como Ferro (Fe), Manganês (Mn), Zinco (Zn), fosfato, sulfato e glicerol, e outros. Assim, o Cádmio entra nas células através destas proteínas transportadoras, tais como Zrt1p, Smf1p e Smf2p, Fet4p e Mid1p, que transportam, respectivamente, Zn, Mn, Fe e Cálcio (Ca). 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 glutathione e gera o bis glutathionato de Cádmio [(GS)<sub>2</sub>Cd], 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 glutathione (LI et al., 1997). De acordo com a literatura, PCA1p 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. Dessa maneira, a proteína ALR1 (Alr1p), que é responsável pela absorção de Magnésio (Mg) nas células, também pode contribuir para a tolerância do Cd, visto que a mutação do gene ALR1 resulta na sensibilidade ao Cd e no aumento intracelular de sua concentração. De fato, a integração de análises filogenéticas e dados biológicos sugere claramente que Alr1p atua na detoxificação de metais pesados da célula, além do mais, um estudo realizado por Kern e colaboradores, 2005, demonstrou que a proteína ALR1 possui a capacidade de executar o efluxo de Cd para o meio extracelular e por isso está envolvida no processo de detoxificação celular ao Cádmio (KERN et al., 2005). A figura 1 ilustra a presença de metais pesados no meio ambiente, e entrada no ambiente intracelular por meio da captação de metais essenciais, os processos de intoxicação pelo Cd e de detoxificação celular e o possível efluxo de Cd por meio da proteína ALR1.



**Figura 1. Modelo proposto para a função da proteína ALR1 em um ambiente contaminado por Cádmio (Cd) (KERN et al., 2005).** (A) O Cd é sequestrado para o meio intracelular e se ligam à glutationa (GSH) gerando o compleo  $[(GS)_2Cd]$ . Estes complexos são compartimentalizados em vacúolos ou são expulsos para o meio extracelular. (B) O acúmulo de Cd na célula esgota as reservas citoplasmáticas antioxidantes levando a danos celulares por meio de espécies reativas de oxigênio, além de bloquear o processo de compartimentalização. Abreviações: ambiente extracelular (Out), núcleo (Nuc), mitocôndria (Mit), citoplasma (Cyt), membrana plasmática (PM), membrana vacuolar (VM). O símbolo (?) representa um caminho hipotético.

ALR1p é uma proteína de membrana plasmática integral e essencial para a captação de  $Mg^{2+}$  em células de levedura, mas também possui afinidade para captação de outros cátions divalentes como  $Ni^{2+}$ ,  $Mn^{2+}$ ,  $Zn^{2+}$  and  $Co^{2+}$  (LIM et al., 2011). Trata-se de um transportador passivo, tal como um canal, cuja função depende da manutenção da diferença de potencial elétrico através da membrana plasmática. Essa proteína possui 859 resíduos de aminoácidos e é codificada pelo gene ALR1, cujo nome deriva da sua resistência ao

Alumínio, visto que foi primeiramente identificada a partir da sua capacidade de aumentar a tolerância de células de levedura à presença de  $Al^{3+}$  (LIU et al., 2002). Além disso, a ALR1p foi o primeiro candidato conhecido para o sistema de transporte de  $Mg^{2+}$  em células eucarióticas, e na literatura, é conhecida como um homólogo distante da superfamília CorA transportadora de  $Mg^{2+}$ . A superfamília CorA também é conhecida como Transportadores de Íons Metálicos - MIT (do inglês, *Metal Ion Transporter*), e corresponde a um importante grupo de transportadores de Magnésio em procaríotos e eucariotos, cujos membros são caracterizados pela presença de dois segmentos transmembrana adjacentes (TM1 e TM2) próximas a porção C-terminal da proteína, e um motivo proteico formado de três resíduos de aminoácidos, Glicina (G), Metionina (M) e Asparagina (N), e por isso conhecido como GMN. Este motivo encontra-se ao final do TM1. Apesar desta superfamília abranger todos os reinos, existem somente dois organismos, *Methanocaldococcus jannaschii* e *Thermotoga marítima*, cujas estruturas proteicas completas estão descritas na literatura. Segundo estes dados, a estrutura da superfamília CorA é formada por um homopentâmero nos quais os segmentos transmembranares C-terminais agrupam-se em um conjunto total de 10 segmentos para formar o poro que atravessa a membrana plasmática. As regiões N-terminais dos monômeros constituem uma região citoplasmática cujo formato se assemelha a um funil, além de incorporarem vários sítios de ligação a metais (PISAT; PANDEY; MACDIARMID, 2009). Segundo a literatura, entre as estruturas completas existentes da família CorA, a proteína da *Thermotoga marítima* é a mais utilizada para estudar a superfamília CorA, tornando-se assim um modelo representativo. A figura 2 ilustra a estrutura da superfamília CorA segundo o modelo de *Thermotoga marítima*.



**Figura 2. Representação da superfaília CorA (ESHAGHI, 2006).** A esquerda é o monômero, colorido em uma escala de azul para vermelho, de acordo com a porção N e C-terminal, respectivamente. Na direita tem-se

o homo-pentâmero, colorido segundo as cinco cadeias. Pode-se observar a estrutura de um funil e duas regiões transmembrana, além da hélice que forma o poro do canal.

## 2) OBJETIVO

Desenvolver biorremediadores e biossensores para detecção e avaliação da toxicidade em ambientes contaminados por metais pesados utilizando biologia computacional e de sistemas tendo como alvo proteínas de *Saccharomyces cerevisiae*.

## 3) METODOLOGIA

### 3.1) PREDIÇÃO DO MOLDE

Para descobrir o melhor molde a ser utilizado na modelagem comparativa foi realizado um BLAST (ALTSCHUL et al., 1997) pelo site UniProtKB (BATEMAN et al., 2015), por meio do qual buscou-se homólogos da proteína ALR1 no banco de dados de proteína - PDB (do inglês, *Protein Data Bank*). Além disso, como a ALR1p pertence a superfamília CorA, foi realizada uma busca no PDB por todas as possíveis estrutura homólogas desta proteína. Para esta busca foi utilizado o termo “Magnesium transport protein CorA”, como o próprio banco sugere.

### 3.2) PREDIÇÃO DA ESTRUTURA SECUNDÁRIA

A estrutura secundária foi predita por meio do consenso entre muitos programas que realizam este tipo de predição. Os programas utilizados foram: PSIPRED (BUCHAN et al., 2013), Jpred (DROZDETSKIY et al., 2015), Jufo (MEILER; BAKER, 2003), RePROF (ROST; LIU, 2003), PROFsec (ROST; LIU, 2003), NetSurf (PETERSEN et al., 2009), PREDATOR (FRISHMAN; ARGOS, 1996), DSC (KING; STERNBERG, 1996), GOR V (SEN et al., 2005), SSpro (POLLASTRI et al., 2002) e PHD (KING; STERNBERG, 1996). Segundo a literatura, combinar o resultado de múltiplos algoritmos de predição secundária aumenta a acurácia da predição (PALOPOLI et al., 2009). Desta forma, cada resíduo de aminoácido da proteína

ALR1 teve sua estrutura secundária predita por cada programa, sendo o tipo de estrutura consenso obtido por maioria simples.

### 3.3) PREDIÇÃO DE REGIÕES INTRINSICAMENTE DESORDENADAS

Em associação com a estrutura secundária foi realizado o estudo de regiões intrinsecamente desordenadas - IDRs (do inglês, *Intrinsic Protein Regions*) da proteína ALR1, com o objetivo de conhecer seu perfil de desordem, visto que IDRs não possuem uma estrutura terciária nativa e por isso não podem ser modeladas (OROSZ; OVÁDI, 2011). Esta predição foi realizada por meio dos banco de dados MobiDB 2.0 (POTENZA et al., 2015) e D<sup>2</sup>P<sup>2</sup> (OATES et al., 2013), acumulando um total de 12 preditores de desordem.

### 3.4) ALINHAMENTO DE SEQUÊNCIA E MODELAGEM COMPARATIVA

Após a busca pelo molde, foi realizado um alinhamento de sequência por meio dos programas CLUSTALO (SIEVERS et al., 2011) e MUSCLE (EDGAR, 2004), usando o molde de maior qualidade e a proteína ALR1, sendo os alinhamentos gerados avaliados pelo programa TCS (CHANG; DI TOMMASO; NOTREDAME, 2014) e o de maior identidade foi selecionado. Então, este alinhamento foi utilizado na modelagem comparativa, cujo modelo construído era prontamente validado pela sua estrutura secundária. O processo de alinhamento, modelagem e validação, foi repetido inúmeras vezes até o modelo obtido possuir uma total concordância com a predição da estrutura secundária, sendo cada alinhamento avaliado por meio do programa TCS, onde não foi permitido alinhamento com valor inferior ao alinhamento original, logo, todos os alinhamentos refinados manualmente foram de qualidade superior ou idêntica ao alinhamento gerado pelos programas já consagrados na literatura, CLUSTALO ou MUSCLE. Tal procedimento já se mostrou útil na resolução de vários problemas de alinhamento entre a estrutura e sequências de proteínas (GUENTHER et al., 1997; SÁNCHEZ; ŠALI, 1997). Por último, o alinhamento validado foi usado na modelagem comparativa para construir um modelo por meio do programa Modeller (ESWAR et al., 2006). Modelagem comparativa constrói uma estrutura tridimensional de uma proteína alvo, nesse caso, ARL1p, baseado no alinhamento de sequência entre proteínas, onde uma é o alvo, que é a proteína a ser modelada, e a outra é o molde, homóloga à

proteína alvo. O processo de modelagem comparativa consiste em alinhar a proteína alvo com o molde, reconhecer a estrutura nativa da proteína molde, construir o modelo e avaliá-lo (MARTÍ-RENOM et al., 2000).

### 3.5) REFINAMENTO E VALIDAÇÃO

O modelo foi globalmente e localmente otimizado por meio dos programas GalaxyRefine (HEO; PARK; SEOK, 2013) e Modeller, respectivamente. O programa GalaxyRefine refina a proteína alvo por meio da reconstrução das cadeias laterais dos resíduos de aminoácidos e subsequente relaxamento da estrutura nativa da proteína por meio de dinâmica molecular (HEO; PARK; SEOK, 2013). Os resíduos remanescentes em regiões locais de baixa qualidade estereoquímica foram então refinados pelo programa Modeller. Após o refinamento, o modelo ALR1 foi validado pelos programas PROCHECK (LASKOWSKI et al., 1993), RAMPAGE (LOVELL et al., 2003), ERRAT (COLOVOS; YEATES, 1993) e ProSA-web (WIEDERSTEIN; SIPPL, 2007) e Tm-align (ZHANG; SKOLNICK, 2005).

### 3.6) PREDIÇÃO DA ESTRUTURA QUATERNÁRIA

Uma vez modelado e validado o monômero da proteína ALR1, foi realizada sua oligomerização pelo programa GalaxyGemini. Este programa, a partir de um monômero como dado de entrada, analisa sua homopolimerização baseado em oligômeros homólogos, segundo uma similaridade de estruturas terciária e quaternária (LEE et al., 2013).

### 3.7) ANÁLISE FILOGENÉTICA

O grau de conservação evolutiva da proteína ALR1 foi obtido pelo programa ConSurf (ASHKENAZY et al., 2016). Este programa realiza uma busca por homólogos usando o algoritmo BLAST, o qual permite comparar uma sequência alvo com milhares de sequências disponíveis em bancos de dados. No caso do ConSurf, a similaridade entre a sequência alvo e as demais é utilizada para conhecer a conservação dos resíduos de aminoácido da proteína alvo. Esta análise foi realizada contra o banco de dados UNIPROT com um uma

amplitude de 95% - 35% de identidade entre as sequências homólogas. Para o alinhamento múltiplo de sequência foi escolhido o método MUSCLE.

### 3.8) ANÁLISE DAS INTERAÇÕES PROTEÍNA-PROTEÍNA

No que tange o interatoma, foi analisada a relação física da proteína ALR1 com as demais proteínas de *Saccharomyces cerevisiae*, por meio de diversos bancos de dados: BioGRID (STARK et al., 2006), IntAct (KERRIEN et al., 2012), DIP (XENARIOS et al., 2002), Mint (LICATA et al., 2012), String (SZKLARCZYK et al., 2015), Mentha (CALDERONE; CASTAGNOLI; CESARENI, 2013), I2D (KOTLYAR et al., 2016), APID (PRIETO; DE LAS RIVAS, 2006) e PINA (COWLEY et al., 2012). Adicionalmente, todo o proteoma da *S. cerevisiae*, obtido no UNIPROT, foi analisado pelo programa GeneMANIA. Este programa vasculha bancos de dados com o objetivo de encontrar genes relacionados com o a proteína alvo, o que inclui as interações proteína-proteína (WARDE-FARLEY et al., 2010). As interações físicas da proteína ALR1 foram compiladas e então construiu-se um grafo, pelo programa Cytoscape 3.4.0, onde foram reunidas todas as interações obtidas. Cytoscape é uma interface gráfica que permite visualizar e construir redes de interação molecular e vias biológicas (SHANNON et al., 2003).

### 3.9) ANÁLISE DO PORO

Para a análise do poro foram utilizadas três abordagens. Primeiramente, foram conhecidos os resíduos de aminoácidos da proteína ALR1 que formam as duas regiões transmembrana, característica da superfamília CorA, e para isso foi utilizado o programa Posicionamento de Proteínas em Membranas – PPM (LOMIZE et al., 2006a) (do inglês, *Positioning of Proteins in Membranes*), do banco de dados Orientação de Proteínas na Membrana - OPM (do inglês, *Orientations of Proteins in Membranes*) (LOMIZE et al., 2006b). Sabendo que o poro é formado pela primeira hélice transmembrana (TM1), após saber quais resíduos formam esta região, a TM1 foi analisada pelo programa NetWheel (MÓL; FONTES, 2016). Este programa analisa a alfa hélice de forma bidimensional, e assim é possível conhecer como os diferentes aminoácidos se distribuem ao longo da alfa hélice. Por último,

as dimensões do poro da proteína ALR1 foram medidas por meio do programa Mole 2.0 (SEHNAL, 2013).

## 4) RESULTADOS E DISCUSSÃO

### 4.1) BUSCA PELO MOLDE E MODELAGEM COMPARATIVA

A busca pelo molde apresentou um total de 11 estruturas candidatas à molde provenientes de três organismos: *Archaeoglobus fulgidus*, *Methanocaldococcus jannaschii* e *Thermotoga marítima*. Porém, apesar de um total de 11 estruturas e três organismos encontrados, somente três estruturas provenientes de dois organismos, *Thermotoga marítima* e *Methanocaldococcus jannaschiia*, apresentam a proteína CorA completa. Além disso, o BLAST contra o *Protein Data Bank* demonstrou que o organismo *Thermotoga marítima* possui a maior similaridade de sequência com a ALR1, 24,1% contra 18,3% do organismo *Methanocaldococcus jannaschiia*, e por isso o organismo *Thermotoga marítima* foi selecionado como molde para a modelagem comparativa. O organismo *Archaeoglobus fulgidus* apresentou uma identidade de 20,8% com a proteína CorA. A tabela 1 reúne, além da estrutura molde (PDB ID: 4i0u), todas as 10 demais estruturas candidatas à molde.

**Tabela 1 – Resultado da busca pelo molde para a modelagem comparativa**

Organismo	UNIPROT ID	PDB ID	Estrutura CorA
<i>Thermotoga marítima</i>	Q9WZ31	2HN2	Incompleta
		2IUB	Incompleta
		2BBH	Incompleta
		2BBJ	Incompleta
		4I0U*	Completa
		4EEB	Incompleta
		4EED	Incompleta
<i>Methanocaldococcus jannaschii</i>	Q58439	4CY4	Completa
		4EGW	Incompleta
		4EV6	Completa
<i>Archaeoglobus fulgidus</i>	O29472	2HN1	Incompleta

\*Molde selecionado para a modelagem comparativa.

Se duas sequências distintas podem ser alinhadas com certo grau de similaridade é possível assumir que elas compartilharam um ancestral comum. Cabe ressaltar que a homologia não requer necessariamente alta identidade entre as sequências, uma vez que o

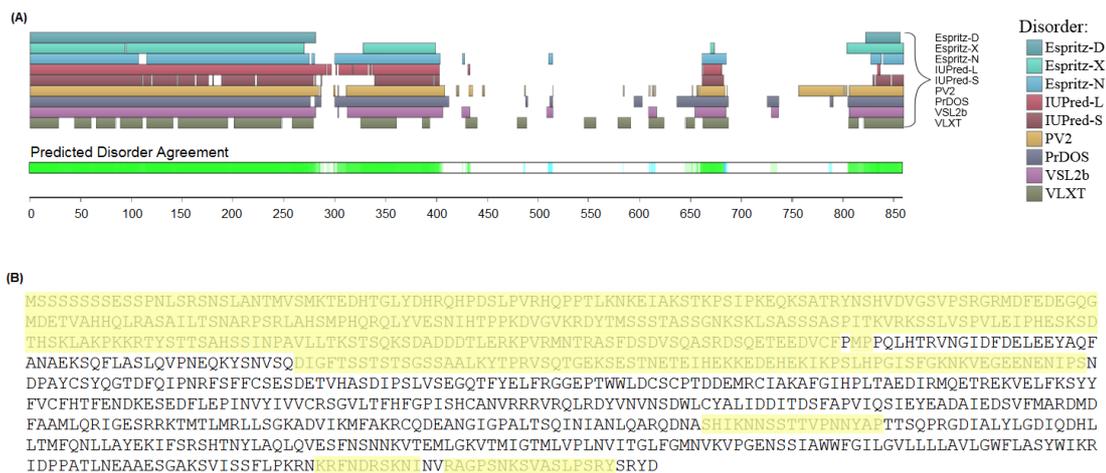
valor dessa identidade dependerá da taxa de evolução do organismo ou da espécie. No caso do proteína em estudo, ALR1, por ser uma proteína de *Saccharomyces cerevisiae* tendo como molde uma proteína CorA de procarioto, a taxa de evolução é algo a ser considerado, o que explica o baixo grau de identidade entre as sequências. A modelagem computacional pode ser dividida em dois métodos: independente de estrutura ou baseado em estrutura. O que vai ditar a escolha do método a ser aplicado é a presença, ou não, de estruturas resolvidas experimentalmente e depositadas em bancos de dados de estruturas. Quanto ao método baseado em estrutura, no qual a modelagem comparativa faz parte, por utilizar uma proteína molde para inferir a estrutura de uma proteína desconhecida, apresenta um limite atual de identidade de sequência em torno de 20%, valor inferior aos 24,1% de identidade entre a proteína ALR1 e seu homólogo procarioto TmCorA. Além disso, como a estrutura terciária de proteínas é mais conservada ao longo da evolução que a estrutura primária, proteínas com identidade muito baixa entre suas sequências podem possuir estruturas terciárias muito semelhantes (VERLI, 2014).

## 4.2) PREDIÇÃO DE DESORDEM E ESTRUTURA SECUNDÁRIA

A ALR1p possui 859 resíduos de aminoácidos, mais que o dobro do tamanho do seu molde (351 AAs), e somente um único domínio, o CorA. Porém, de acordo com o perfil de desordem estrutural da ALR1p apresentado pelos bancos de dados D<sup>2</sup>P<sup>2</sup> e MobiDB 2.0, nas figuras 3 e 4, respectivamente, é possível observar que salvo o domínio CorA, quase toda a proteína é desordenada. Fora deste domínio, há ainda a presença de uma pequena região ordenada próxima a porção N-terminal da proteína. Sabendo que regiões intrinsecamente desordenadas não possuem uma estrutura terciária estável, apenas a região correspondente ao domínio CorA foi modelada. Esta região engloba os resíduos 403-800 da proteína ALR1 e possui uma região desordenada. A falta de uma estrutura terciária estável é derivada dos aminoácidos que formam as IDRs, cuja composição possui uma alta porcentagem de resíduos polares e uma baixa porcentagem de hidrofóbicos, característica esta que dificulta o enovelamento proteico, até mesmo a formação de uma estrutura secundária de maior complexidade, como hélices e folhas beta (LINDING et al., 2003). De fato, todas as regiões desordenadas da proteína ALR1 se apresentaram como alças na predição da estrutura secundária.



**Figura 3. Predição de Regiões Intrinsecamente Desordenadas (IDRs) segundo o MobiDB 2.0 para a proteína ALR1 (ALR1p).** IDRs estão ilustradas em laranja, estrutura ordenada em azul e do domínio CorA em verde. **(A)** Predição de desordem por diferentes algoritmos e a região consenso abaixo do domínio CorA. **(B)** Predição de desordem segundo o consenso para a estrutura primária da ALR1p.



**Figura 4. Predição de Regiões Intrinsecamente Desordenadas (IDRs) segundo o D<sup>2</sup>P<sup>2</sup> para a proteína ALR1 (ALR1p).** **(A)** Os algoritmos de desordem estão expresso em diferentes cores. O nível de concordância entre os algoritmos está ilustrado em um gradiente de intensidade, na barra abaixo. Os segmentos em verde representam desordem fora do domínio CorA, e em azul, dentro deste domínio. Todas essas representações estão alinhadas com a cadeia peptídica representada em preto. **(B)** Predição de desordem, em amarelo, segundo o consenso dos diferentes preditores para a estrutura primária da ALR1p.

De acordo com o MobiDB 2.0, quase metade da ALR1p é desordenada, apresentando um grau de desordem no valor de 47,96%, assim, embora a ALR1p possua um total de 858 resíduos de aminoácidos, apenas pouco mais da metade dos seus resíduos é ordenada. Segundo a literatura, a frequência de desordem proteica aumenta de eubactéria para arqueobactéria para eucarioto, com uma frequência de 6-33%, 9-37% e 35-51%, respectivamente, sendo maior em eucariotos multicelulares do que unicelulares. Em geral, quanto maior a complexibilidade do organismo mais frequente é a desordem proteica (DUNKER et al., 2002; FENG et al., 2006). Esse comportamento pode ser explicado pela alta flexibilidade desse tipo de proteína que pode proporcionar uma melhor resposta a mudanças do ambiente do que proteínas rígidas e por uma maior necessidade de sinalização e regulação, que são funções associadas às IDRs (DUNKER et al., 2001).

Ainda, é possível observar que o domínio CorA localiza-se na região C-terminal da proteína ALR1, e como mencionado anteriormente, ocupa os resíduos de aminoácidos 403-800, o que lhe confere uma amplitude de 398 aminoácidos, valor bem próximo aos 351 aminoácidos do seu homólogo procarioto. Um domínio é uma parte da sequência polipeptídica de enovelamento independente (e, potencialmente, de função também independente). Assim, se um domínio for recortado de um gene e expresso separadamente ele deve, em princípio, manter suas características estruturais (VERLI, 2014). Dessa forma, todos os demais 460 resíduos de aminoácidos, que não pertencem ao domínio CorA, não devem exercer nenhum papel funcional ou estrutural em relação a este domínio. Desses 460, aproximadamente, 400 estão na porção N-terminal e 60 na porção C-terminal, separados pelo domínio CorA. Neste sentido, em um estudo feito por Jong-min e Richard, 2005, não foi encontrada nenhuma função para estas extremidades em bancos de dados para proteínas homólogas, além do truncamento da ALR1p demonstrar que os primeiros 239 resíduos de aminoácidos da porção N-terminal, e os 53 resíduos C-terminais, não são essenciais para a captação de magnésio pela proteína ALR1. Porém, para que ocorra uma captação efetiva de Magnésio pela proteína ALR1, parte do segmento de resíduos 240-322 da porção N-terminal precisa estar presente. Esta região não compreende o domínio CorA (LEE; GARDNER, 2006). Curiosamente, a única região ordenada da proteína ALR1 externa ao domínio CorA abrange os resíduos de aminoácidos 282-328, e parece corresponder a uma alfa hélice, segundo a predição da estrutura secundária.

### 4.3) REFINAMENTO E VALIDAÇÃO

A geração de estruturas por modelagem comparativa envolve uma série de substituições, inserções e deleções de aminoácidos. Dessa forma, mesmo com uma identidade de sequência elevada, a cadeia lateral dos aminoácidos é menos precisa do que a estrutura da cadeia principal, enquanto que com uma identidade de sequência mais baixa, as estruturas geradas podem ter erros em ambas as cadeias laterais e principal. Os programas de validação analisam a qualidade de uma estrutura proteica, e se o modelo gerado pela modelagem comparativa não apresentar uma qualidade mínima, sendo aprovado pelos programas de validação, este modelo precisa ser refinado (CALIXTO, 2013; HEO; PARK; SEOK, 2013). Assim, a estrutura final, modelada e refinada da proteína ALR1, foi avaliada segundo a geometria de sua cadeia principal pelos programas PROCHECK e RAMPAGE, as interações atômicas não covalentes pelo programa ERRAT, o perfil de energia pelo ProSAweb e o enovelamento pelo Tm-align.

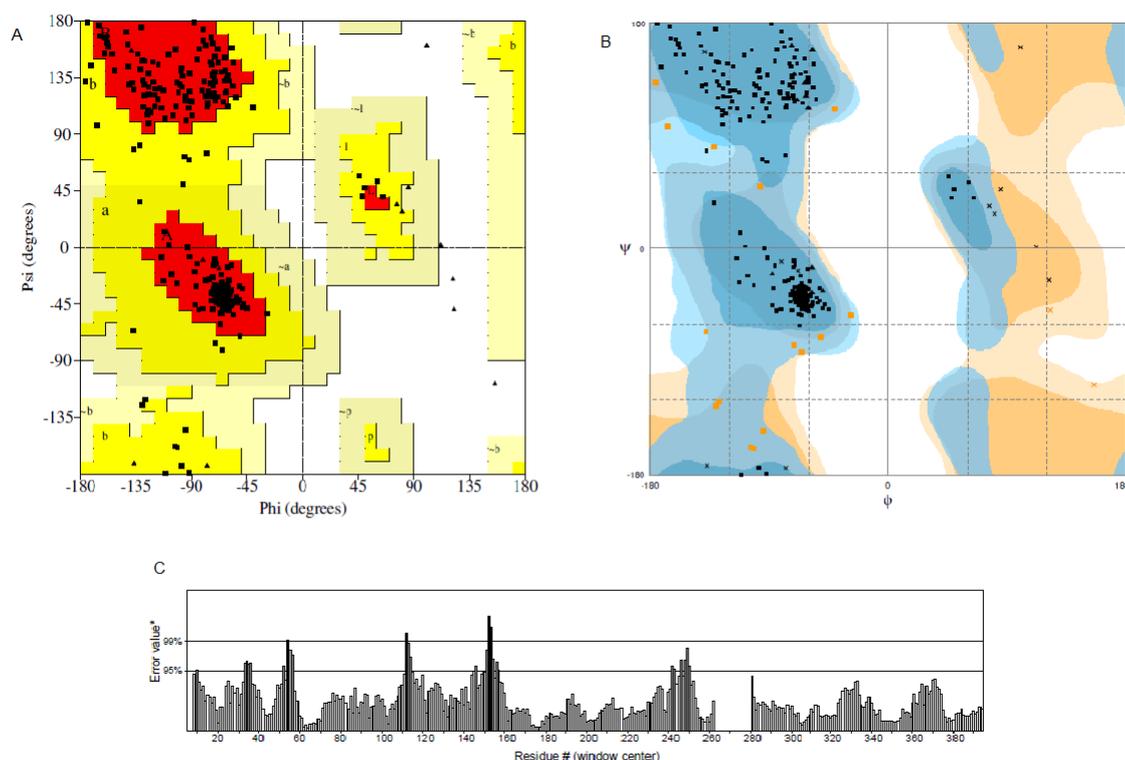
O PROCHECK e o Rampage analisam os ângulos phi-psi para todo resíduo de aminoácido da proteína segundo o Gráfico de Ramachandran, e classificam os modelos de acordo com a porcentagem de resíduos de aminoácidos nas regiões mais favoráveis. Para o Procheck, um modelo de boa qualidade apresenta mais de 90% de seus resíduos nas regiões mais favoráveis, ao passo que o Rampage utiliza um valor aproximado de 98% dos resíduos em regiões favoráveis (LASKOWSKI et al., 1993; LOVELL et al., 2003). O modelo gerado possui 91,9% dos resíduos nas regiões mais favoráveis e 8,1% em regiões permitidas segundo o Procheck, e 95,7% em regiões favoráveis e 4,3% em regiões permitidas segundo o Rampage.

O ERRAT analisa as interações não covalentes entre os átomos nitrogênio (N), carbono (C) e oxigênio/enxofre (O), segundo a nomenclatura do programa, em um total de seis interações (NN, CC, OO, NC, NO, CO), e considera uma estrutura de boa qualidade quando esta possui um score > 50. Este valor é proporcional à resolução da estrutura, de modo que valores em torno de 91% são esperados para estruturas de resolução 2.5-3.0 Å (COLOVOS; YEATES, 1993). O modelo gerado apresentou um score de 93.207. Os resultados do PROCHECK, Rampage e ERRAT estão ilustrados na figura 5.

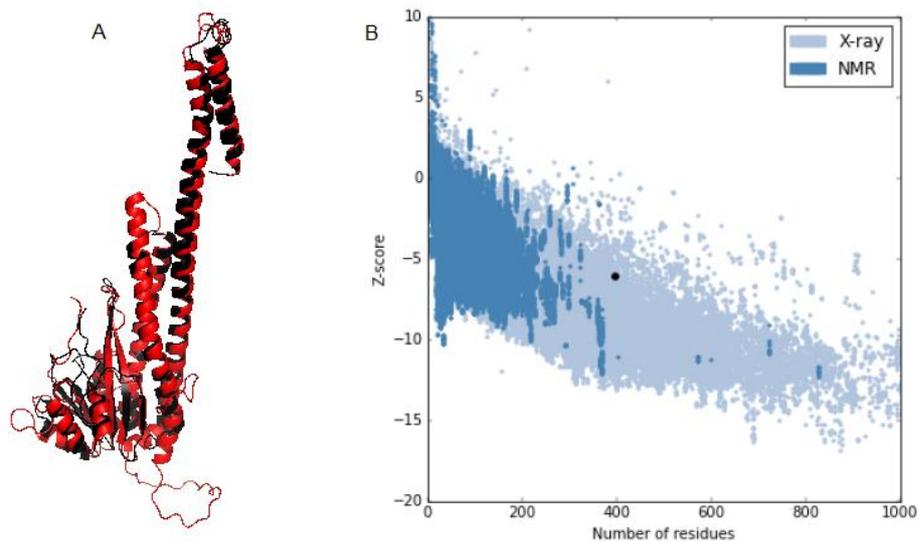
O programa ProSA-web avalia a energia da estrutura e atribui um Z-score, comparando este valor com aqueles obtidos de estruturas experimentais, de acordo com o tamanho da proteína. Se o Z-score estiver fora da amplitude dos valores obtidos de

proteínas experimentais, a estrutura submetida provavelmente contém erros (WIEDERSTEIN; SIPPL, 2007). O modelo gerado apresentou um Z-score = -6.1 Kcal/mol e se encontra dentro da amplitude de valores experimentais.

O TM-align realiza um alinhamento estrutural e avalia a similaridade das estruturas de acordo com os valores de Desvio Médio da Raíz Quadrada - RMSD (do inglês, *Root Mean Square Deviaton*), que mede a distância entre os resíduos correspondentes, e TM-score, que equilibra a distância entre resíduos de uma maneira menos independente do tamanho das proteínas (ZHANG; SKOLNICK, 2005). Valores de RMSD < 2.0 Å e TM-score > 0.5 indicam que as proteínas analisadas apresentam um mesmo enovelamento, o que é esperado para proteínas de uma mesma família. O modelo gerado, quando comparado com o seu molde (PDB ID: 4i0u), apresentou um RMSD = 1.94 Å e um TM-score = 0.88885. Os resultados do ProSA-web e Tm-align estão ilustrados na figura 6.



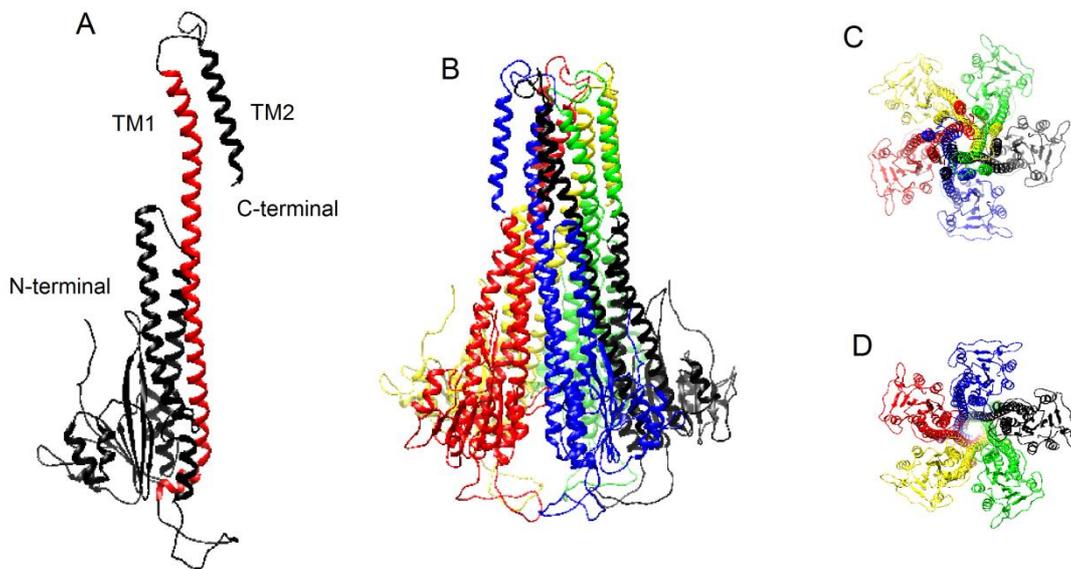
**Figura 5. Validação do modelo in silico da proteína ALR1 segundo o PROCHECK, RAMPAGE e ERRAT. (A)** O gráfico de Ramachandran segundo o PROCHEK apresentou 91,9% dos resíduos em most favoured regions e 8,1% em additional allowed regions. **(B)** O gráfico de Ramachandran segundo o RAMPAGE apresentou 95,7% em favoured region e 4,3% em allowed regions. **(C)** Segundo o ERRAT o modelo obteve um score de 93,207.



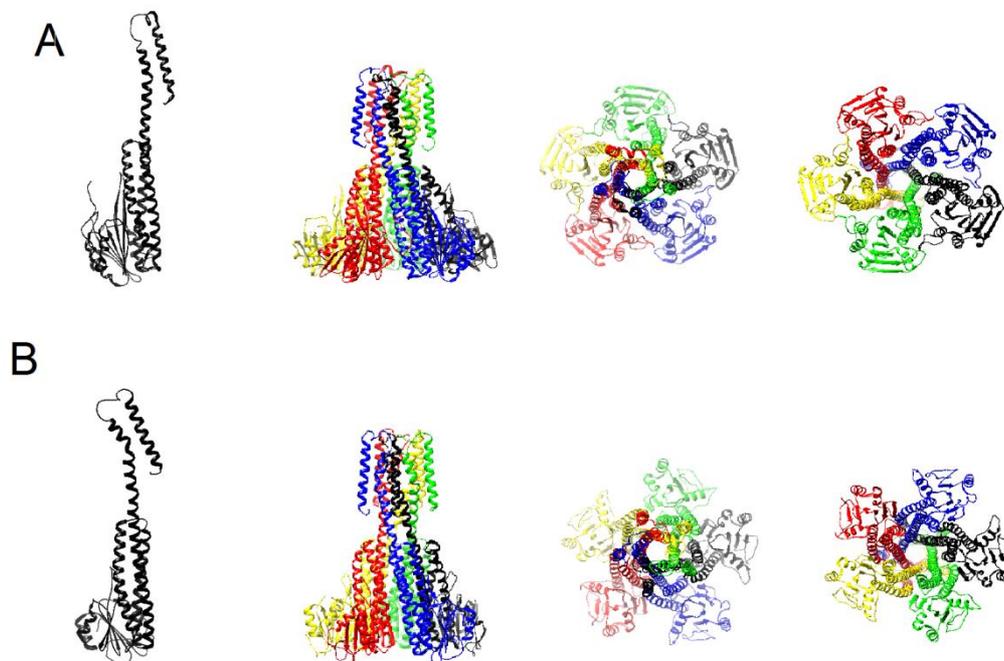
**Figura 6. Validação do modelo in silico da proteínas ALR1 segundo o TM-align e o ProSA-web. (A)** Alinhamento estrutural segundo o TM-align (RMSD = 1,94 Å e um TM-score = 0,88885). Em preto o molde (PDB ID: 4i0uB) e em vermelho o modelo ALR1. **(B)** Validação da estrutura segundo o ProSA-web (Z-score = -6,1) demonstra que o modelo gerado (em preto), apresenta-se dentro da faixa de qualidade do *Protein Data Bank*.

#### 4.4) ANÁLISE DA ESTRUTURA

A estrutura CorA do organismo *Thermotoga maritima* foi a primeira a ser elucidada pela literatura e dessa forma, a maior parte do conhecimento sobre o domínio CorA é proveniente de estudos com a estrutura deste organismo. Posteriormente, até mesmo quando a estrutura MjCorA foi obtida, esta foi estudada comparando-a com a TmCorA, visto que já era conhecido resíduos de aminoácidos essenciais para a função deste domínio, bem como seu mecanismo de ação. Dessa forma, como neste estudo procedeu-se de maneira semelhante, obtendo a estrutura da proteína ALR1 de *S. cerevisiae*, outro homólogo CorA, buscou-se diferenças e semelhanças da ALR1p em relação aos seus homólogos procariotos. Neste sentido, a proteína ALR1, assim como as outras duas proteínas CorA presentes no PDB, também é um homo-pentâmero constituído de cinco monômeros, de acordo com a predição do programa GalaxyGemini e a literatura (PAYANDEH; PFOH; PAI, 2013). As figura 7 e 8 ilustram o modelo ALR1p e seus homólogos procariotos, respectivamente, e a figura 9 apresenta a região N-terminal das proteínas CorA. Os dados da figura 8 foram retirados da literatura.



**Figura 7. Estrutura da proteína ALR1. (A)** Estrutura terciária. Em vermelho é a hélice que forma o poro. Também estão indicados os segmentos transmembrana (TM1 e TM2) e as extremidades N e C-terminal. **(B)** Vista lateral da estrutura quaternária. **(C)** Visão extracelular da estrutura quaternária. **(D)** Visão intracelular da estrutura quaternária. Em (B), (C) e (D), os monômeros estão representados em diferentes cores.

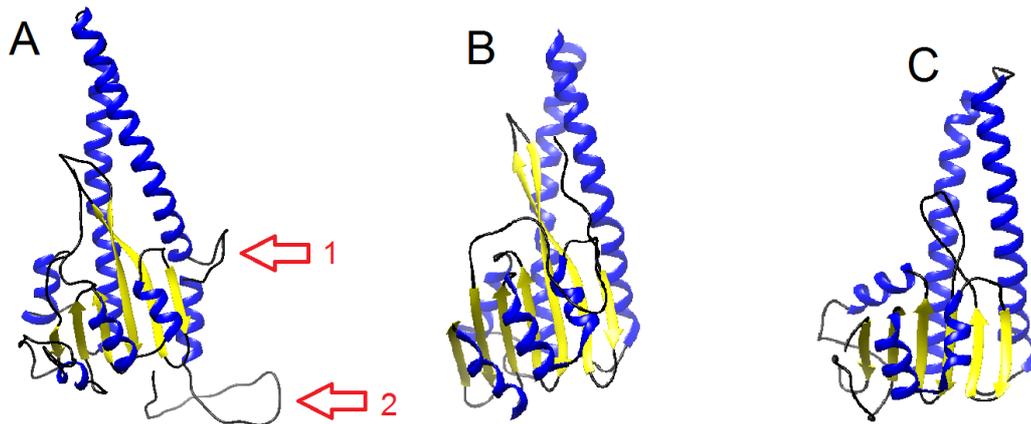


**Figura 8. Estrutura dos homólogos TmCorA (NORDIN et al., 2013) e MjCorA (GUSKOV et al., 2012). (A)** *Thermotoga marítima* (PDB ID: 4i0u). Estrutura terciária e quaternária (vista lateral, extracelular e intracelular).

**(B)** *Methanocaldococcus jannaschii* (PDB ID: 4ev6). Estrutura terciária e quaternária (vista lateral, extracelular e intracelular). Em (A) e (B) as cadeias do homo-pentâmero estão representadas por diferentes cores.

De acordo com as figuras 7 e 8, é possível observar que a ALR1p mantém o dobramento característico da superfamília CorA, observado por seus homólogos procariotos. Nota-se que a proteína ALR1, assim como as demais proteínas CorA, possui uma forma de funil, com uma região N-terminal extensa e larga e uma curta e estreita porção C-terminal, sendo ambas as extremidades localizadas no citoplasma. O poro é composto, na formação da estrutura quaternária, pela maior alfa hélice representada pelos monômeros. Esta hélice, além de formar o poro, liga as extremidades N e C-terminal da proteína e abriga, na altura da membrana plasmática, o primeiro domínio transmembrana (TM1). Vale lembrar que a superfamília CorA é caracterizada pela presença de dois segmentos transmembrana (TM1 e TM2), sendo a tríade GMN (Glicina/Metionina/Asparagina) localizada ao final da TM1.

A região citoplasmática ALR1p possui um conjunto de folhas beta e alfa hélices semelhante aos seus homólogos procariotos, visto que a proteína ALR1 possui um total de nove hélices e seis folhas beta, contra oito hélices e sete folhas beta do organismo *Thermotoga maritima* (TmCorA), e sete hélices e cinco folhas beta do organismo *Methanocaldococcus jannaschii* (MjCorA). De fato, a literatura aponta para a extremidade N-terminal das proteínas CorA como uma região de menor conservação se comparada à extremidade C-terminal. Além da quantidade de estruturas secundárias presentes na porção citoplasmática das proteínas CorA, outro aspecto relevante segundo a literatura, é a maneira conformacional de como estas estruturas secundárias se localizam no espaço. Desta maneira, a literatura descreve este conjunto de estrutura secundária como um sanduíche de folhas beta realizado por dois conjuntos de alfa-hélices. A figura 9 ilustra a região citoplasmática das proteínas CorA, onde é possível observar que este tipo de descrição encontra-se conservado entre a ALR1p e seus homólogos.



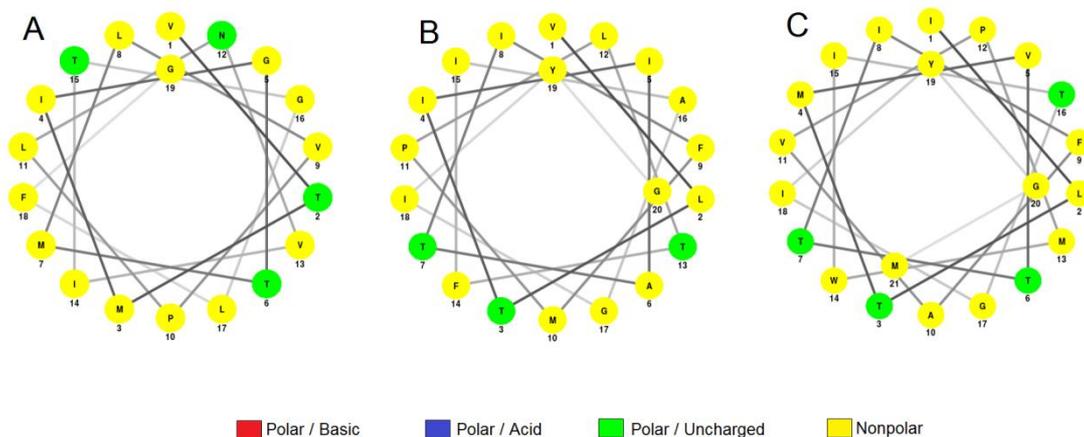
**Figura 9. Extremidade N-terminal dos monômeros da CorA de acordo com a estrutura secundária.** Em amarelo as folhas beta, em azul as hélices e em preto as alças. **(A)** ALR1p. As setas indicam as diferenças significativas na estrutura, quando comparada aos seus homólogos CorA. **(B)** *Thermotoga maritima* (NORDIN et al., 2013). **(C)** *Methanocaldococcus jannaschii* (GUSKOV et al., 2012).

Levando em conta todas as hélices presentes em cada estrutura, com exceção da hélice que forma o poro e da TM2, ausentes na figura 9, todas as demais hélices estão presentes na porção citoplasmática da proteína. A presença de uma hélice a mais do que seu homólogo TmCorA ocorre pois uma das hélices da proteína ALR1 é dividida em duas devido a presença de uma alça, como indicado pela seta 1. Esta alça, juntamente com a estrutura desordenada da alça indicada pela seta 2, constituem as duas grandes diferenças estruturais entre a proteína ALR1 e seus homólogos procariotos. Vale ressaltar que a estrutura indicada pela seta 2 é desordenada e dessa não constitui, necessariamente, uma estrutura secundária do tipo alça, podendo obter outras configurações como alfa-hélice ou folha beta durante o mecanismo de ação da proteína ALR1.

#### 4.5) ANÁLISE DO PORO

Segundo a predição da região transmembrana da ALR1p realizada pelo programa PPM, presente no banco de dados OPM, a região TM1 é composta pelos resíduos de aminoácidos 743-761 (VTMIGTMLVPLNVITGLFG) e a região TM2 pelos resíduos 778-798 (GILGVLLLLAVLGWFLASYWI). A tríade GMN característica da superfamília CorA, na proteína ALR1 abrange os resíduos 761-763, estando logo após o TM1. Lembrando que a ALR1p possui 859 resíduos de aminoácidos e o domínio CorA está localizado na porção C-

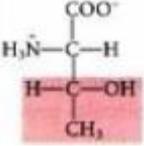
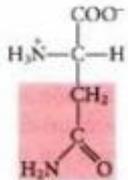
terminal, e a região modelada corresponde aos resíduos 403-800. O TM1 e TM2 da TmCorA abrangem os resíduos 293-312 e 326-345, sendo o GMN localizado nos resíduos 312-314 (LUNIN et al., 2006). Quanto ao homólogo MjCorA, seu TM1 e TM2 abrangem os resíduos 259-279 e 291-311, sendo o GMN localizado nos resíduos 278-280 (GUSKOV et al., 2012). Dessa forma podemos observar que a ALR1p possui a porção C-terminal bastante similar aos seus homólogos, com a tríade GMN localizada logo após o TM1. A figura 10 ilustra a análise do TM1 da ALR1p e de seus homólogos procaríotos, segundo o programa NetWheels.



**Figura 10. Representação bidimensional do TM1 para proteínas CorA. (A) ALR1p. (B) TmCorA (NORDIN et al., 2013).. (C) MjCorA (GUSKOV et al., 2012).**

De acordo com a figura 10, observa-se que não existe resíduo de aminoácido polar no TM1 da ALR1p, de forma semelhante aos seus homólogos. Nota-se, também, que a proteína ALR1, assim como seu homólogo MjCorA, possui quatro resíduos polares não carregados em seu segmento transmembrana TM1, contra três resíduos da TmCorA, porém, somente a ALR1p possui um resíduo de Asparagina (N), ao contrário de todos os resíduos de Treonina (T) existentes. Comparando os aminoácidos polares não carregados presentes neste segmento, de acordo com a tabela 2, conclui-se a Asparagina possui características mais desejáveis, do que a Treonina, para efetuar o transporte de íons, visto que apresenta uma menor hidrofobicidade e um maior potencial para formar pontes de Hidrogênio, além de ter um ponto isoelétrico similar ao da Treonina.

**Tabela 2 - Características dos Aminoácidos Treonina e Asparagina**

	Treonina (T)	Asparagina (N)
Pontes de Hidrogênio	3	5
Ponto Isoelétrico	5,6	5,4
Hidrofobicidade	0,634	0,448
Fórmula		

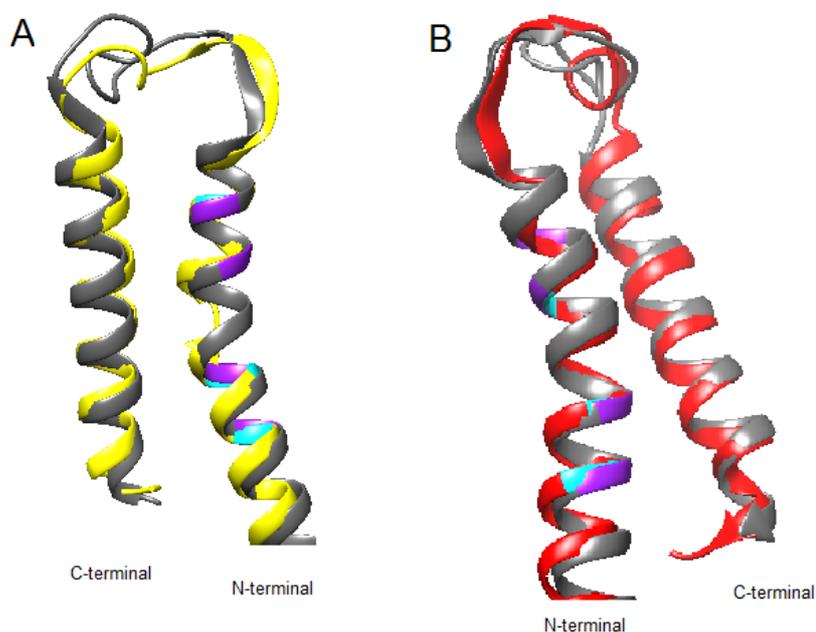
Fonte: NCBI Amino Acid Explorer

([https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa\\_explorer.cgi](https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi))

Ainda, observa-se que todos os resíduos de aminoácidos polares não carregados, presentes na figura 10, encontram-se agrupados em um lado TM1. Essa disposição espacial reflete a porção transmembranar onde o segmento TM1 se encontra, pois sendo a membrana composta por uma bicamada fosfolipídica, espera-se que os resíduos hidrofílicos sejam orientados para o interior do poro e os hidrofóbicos para o exterior, em contato com a membrana. Os resíduos de aminoácidos polares não carregados presente na TmCorA são Treonina (T) 295, T299 e T305, e estão todos envolvidos no transporte de íons segundo a literatura, sendo T305 envolvido, ainda, no processo de reconhecimento de íons na entrada do canal da TmCorA (ESHAGHI, 2006; LUNIN et al., 2006; NORDIN et al., 2013) e na MjCorA, os resíduos presentes na TM1 são T261, T264, T265 e T274, enquanto os três primeiros estão relacionados com o transporte de íons, não existe dados para o resíduo T274 (GUSKOV et al., 2012). Os resíduos polares não carregados apresentados no TM1 da ALR1p são, T744, T748, N754 e T757.

Um alinhamento estrutural pelo Tm-align, ilustrado na figura 11, demonstrou que os resíduos de TmCorA T295, T299, T305 alinham com os resíduos da ALR1p, T744, T748 e N754, e dessa forma, observa-se que somente o resíduo T305 não se encontra conservado no ALR1p. Sabendo que a localização desses resíduos não conservados está envolvida com o reconhecimento de íons na entrada do canal, a proteína ALR1 pode ter uma tendência em caputar uma maior variedade de íons, devido a um maior potencial de pontos de hidrogênio da Asparagina, assim como sua menor hidrofobicidade, em comparação com a Treonina. Não existe correspondente, na TmCorA, do resíduo T757 da ALR1p. Um mesmo alinhamento estrutural, porém com o homólogo MjCorA, revelou que os resíduos de MjCorA,

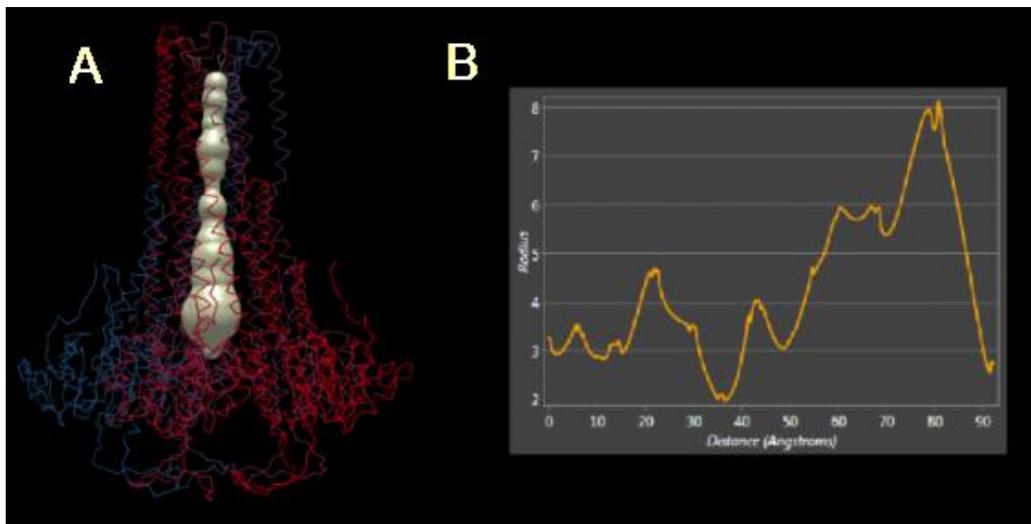
T261, T265 e T274 alinham com os resíduos de ALR1p, T744, T748 e T757. Não existe correspondente, na ALR1p, do resíduo T264 de MjCorA. Embora não existam dados literários que sustentem alguma função para o resíduo T274 de MjCorA, a sua posição similar com o resíduo T305 de TmCorA induz a uma similaridade de função. Assim, T274, similar a T305, pode estar envolvido com o reconhecimento de íons no início do canal, o que implicaria que a proteína ALR1p teria dois resíduos envolvidos nesta função, N754 que ocupa a mesma posição do que T305 de TmCorA, além do T757 que equivale ao T274 de MjCorA.



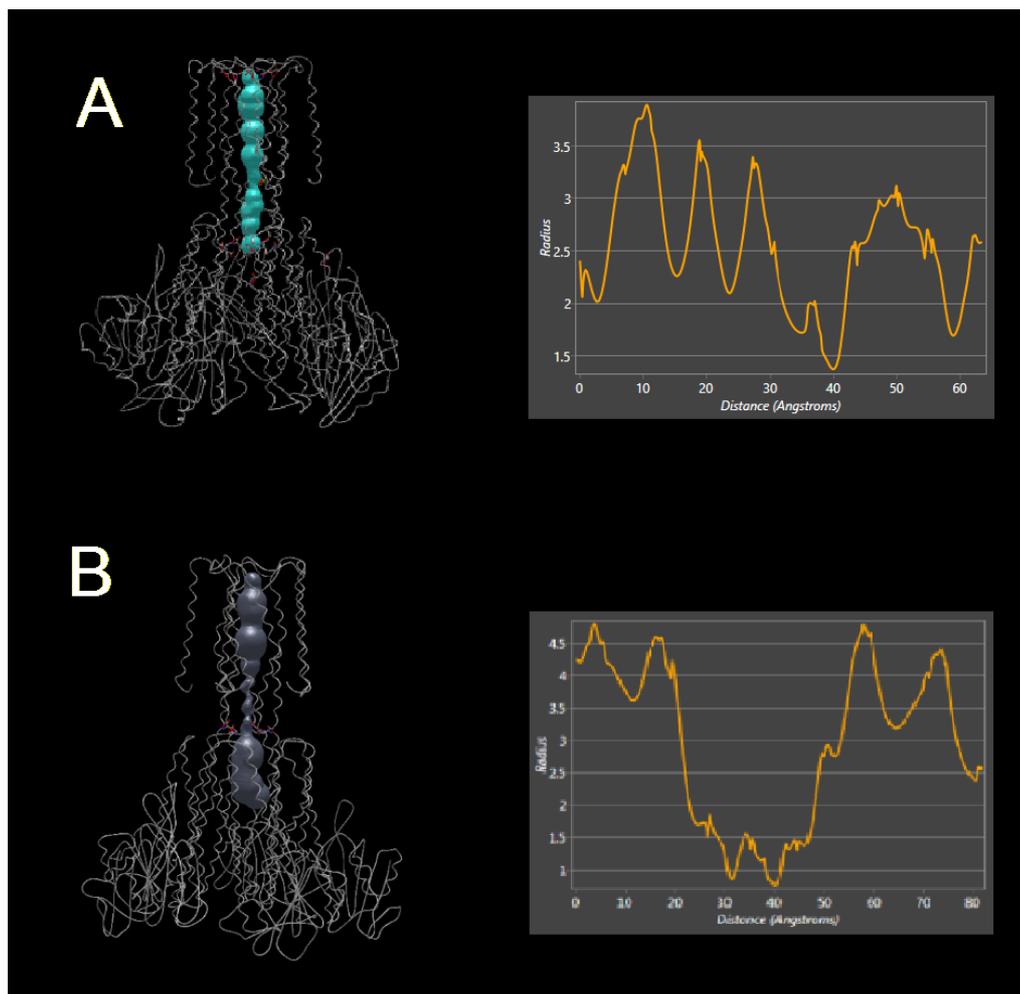
**Figura 11. Alinhamento estrutural entre os resíduos polares não carregados presentes no TM1 das proteínas CorA.** Em azul são os resíduos presentes nos procariotos CorA e em roxo, os resíduos da ALR1p. **(A)** ALR1p (cinza) x MjCorA (amarelo) (GUSKOV et al., 2012). **(B)** ALR1p (cinza) x TmCorA (vermelho) (NORDIN et al., 2013).

A análise das dimensões da ALR1p pelo programa Mole 2.0 revelou que seu poro possui um comprimento de 90 Å, contra 65 Å da TmCorA, ao passo que uma membrana bifosfolipídica possui aproximadamente 30 Å de espessura (PAYANDEH; PAI, 2006). Sendo o resíduo de Asparagina (N), presente na tríade GMN e conservado na superfamília CorA, considerado a entrada do poro, o canal de TmCorA abrange os resíduos N314-277 (PAYANDEH; PAI, 2006) e o da ALR1p 763-641. Para calcular a amplitude do canal da ALR1p foi considerada como saída do poro a sua última constricção, constituída pelos resíduos N239, a qual é formada pela estrutura secundária do tipo alça representada pela

seta 1 na figura 9. A figura 12 e 13 ilustram as dimensões do poro para a proteína ALR1 e seus homólogos, respectivamente.



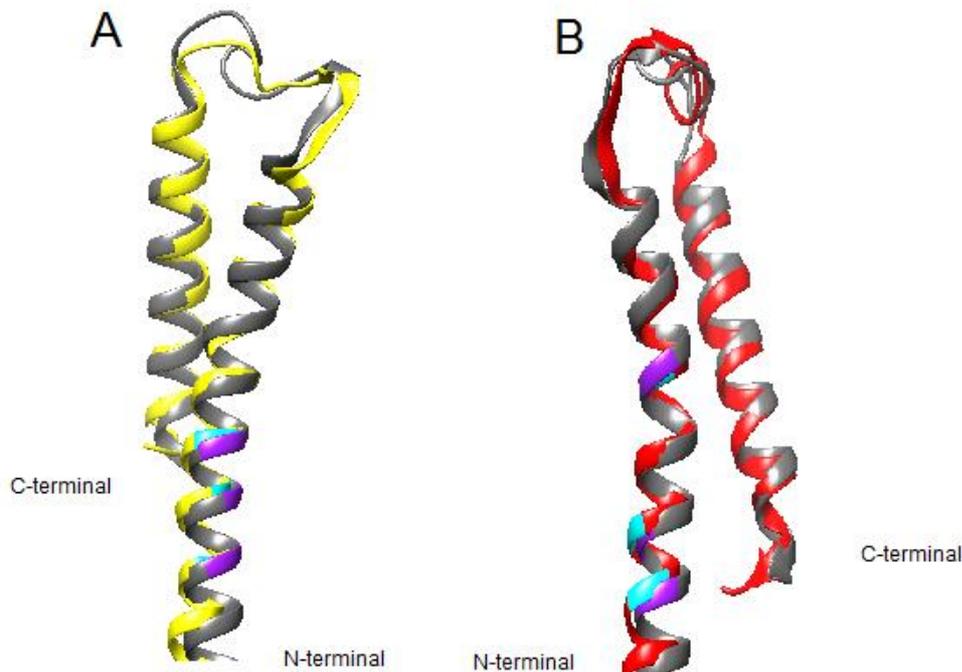
**Figura 12. Poro da proteína ALR1 no sentido C-terminal para N-terminal. (A)** Visualização do Poro. As cadeias da proteína estão representadas em cores diferentes. **(B)** Dimensão do poro da proteína ALR1.



**Figura 13. Poro dos homólogos CorA da proteína ALR1. (A)** TmCorA (NORDIN et al., 2013). Sentido do gráfico: C-terminal para N-terminal **(B)** MjCorA (GUSKOV et al., 2012). Sentido do gráfico: N-terminal para C-terminal.

Observa-se que por volta de 30-50 Å de distância todas as proteínas apresentam um mínimo global para o raio do canal, cujo valor tende a aumentar após esse intervalo e volta a diminuir somente na última constrição do canal, onde as proteínas TmCorA e Alr1p apresentam um valor do raio do poro em torno de 2 Å. Não se sabe em que resíduo de aminoácido terminaria o canal da MjCorA ( GUSKOV et al., 2012; CLEVERLEY et al., 2015), logo, a saída do canal apresentada pelo gráfico é hipotética, para esta proteína em específico. Porém, no que diz respeito ao mínimo global, existem valores de raio específicos por proteína, visto que no caso da TmCorA, este valor é de aproximadamente 1,35 Å, da MjCorA é de 0,77 Å, e da ALR1p, de 1,9 Å. O íon Mg é transportado pelos procaríotos CorA em seu estado parcialmente hidratado, o qual possui um raio de 2,09 Å (NORDIN et al., 2013), porém uma região hidrofóbica no interior de um canal não precisa estar fisicamente fechada para impedir a passagem de íons. Neste sentido, um poro hidrofóbico, com raio de até 4,5 Å, se encontra completamente fechado a íons. Tal poro se abre parcialmente caso o raio aumente para um valor aproximado de 5,5 Å, e se abrirá completamente se o raio for maior do que 7 Å (BECKSTEIN et al., 2003).

Segundo o programa Mole 2.0, os resíduos que formam o poro da ALR1p são: Glicina (G) 758, Asparagina (N) 754, Valina (V) 751, Valina 743, Leucina (L) 740, Valina 736, Asparagina 733, Serina (S) 729, Glutamina (Q) 726, Leucina 722, Asparagina 642. Observando estes resíduos, nota-se a presença de três perfis ao longo do poro da ALR1p. A primeira porção do canal é formado pelos resíduos G758 e N754 e confere um caráter hidrofílico ao canal, seguido por uma porção hidrofóbica, formada pelos resíduos V751, V743, L740 e V736. No terceiro e último perfil o canal volta a ser hidrofílico, sendo formado pelos resíduos N733, S729, Q726, L722 e N642. A região onde o poro apresenta o valor de mínimo global corresponde à porção hidrofóbica da proteína. Resultado semelhante ocorre com seus homólogos, visto que segundo a literatura, o fechamento do canal da TmCorA e MjCorA ocorre por meio de três resíduos hidrofóbicos. No primeiro organismo, estes resíduos são Metionina (M) 302, L294 e L291 (NORDIN et al., 2013), e no segundo, L260, M257 e M253 (GUSKOV et al., 2012). Um alinhamento estrutural, segundo a figura 14 entre a ALR1p e e seus homólogos demonstra que os resíduos que compõe a região de mínimo global estão alinhados entre si.



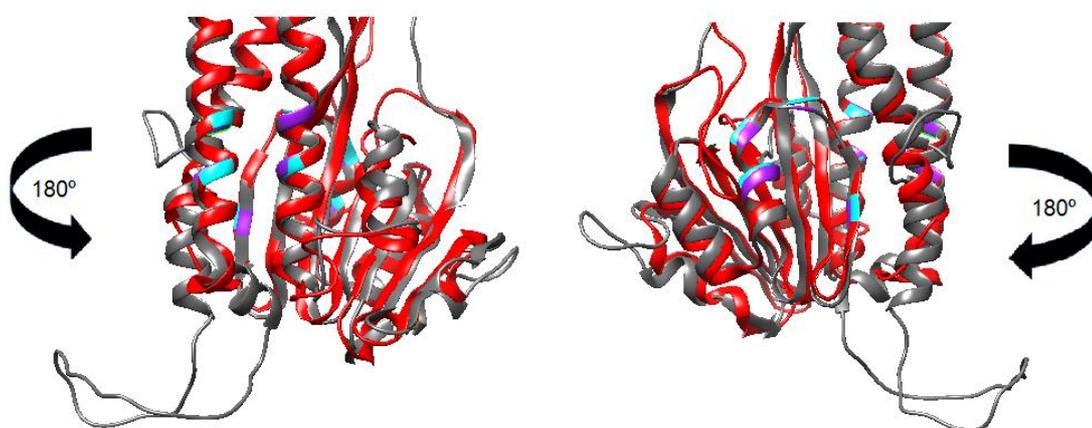
**Figura 14. Alinhamento estrutural entre os resíduos hidrofóbicos.** Em azul são os resíduos hidrofóbicos dos homólogos procariotos, e em roxo, os resíduos da ALR1p. **(A)** ALR1p (cinza) x MjCorA (amarelo) (GUSKOV et al., 2012). **(B)** ALR1p (cinza) x TmCorA (vermelho) (NORDIN et al., 2013).

Segundo o alinhamento estrutural dos resíduos hidrofóbicos, observa-se que os resíduos da ALR1p V743, L740 e V736 alinham com os resíduos L260, M257 e M253 da MjCorA, enquanto que os resíduos da TmCorA M302, L294 e L291 alinham com os resíduos da ALR1p, V751, V743 e L740. Embora seus homólogos CorA possuam apenas três resíduos envolvidos no fechamento do canal, a ALR1p deve possuir quatro resíduos com esta função, visto que M253 da MjCorA e M302 de TmCorA não se alinham entre si, porém os dois alinham com diferentes resíduos hidrofóbicos da ALR1p, V736 e V751, respectivamente. A Valina é um aminoácido mais hidrofóbico do que a Metionina, além desta última possuir um átomo de Enxofre em sua estrutura. A polaridade do átomo de Enxofre da Metionina permite a ligação de íons metálicos e facilita seu transporte (YU, 2011), além da mutação de uma Treonina por Metionina, entre as Treoninas que atuam no transporte de íons da TmCorA, não alterar o transporte de íons (NORDIN et al., 2013).

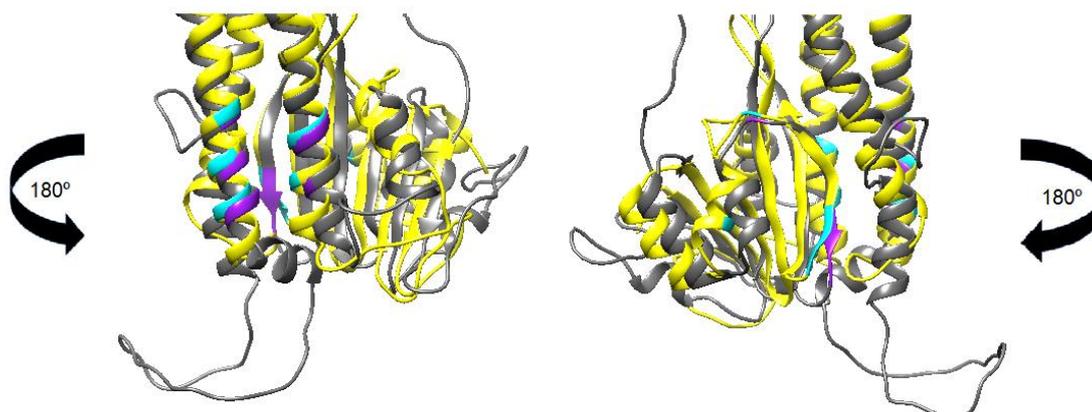
#### 4.6) ANÁLISE DA REGIÃO CITOPLASMÁTICA

Consoante a literatura, existem dois sítios de ligação ao metal nos homólogos procariotos TmCorA e MjCorA, na região citoplasmática, cujas composições não são

conservadas para um único resíduo específico, mas sim para resíduos de propriedades químicas semelhantes (PAYANDEH; PAI, 2006; GUSKOV et al., 2012). Para TmCorA, estes sítios são formados pelos resíduos Ácido Glutâmico (E) 88, Ácido Aspártico (D) 89, N92, Glutamina (Q) 95, E100, D175, D179, D253 e Histidina (H) 257 (ESHAGHI, 2006; PAYANDEH; PAI, 2006). Para MjCorA, D54, Q62, E68, E69, D70, Tirosina (Y) 137, N141, E142, E215, N216, D219 e D223 (GUSKOV et al., 2012). Por meio de um alinhamento estrutural entre os homólogos CorA e a ALR1p é possível visualizar quais resíduos de aminoácidos da proteína ALR1 são equivalentes ou estão próximos, sendo resíduos polares carregados ou não, dos resíduos apresentados pelos homólogos.



**Figura 15. Alinhamento estrutural entre ALR1 e TmCorA para sítio de ligação a metal.** ALR1p está em cinza e seus resíduos destacados em roxo. TmCorA (NORDIN et al., 2013) está em vermelho e seus resíduos destacados em azul.

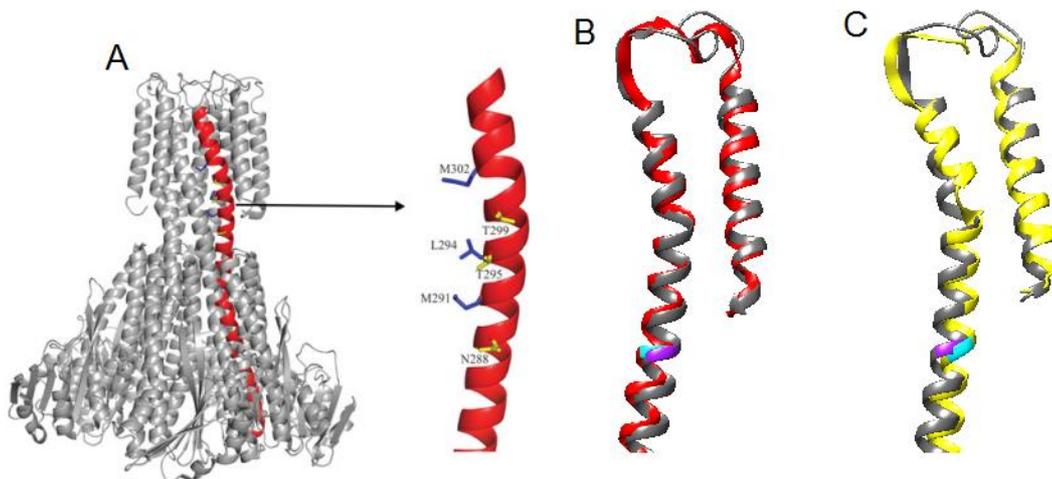


**Figura 16. Alinhamento estrutural entre ALR1 e MjCorA para sítio de ligação a metal.** ALR1p está em cinza e seus resíduos destacados em roxo. MjCorA (GUSKOV et al., 2012) está em amarelo e seus resíduos destacados em azul.

A partir da análise das figuras 15 e 16 é possível observar potenciais resíduos para um sítio de ligação ao metal na proteína ALR1, além de identificar, na região citoplasmática das proteínas CorA, somente dois sítios de ligação a metal, facilmente visualizados por regiões demarcadas em azul e roxo nas figuras. Quando comparado contra o TmCorA, a ALR1p apresenta os seguintes resíduos equivalentes: E482, D483, M486, T489, E494, D569, D573, T702, N706. Desses, somente a Metionina 486 é considerada um resíduo hidrofóbico, porém seu átomo de Enxofre lhe confere um pequeno perfil polar, tanto que, em um estudo indicado na literatura, a mutação de um resíduo hidrofílico por um hidrofóbico, T299M, não causou uma inibição significativa e os mutantes TmCorA demonstraram uma atividade de transporte similar ao modelo selvagem (NORDIN et al., 2013). Se comparada ao MjCorA, os resíduos equivalentes são: T480, E488, E494, L495, Fenilalanina (F) 496, Y565, D569, D570, D698, H699, T702 e N706. A análise da ALR1p contra a MjCorA resultou em dois resíduos hidrofóbicos, L495 e F496, e não são considerados sítios de ligação ao metal. Dessa forma, ao todo a ALR1p possui um total de 15 possíveis sítios para ligação ao metal: T480, E482, D483, M486, E488, T489, E494, Y565, D569, D570, D573, D698, H699, T702 e N706. A TmCorA e MjCorA possuem um total de aproximadamente 9 e 12 sítios de ligação a metal segundo a literatura.

#### 4.7) ANÁLISE DE CONSERVAÇÃO

O mecanismo de ação das proteínas CorA homólogas da ALR1p consiste em dois grupos de resíduos de aminoácidos, hidrofóbicos e hidrofílicos, que se revezam no interior do canal segundo a disponibilidade de íons Mg no ambiente intracelular. Em altas concentrações, os sítios de ligação ao metal, presentes na região citoplasmática da proteína CorA, se tornam ocupados pela ligação ao Magnésio, ao passo que em baixas concentrações, estes íons não estão presentes para se ligarem aos sítios de ligação ao metal. Como estes sítios são adjacentes à hélice que forma o poro, os íons Mg quando se ligam ou se liberam destes sítios, causam uma torção nesta hélice em torno do seu próprio eixo, expondo resíduos hidrofóbicos e hidrofílicos presentes no TM1 de maneira alternada, abrindo e fechando o canal. A figura 17 ilustra este mecanismo.

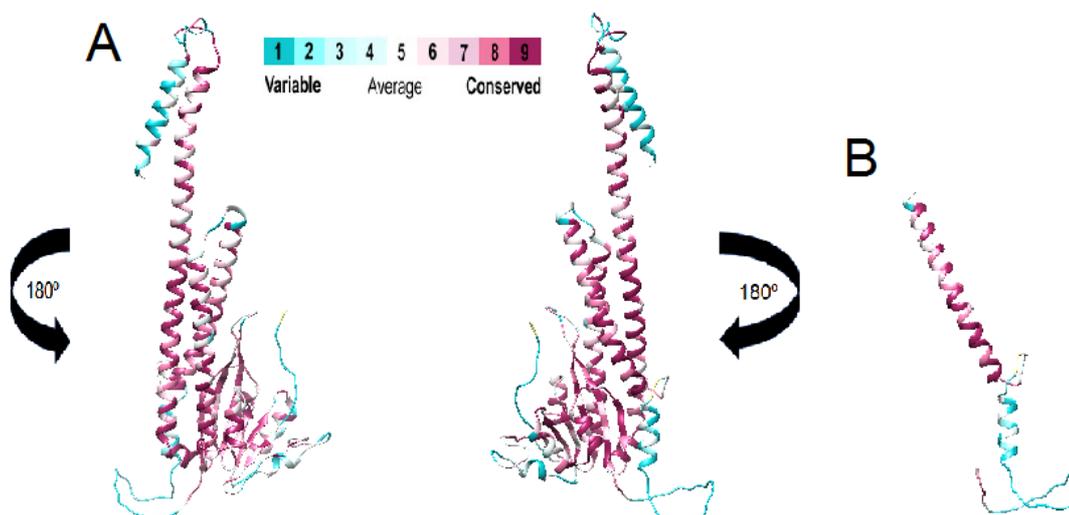


**Figura 17. Mecanismo de ação da TmCorA e alinhamento estrutural do resíduo N288. (A)** Resíduos envolvidos no mecanismo de ação da TmCorA. **(B)** ALR1p (cinza) x TmCorA (vermelho) (NORDIN et al., 2013). **(C)** ALR1p (cinza) x MjCorA (amarelo) (GUSKOV et al., 2012). Em (B) e (C) o resíduo da ALR1p está em roxo e o do seu homólogo em azul.

De acordo com a figura 17, para o homólogo TmCorA, os resíduos de aminoácidos hidrofóbicos envolvidos no mecanismo de ação são, M302, L294 e M291, e os hidrofílicos são T299, T295, e N288. No organismo *M. jannaschii*, os resíduos hidrofóbicos são M253, M257 e L260, e os hidrofílicos, N254, T261 e T265. Ao longo deste trabalho, a partir da análise do poro da proteína ALR1p por meio do estudo do segmento TM1 e das dimensões do poro, verificou-se a similaridade dos resíduos hidrofílicos e hidrofóbicos, respectivamente. Dessa maneira, entre os resíduos envolvidos no mecanismo de ação das proteínas CorA de procaríoto, salvo o resíduo N288, todos os demais já foram analisados. Novamente, por meio de um alinhamento de estrutura, ilustrado na figura 15, observa-se que o resíduo N288 de TmCorA alinha com T737 da ALR1p e com N254 da MjCorA, embora exista um resíduo de Asparagina (N), direcionado para o canal, logo abaixo do T737, o N733, como demonstrado pelo programa Mole 2.0.

Segundo as análises anteriores, os homólogos procariotos apresentam em seu TM1 somente resíduos de Treonina, em um total de quatro para a MjCorA e três para a TmCorA, ao passo que a proteína ALR1 apresenta um resíduo de Arginina e três de Treonina, T744, T748, N754 e T757. Quanto aos resíduos hidrofóbicos envolvidos na dimensão do poro e no mecanismo de ação, enquanto os homólogos procariotos apresentam um total de três, M302, L294 e L291 para a TmCorA, e L260, M257 e M253 para o MjCorA, a ALR1p parece ter quatro resíduos envolvidos nessa função, V751, V743, L740 e V73, além da preferência por resíduos de Valina, ao invés de Metionina ou Leucina.

Porém, a ALR1p é evolutivamente muito distante das homólogas em *Methanocaldococcus jannaschii* e da *Thermotoga marítima*, seus homólogos procariotos, e as diferenças encontradas podem ser benéficas para a proteína ALR1, em um processo de evolução. Dessa forma, sabendo que resíduos de aminoácidos que são estruturalmente ou funcionalmente importantes para as proteínas se encontram conservados, uma análise da proteína ALR1, com homólogos mais próximos, pode ajudar a confirmar a importância de alguns resíduos específicos. A análise filogenética segundo o ConSurf mede o grau de conservação dos resíduos de aminoácido segundo uma escala que varia de 1-9, onde 1 é o mais variável possível e 9 o mais conservado. A figura 18 ilustra a análise filogenética para toda a proteína ALR1p. e a tabela 3 apresenta os resultados desta análise de maneira mais específica, segundo os resíduos importantes apontados pelos homólogos CorA.



**Figura 18. Análise filogenética para a ALR1p. (A)** Proteína inteira. **(B)** Parte estrutural da proteína ALR1p, que a diferencia dos homólogos CorA.

Por meio da figura 18 observa-se que na região citoplasmática, de um total de sete alfa-hélices e seis folhas beta, quatro hélices não são conservadas, porém as duas regiões de ligação a metal permanecem conservadas. A ALR1p possui uma hélice a mais do que seu homólogo TmCorA, além de uma região desordenada, ambas representadas na figura 18A, e com maior destaque na 18B, onde percebe-se uma região de estrutura secundária do tipo alça hélice alça. Este região é a área de maior diferença estrutural entre a ALR1p e seus homólogos CorA, já que todas as proteína CorA possuem um conjunto de hélices e folha beta em seu domínio citoplasmático. A primeira alça desta região forma a saída do poro segundo o programa Mole 2.0 e a segunda é desordenada e de função desconhecida. A

análise da conservação estrutural do domínio transmembrana demonstra que somente o TM2 é completamente variável e o TM1 mistura resíduos de conservação média a elevada, porém nenhum variável.

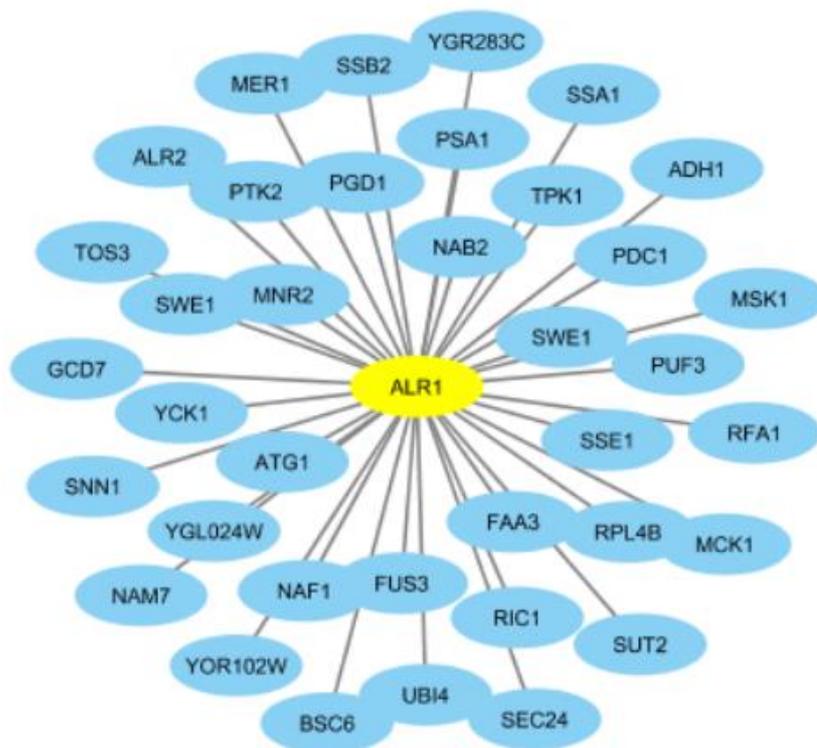
A respeito da análise de conservação funcional da ALR1p, todos os resíduos equivalentes àqueles importantes e essenciais para a função dos seus homólogos encontram-se com elevado grau de conservação, e dessa forma, o mecanismo de ação das proteínas CorA deve estar conservado. A tabela 3 apresenta os resultados da análise de conservação filogenética de maneira mais específica, segundo os resíduos importantes apontados pelos homólogos CorA.

**Tabela 3 - Grau de Conservação dos Aminoácidos Importantes da ALR1p**

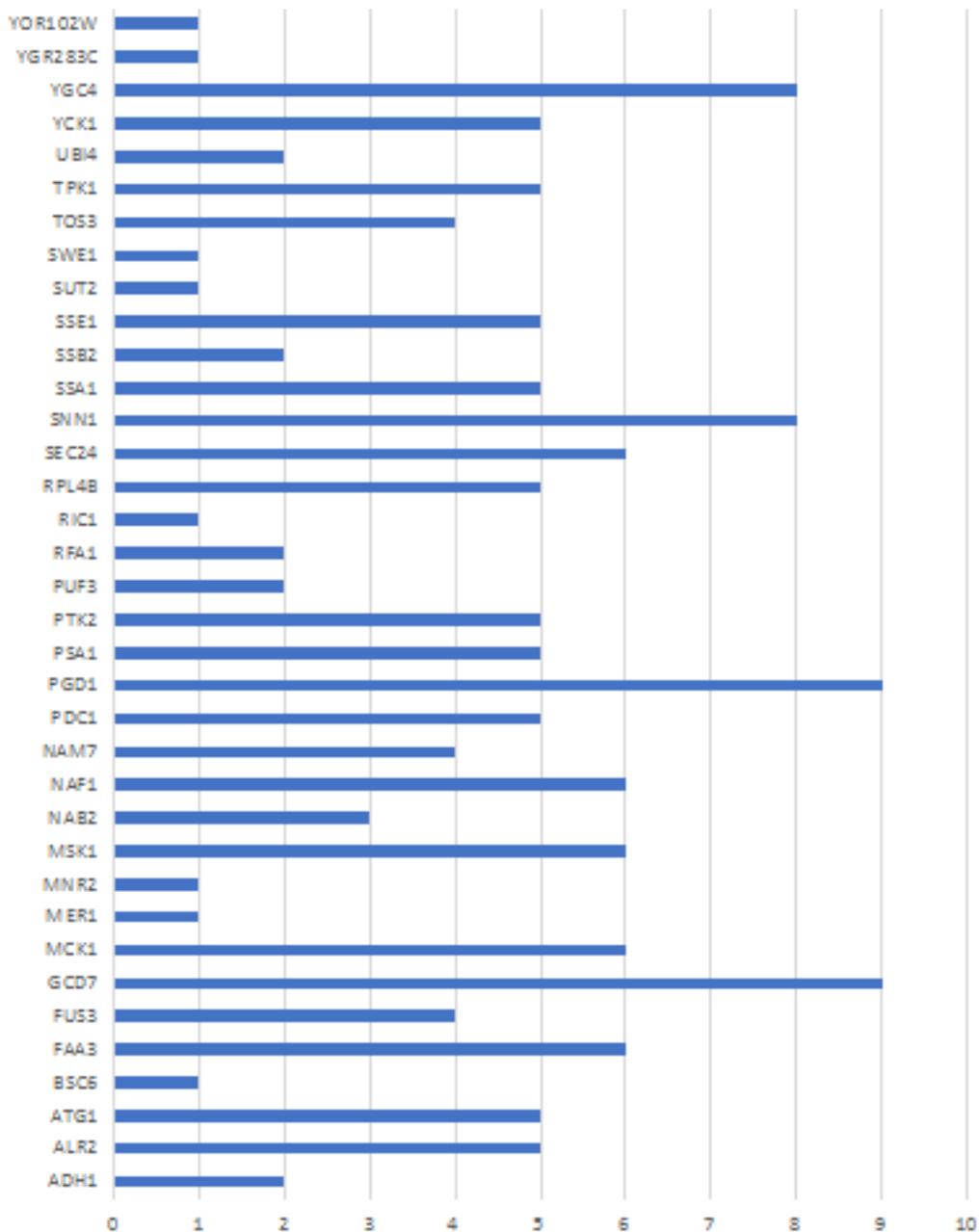
Resíduo	Grau	Função
T737	8	Resíduo hidrofílico na hélice do poro
T744	9	
T748	9	
N754	9	
T757	8	
V751	7	Resíduo hidrofóbico na hélice do poro
V743	8	
L740	9	
V736	7	
T480	9	Sítio de ligação a metal
E482	9	
D483	9	
M486	8	
E488	9	
T489	6	
E494	9	
Y565	9	
D569	9	
D570	8	
D573	9	
D698	9	
H699	9	
T702	9	
N706	8	

#### 4.8) ANÁLISE DO INTERATOMA FÍSICO

A Biologia de Sistemas estuda a interação entre as diversas proteínas e genes encontrados em uma célula, dentro do contexto de uma rede de interação ampla e dinâmica, sendo a construção desta rede essencial para facilitar a compreensão dos complexos sistemas biológicos (COSTANZO et al., 2010). Neste sentido, existem dois tipos de interação, a proteica e a genética. As interações proteicas reúnem a maquinaria molecular e sustentam todas as respostas celulares, enquanto que as interações genéticas revelam as relações funcionais entre os genes (STARK et al., 2006). Dentro deste contexto, obteve-se, a partir de 10 bancos de dados, o interatoma da proteína ALR1 acerca das interações proteicas e observou-se um total de 37 interações. As figuras 19 e 20 ilustram o resultado do interatoma da proteína ALR1 de uma maneira geral e específica de acordo com a frequência das proteínas nos bancos de dados, respectivamente.



**Figura 19. Interatoma acerca das interações proteicas da ALR1p.**



**Figura 20. Proteínas do interatoma da ALR1p segundo a frequência nos bancos de dados de interação.**

De acordo com a figura 20 observa-se que não existe um bando de dados central que archive todas as interações proteína-proteína publicadas, visto que os bancos de dados capturaram apenas uma fração dos dados publicados, e portanto, suas informações são complementares e podem ser unificadas para aumentar e melhorar o conhecimento sobre o interatoma. De fato, observa-se que a interseção e a sobreposição entre esses bancos de dados não é grande, pois nenhuma proteína foi predita pelos 10 bancos de dados ao mesmo

tempo, e, de um total de 37 proteínas, somente duas foram preditas por 9 bancos de dados e outras duas por oito bancos, sendo a grande maioria das proteínas concentrada na faixa de 3-6 bancos de dados, o que representa uma baixa amplitude para um total de 10 bancos de dados utilizados. De maneira contrária, as oito proteínas representadas por apenas um banco de dados precisam ser ainda validadas por experimentos futuros, visto que, associada a esta única predição, as interações proteicas podem ser determinadas por um conjunto diversificado de técnicas experimentais, incluindo técnicas de pouca acurácia que possuem uma alta taxa de falsos positivos (KERRIEN et al., 2012).

Ainda, de um total de 37 proteínas analisadas, somente duas foram encontradas na literatura associadas com a resposta ao Cádmio. Dagmar e colaboradores, 2014, analisaram o transcriptoma do organismo *S. cerevisiae* em resposta a exposição aguda a 10 metais, Ag<sup>+</sup>, Al<sup>3+</sup>, As<sup>3+</sup>, Cd<sup>2+</sup>, Co<sup>2+</sup>, Hg<sup>2+</sup>, Mn<sup>2+</sup>, Ni<sup>2+</sup>, V<sup>3+</sup>, and Zn<sup>2+</sup>, estabelecendo uma diferença mínima de 50% para o grupo controle não tratado com o objetivo de definir uma mudança na expressão gênica. Foram encontrados 172 genes que tiveram seu perfil de expressão gênica alterado, sendo os genes SSA1 e SSE1 também encontrados no interactoma da proteína ALR1, preditos por cinco bancos de dados. SSA1 e SSE1 são chaperonas induzidas pela intoxicação pelo Cádmio e auxiliam o correto dobramento das proteínas, inibem e resgatam os agregados proteicos, e degradam proteínas anormais, ocasionando uma resposta ao Cádmio, visto que a degradação de aberrações proteicas contribui para a tolerância e sobrevivência celular (HOSINER et al., 2014).

## 5) CONCLUSÃO

A proteína ALR1 é conhecida na literatura por transportar uma variedade de metais, sendo estes metais tóxicos ou essenciais, e embora a atividade de efluxo de Cádmio já tenha sido comprovada, sua função primária é permitir o influxo de íons Magnésio. De acordo com este trabalho conclui-se que a ALR1p é de fato um homólogo distante dos organismos *Methanocaldococcus jannaschii* e *Thermotoga marítima*, visto que manteve a estrutura característica da superfamília CorA (2 TMs + GMS), além de possuir resíduos de aminoácidos conservados na mesma posição de resíduos considerados essenciais para seus homólogos, e assim, o mecanismo de ação deve estar preservado, ainda que existam pequenas alterações na estrutura, quando comparada a seus homólogos procaríotos. Não

se sabe o motivo pelo qual a ALR1p aparenta ter uma menor seletividade para captação de íons, porém, foi estabelecido que esta proteína apresenta dois resíduos conservados na entrada do poro cujas funções, em seus homólogos, é a captação de íons. Os homólogos apresentam somente um único resíduo nesta posição, e um raio do canal com um valor menor se comparados à proteína ALR1. O interatoma desta proteína demonstrou somente duas outras proteínas, SSA1 e SSE1, em conformidade com uma intoxicação aguda pelo Cádmiu, embora nenhuma dessas duas proteínas sejam específicas para uma resposta ao Cd. Dessa forma, sendo a ALR1p uma proteína canal de baixa seletividade, segundo a literatura, associada ao fato de já ter sido relacionada com o efluxo de Cd, e somando os resultados encontrados neste estudo, acredita-se que a ALR1p deva ser capaz de mediar o efluxo de Cádmiu, porém estudos posteriores são necessários para comprovar seu mecanismo de ação, visto que assim como seus homólogos, a proteína ALR1 se encontra em seu estado conformacional fechado.

## 6) REFERÊNCIAS BIBLIOGRÁFICAS

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.

ALTSCHUL, S. F. et al. Gapped BLAST and PS I-BLAST: a new generation of protein database search programs. **Nucleic Acids Res**, v. 25, n. 17, p. 3389–3402, 1997.

ASHKENAZY, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic acids research**, p. 1–7, 2016.

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

BATEMAN, A. et al. UniProt: A hub for protein information. **Nucleic Acids Research**, v. 43, n. D1, p. D204–D212, 2015.

BECKSTEIN, O. et al. Ion channel gating: Insights via molecular simulations. **FEBS Letters**, v. 555, n. 1, p. 85–90, 2003.

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.

BREIEROVÁ, E. et al. Biosorption of cadmium ions by different yeast species. **Zeitschrift fur Naturforschung - Section C Journal of Biosciences**, v. 57, n. 7–8, p. 634–639, 2002.

BUCHAN, D. W. A et al. Scalable web services for the PSIPRED Protein Analysis Workbench. **Nucleic acids research**, v. 41, n. Web Server issue, p. 349–357, 2013.

CALDERONE, A.; CASTAGNOLI, L.; CESARENI, G. Mentha: a Resource for Browsing Integrated Protein-Interaction Networks. **Nature methods**, v. 10, n. 8, p. 690, 2013.

CALIXTO, P. H. M. ASPECTOS GERAIS SOBRE A MODELAGEM COMPARATIVA DE PROTEÍNAS. **Ciência Equatorial**, v. 3, 2013.

CARNEIRO, D. D. A.; GARIGLIO, L. P. A biorremediação como ferramenta para a descontaminação de ambientes terrestres e aquáticos. **Tecer**, v. 3, p. 82–95, 2010.

CHANG, J. M.; DI TOMMASO, P.; NOTREDAME, C. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. **Molecular Biology and Evolution**, v. 31, n. 6, p. 1625–1637, 2014.

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.

CLEVERLEY, R. M. et al. The Cryo-EM structure of the CorA channel from *Methanocaldococcus jannaschii* in low magnesium conditions. **Biochimica et Biophysica Acta - Biomembranes**, v. 1848, n. 10, p. 2206–2215, 2015.

COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Sci**, v. 2, n. 9, p. 1511–1519, 1993.

COSTANZO, M. et al. The genetic landscape of a cell. **Science (New York, N.Y.)**, v. 327, n. 5964, p. 425–31, 2010.

COUTINHO, P. W. R. et al. Alternative Soil Remediation: Bioremediation and Phytoremediation. **Nucleus**, v. 12, n. 1, p. 59–68, 2015.

COWLEY, M. J. et al. PINA v2.0: Mining interactome modules. **Nucleic Acids Research**, v. 40, n. D1, p. 862–865, 2012.

DFG, *Deutsche Forschungsgemeinschaft*. Cadmium and its compounds (in the form of inhale dusts/aerosols). **The MAK collection for occupational health and safety**, vol 22. D. Forschungsgemeinschaft, Wiley-VCH, Weinheim, 2006.

DROZDETSKIY, A. et al. JPred4: A protein secondary structure prediction server. **Nucleic Acids Research**, v. 43, n. W1, p. W389–W394, 2015.

DUFFUS, J. H. “HEAVY METALS”—A MEANINGLESS TERM? (IUPAC Technical Report) (IUPAC Technical Report). v. 74, n. 5, p. 793–807, 2002.

DUNKER, A. K. et al. Intrinsically disordered protein. **Journal of Molecular Graphics and Modelling**, v. 19, n. 1, p. 26–59, 2001.

DUNKER, A. K. et al. Intrinsic Disorder and Protein Function. **Proteins**, v. 41, n. 21, p. 6573–

6582, 2002.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.

ESHAGHI, S. Crystal Structure of a Divalent Metal Ion Transporter CorA at 2.9 Angstrom Resolution. **Science**, v. 313, n. 5785, p. 354–357, 2006.

ESWAR, N. et al. **Comparative protein structure modeling using Modeller**. [s.l: s.n.]. v. Chapter 5.

FENG, Z. P. et al. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. **Molecular and Biochemical Parasitology**, v. 150, n. 2, p. 256–267, 2006.

FEREIDOUNI, M.; DANESHI, A.; YOUNESI, H. Biosorption equilibria of binary Cd(II) and Ni(II) systems onto *Saccharomyces cerevisiae* and *Ralstonia eutropha* cells: Application of response surface methodology. **Journal of Hazardous Materials**, v. 168, n. 2–3, p. 1437–1448, 2009.

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.

FONTENELLE, M. Aspectos Da Política Nacional Do Meio Ambiente : O Estudo De Impacto Ambiental Como Instrumento. p. 271–302, 2004.

FRISHMAN, D.; ARGOS, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. **Protein Engineering**, v. 9, n. 2, p. 133–142, 1996.

GAYLARDE, C. C.; BELLINASSO, M. D. L.; MANFIO, G. P. Biorremediação: Aspectos Biológicos e Técnicos da Biorremediação de Xenobióticos. **Biotecnologia Ciência e Desenvolvimento**, p. 36–43, 2005.

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.

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

GUENTHER, B. et al. Crystal structure of the  $\beta$ -subunit of the clamp-loader complex of *E. coli* DNA polymerase III. **Cell**, v. 91, n. 3, p. 335–345, 1997.

GUSKOV, A. et al. Structural insights into the mechanisms of Mg<sup>2+</sup> uptake, transport, and gating by CorA. **Proceedings of the National Academy of Sciences**, v. 109, n. 45, p. 18459–18464, 2012.

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

HEO, L.; PARK, H.; SEOK, C. GalaxyRefine: Protein structure refinement driven by side-

- chain repacking. **Nucleic acids research**, v. 41, n. Web Server issue, p. 384–388, 2013.
- HOSINER, D. et al. Impact of acute metal stress in *Saccharomyces cerevisiae*. **PLoS ONE**, v. 9, n. 1, p. 1–14, 2014.
- HU, J. et al. Renal cell carcinoma and occupational exposure to chemicals in Canada. **Occupational Medicine**, v. 52, n. 3, p. 157–164, 2002.
- IARC, International Agency for Research on Cancer. **Monograph summary, Volume 100C**, 2012.
- 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.
- 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.
- KERRIEN, S. et al. The IntAct molecular interaction database in 2012. **Nucleic Acids Research**, v. 40, n. D1, p. 841–846, 2012.
- KING, R. D.; STERNBERG, M. J. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. **Protein science : a publication of the Protein Society**, v. 5, n. 11, p. 2298–2310, 1996.
- KOTLYAR, M. et al. Integrated interactions database: Tissue-specific view of the human and model organism interactomes. **Nucleic Acids Research**, v. 44, n. D1, p. D536–D541, 2016.
- LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. November, p. 283–291, 1993.
- LEE, H. et al. GalaxyGemini: A web server for protein homo-oligomer structure prediction based on similarity. **Bioinformatics**, v. 29, n. 8, p. 1078–1080, 2013.
- LEE, J. M.; GARDNER, R. C. Residues of the yeast ALR1 protein that are critical for magnesium uptake. **Current Genetics**, v. 49, n. 1, p. 7–20, 2006.
- LEI, Y.; CHEN, W.; MULCHANDANI, A. Microbial biosensors. **Analytica Chimica Acta**, v. 568, n. 1–2, p. 200–210, 2006.
- LI, C. et al. Effect of NaCl on the heavy metal tolerance and bioaccumulation of *Zygosaccharomyces rouxii* and *Saccharomyces cerevisiae*. **Bioresource Technology**, v. 143, p. 46–52, 2013.
- 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.

LICATA, L. et al. MINT, the molecular interaction database: 2012 Update. **Nucleic Acids Research**, v. 40, n. D1, p. 857–861, 2012.

LIM, P. H. et al. Regulation of *alr1* mg transporter activity by intracellular magnesium. **PLoS ONE**, v. 6, n. 6, 2011.

LINDING, R. et al. Protein disorder prediction: Implications for structural proteomics. **Structure**, v. 11, n. 11, p. 1453–1459, 2003.

LIU, G. J. et al. Large Mg<sup>2+</sup>-dependent currents are associated with the increased expression of ALR1 in *Saccharomyces cerevisiae*. **FEMS Microbiology Letters**, v. 213, n. 2, p. 231–237, 2002.

LOMIZE, A. L. et al. Positioning of proteins in membranes: A computational approach. **Protein Science**, v. 15, n. 6, p. 1318–1333, 2006a.

LOMIZE, M. A. et al. OPM: Orientations of proteins in membranes database. **Bioinformatics**, v. 22, n. 5, p. 623–625, 2006b.

LOVELL, S. C. et al. Structure validation by C $\alpha$  geometry: phi,psi and C $\beta$  deviation. **Proteins**, v. 50, n. 3, p. 437–450, 2003.

LUNIN, V. V. et al. Crystal structure of the CorA Mg<sup>2+</sup> transporter. **Nature**, v. 440, n. 7085, p. 833–837, 2006.

MARTÍ-RENO, M. A. et al. Comparative protein structure modeling of genes and genomes. **Annual review of biophysics and biomolecular structure**, v. 29, p. 291–325, 2000.

MEILER, J.; BAKER, D. Coupled prediction of protein secondary and tertiary structure. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 21, p. 12105–12110, 2003.

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

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.

Mól, A. R, Castro, M, S. Fontes, W. NetWheels: A web application to create high quality peptide helical wheel and net projections, 2016.

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

NORDIN, N. et al. Exploring the structure and function of *Thermotoga maritima* CorA reveals the mechanism of gating and ion selectivity in Co<sup>2+</sup>/Mg<sup>2+</sup> transport. **Biochemical Journal**, v. 451, n. Pt 3, p. 365–374, 2013.

NTP, National Toxicology Program. **Tenth Report on Carcinogens, Department of Health and Human Services, III-42–III-44, Triangle Park, NC, 2000.**

OATES, M. E. et al. D2P2: Database of disordered protein predictions. **Nucleic Acids Research**, v. 41, n. D1, p. 508–516, 2013.

OBER, Joyce A. **Mineral commodity summaries 2016.** US Geological Survey, 2016.

OROSZ, F.; OVÁDI, J. Proteins without 3D structure: Definition, detection and beyond. **Bioinformatics**, v. 27, n. 11, p. 1449–1454, 2011.

PALOPOLI, L. et al. Improving protein secondary structure predictions by prediction fusion. **Information Fusion**, v. 10, n. 3, p. 217–232, 2009.

PAYANDEH, J.; PAI, E. F. A structural basis for Mg<sup>2+</sup> homeostasis and the CorA translocation cycle. **The EMBO journal**, v. 25, n. 16, p. 3762–73, 2006.

PAYANDEH, J.; PFOH, R.; PAI, E. F. The structure and regulation of magnesium selective ion channels. **Biochimica et Biophysica Acta (BBA) - Biomembranes**, v. 1828, n. 11, p. 2778–2792, 2013.

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.

PETERSEN, B. et al. A generic method for assignment of reliability scores applied to solvent accessibility predictions. **BMC Structural Biology**, v. 9, n. 1, p. 51, 2009.

PISAT, N. P.; PANDEY, A.; MACDIARMID, C. W. MNR2 regulates intracellular magnesium storage in *Saccharomyces cerevisiae*. **Genetics**, v. 183, n. 3, p. 873–884, 2009.

POLLASTRI, G. et al. **Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles** *Proteins: Structure, Function, and Bioinformatics*, 2002.

POTENZA, E. et al. MobiDB 2.0: An improved database of intrinsically disordered and mobile proteins. **Nucleic Acids Research**, v. 43, n. D1, p. D315–D320, 2015.

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.

PRIETO, C.; DE LAS RIVAS, J. APID: Agile protein interaction DataAnalyzer. **Nucleic Acids Research**, v. 34, n. WEB. SERV. ISS., p. 298–302, 2006.

ROST, B.; LIU, J. The PredictProtein server. **Nucleic Acids Research**, v. 31, n. 13, p. 3300–3304, 2003.

SÁNCHEZ, R.; ŠALI, A. **Advances in comparative protein-structure modelling** *Current*

**Opinion in Structural Biology**, 1997.

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.

SEHNAL, D. MOLE 2 . 0 : advanced approach for analysis of biomacromolecular channels. p. 1–13, 2013.

SEN, T. Z. et al. GOR V server for protein secondary structure prediction. **Bioinformatics**, v. 21, n. 11, p. 2787–2788, 2005.

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. v. 28, n. 1, p. 45–54, 2013.

SHANNON, P. et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 2003.

SHIRAISHI, 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.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular systems biology**, v. 7, n. 1, p. 539, 2011.

SILVA, R. R. R. et al. O Processo De Biorremediação Na Recuperação De Solos Contaminados. p. 1–5, 2008.

STARK, C. et al. BioGRID: a general repository for interaction datasets. **Nucleic acids research**, v. 34, n. Database issue, p. D535-9, 2006.

SZKLARCZYK, D. et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. **Nucleic Acids Research**, v. 43, n. D1, p. D447–D452, 2015.

VERLI, H. **Bioinformática da Biologia à Flexibilidade Molecular**. [s.l: s.n.]. v. 53

WARDE-FARLEY, D. et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. **Nucleic Acids Research**, v. 38, n. SUPPL. 2, p. 214–220, 2010.

WEBER, Bruna; SANTOS, Antenor. Utilização da Biorremediação como Ferramenta para o controle da Degradação Ambiental Causada pelo Petróleo e seus Derivados. **Engenharia Ambiental: Pesquisa e Tecnologia**, v. 10, n. 1, 2013.

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.

XENARIOS, I. et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. **Nucleic acids research**, v. 30, n. 1, p. 303–305, 2002.

YU, E. Crystal Structures of the CusA Efflux Pump Suggest Methionine Mediated Metal Transport. **Biophysical Journal**, v. 100, n. 3, p. 382a, 2011.

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.