

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-graduação em Bioinformática

Julia Teixeira Rodrigues

**AVALIAÇÃO DOS PRODUTOS DE EXPRESSÃO DAS REGIÕES GÊNICAS
ENVOLVIDAS NA DEGRADAÇÃO DO ÁCIDO ÚRICO**

Belo Horizonte

2024

Julia Teixeira Rodrigues

**AVALIAÇÃO DOS PRODUTOS DE EXPRESSÃO DAS REGIÕES GÊNICAS
ENVOLVIDAS NA DEGRADAÇÃO DO ÁCIDO ÚRICO**

Dissertação de Mestrado apresentada
ao Programa de Pós-Graduação
Interunidades em Bioinformática da
Universidade Federal de Minas Gerais
como requisito parcial para a obtenção
do título de Mestre em Bioinformática.

Orientadora: Prof^ª. Dr^ª. Mariana
Torquato Quezado de Magalhães

Co-orientadora: Prof^ª. Dr^ª. Glória
Regina Franco

Belo Horizonte

2024

043

Rodrigues, Julia Teixeira.

Avaliação dos produtos de expressão das regiões gênicas envolvidas na degradação do ácido úrico [manuscrito] / Julia Teixeira Rodrigues. – 2024.

121 f. : il. ; 29,5 cm.

Orientador: Mariana Torquato Quezado de Magalhães. Coorientador: Glória Regina Franco.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Ácido Úrico. 3. Hiperuricemia. 4. Urato Oxidase. 5. Expressão Gênica. I. Magalhães, Mariana Torquato Quezado de. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS

ATA

INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

Julia Teixeira Rodrigues

Às oito horas e trinta minutos do dia **19 de setembro de 2024**, reuniu-se, por videoconferência através do aplicativo zoom, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**AVALIAÇÃO DOS PRODUTOS DE EXPRESSÃO DAS REGIÕES GÊNICAS ENVOLVIDAS NA DEGRADAÇÃO DO ÁCIDO ÚRICO**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, a Presidente da Comissão, **Dra. Mariana Torquato Quezado de Magalhaes**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professora/Pesquisadora	Instituição	Indicação
Dra. Mariana Torquato Quezado de Magalhaes - Orientadora	Universidade Federal de Minas Gerais	Aprovada
Dra. Glória Regina Franco	Universidade Federal de Minas Gerais	Aprovada
Dr. Anderson de Sá Pinheiro	Universidade Federal do Rio de Janeiro	Aprovada
Dra. Flavia Figueira Aburjaile	Universidade Federal de Minas Gerais	Aprovada
Dra. Rafaela Salgado Ferreira	Universidade Federal de Minas Gerais	Aprovada

Pelas indicações, a candidata foi considerada: **Aprovada**

O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.



Documento assinado eletronicamente por **Rafaela Salgado Ferreira, Chefe de departamento**, em 19/09/2024, às 13:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gloria Regina Franco, Professora do Magistério Superior**, em 19/09/2024, às 14:18, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Flávia Figueira Aburjaile, Professora do Magistério Superior**, em 19/09/2024, às 14:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 19/09/2024, às 14:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Anderson de Sá Pinheiro, Usuário Externo**, em 19/09/2024, às 17:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3564224** e o código CRC **3763A018**.

AGRADECIMENTOS

À minha família, em especial meus pais, Adriana e Klever, minha irmã, Carol, e meus companheiros, Dudu e o Magneto, por todo suporte e apoio verbais e não-verbais. À minha madrinha, Shirley Simone, e aos meus tios Heberth e Daniela, que nunca falharam em fornecer os conselhos que eu precisava no exato momento que eram requeridos.

Às minhas pessoas favoritas do LBM, Camila, Flávio, Ana e Amanda, que sempre estiveram lá para auxiliar não só nos experimentos, mas em todos momentos de dificuldade, preenchendo meus dias no laboratório com olhares, risadas e, quando necessário, sermões. Aos alunos de graduação mais eficientes do ICB, Gabriel, Aninha e Pedrinho, sempre interessados e dispostos a ajudar.

Aos meus amigos do PPG, cujos laboratórios visito com uma frequência alarmante, agradeço todas conversas sobre ciência, assim como a troca de informações de interesse público. Em especial, gostaria de agradecer à Duda, que foi praticamente uma irmã para mim, me consolando, apoiando e dizendo duras, e necessárias, verdades. Às minhas amigas da graduação, especialmente a Giovanna, a Marina, a Livia e a Carol, com quem compartilhei muitas histórias durante a trajetória na faculdade de farmácia.

Aos integrantes do LGB, principalmente a prof^a. Glória e seus alunos, Iza, Gabi, Bárbara, Dáfne e Freddy, que me ajudaram a compreender as análises dos dados de transcriptoma, e a Dani, que me auxiliou na realização do western blot.

À todos os professores e funcionários técnicos, especialmente o Lucas, o Maurício, o Jamil, a Ivana e a Adriana, que auxiliaram direta e indiretamente não só no desenvolvimento deste trabalho, mas na construção da minha trajetória na ciência.

À Sheila e ao Tiago, deixo um agradecimento especial, por terem ajudado, de forma transparente e impecável, com as inúmeras burocracias da pós-graduação e, além disso, por todas as dúvidas sanadas, e-mails respondidos em um intervalo de 10 minutos e, principalmente, pelas palavras de incentivo.

À prof. Mariana que atuando como professora, orientadora e amiga nas horas vagas, me guiou ao longo da minha iniciação científica e do meu mestrado. Obrigada por ter mostrado a importância de resultados negativos, ensinado a duvidar dos resultados positivos e, acima de tudo, por abrir as portas para o fantástico mundo das proteínas.

À Universidade Federal de Minas Gerais, às agências de fomento, CAPES, CNPq e Fapemig, por tornarem possível a realização deste trabalho.

RESUMO

Em hominídeos o produto final do catabolismo das purinas é o ácido úrico, ao contrário de outros mamíferos placentários, nos quais, através de uma série de reações catalisadas pelas urato oxidase (uricase), HIU hidrolase (HIUase) e OHCU descarboxilase, o ácido úrico é degradado em (S)-alantoína. Em *Homo sapiens*, estudos sobre a via de degradação do ácido úrico revelaram que mutações no gene da uricase levaram a uma redução na pressão seletiva para a manutenção das regiões gênicas das outras enzimas da via. A falta de expressão dessas enzimas leva, em algumas situações, ao acúmulo de ácido úrico. Apesar da aparente ausência das proteínas em *H. sapiens*, nossas análises mostraram que há a transcrição das regiões gênicas da uricase (UOX), HIUase (URAHP), OHCU descarboxilase (URAD) e alantoicase (ALLC). Enquanto alguns transcritos de UOX e URAHP são considerados não produtivos, os genes URAD e ALLC são produtivos, ou seja, produzem transcritos codificadores de proteínas. Neste estudo foi feita a análise de dados da literatura, assim como de base de dados pública, para levantar informações sobre os produtos (RNA e proteína) e perfis de expressão das regiões UOX, URAHP, URAD e ALLC, seguida de análises de dados de RNA-seq provenientes de estudos caso-controle com pacientes acometidos com afecções associadas à hiperuricemia, para avaliar se a expressão dos transcritos destas regiões gênicas poderia contribuir para a regulação do catabolismo das purinas e a patogênese de doenças associadas ao acúmulo de ácido úrico. Tendo a hipótese inicial do trabalho se mostrado nula e considerando a potencial presença dos produtos proteicos da URAD e da ALLC, partiu-se para avaliação da possibilidade de produção de enzimas ativas. Assim, foram realizados alinhamentos da estrutura primária, que indicaram que o produto proteico da URAD parecia ser mais promissor, sendo, então, realizada a modelagem e avaliação da estrutura terciária e a expressão em sistema heterólogo da OHCU descarboxilase recombinante de *H. sapiens*. Tendo sido obtida a proteína recombinante, foram realizados ensaios enzimáticos, além de experimentos de para a caracterização utilizando espectrometria de massa, dicroísmo circular e ressonância magnética nuclear, tendo o resultado geral apontado para produção de uma enzima ativa e cuja estrutura é condizente com as demais descarboxilases da literatura.

Palavras chave: ácido úrico, hiperuricemia, RNA-seq, OHCU descarboxilase

ABSTRACT

In hominids, the final product of purine catabolism is uric acid, unlike in other placental mammals where uric acid is degraded into (S)-allantoin through a series of reactions catalyzed by urate oxidase (uricase), HIU hydrolase (HIUase), and OHCU decarboxylase. In *Homo sapiens*, studies on the uric acid degradation pathway revealed that mutations in the uricase gene have led to reduced selective pressure for maintaining the gene regions of other enzymes in the pathway. The lack of expression of these enzymes can lead, in some situations, to the accumulation of uric acid. Despite the apparent absence of the proteins in *H. sapiens*, our analyses showed transcription of the gene regions for uricase (UOX), HIUase (URAHP), OHCU decarboxylase (URAD), and allantoicase (ALLC). While some transcripts of UOX and URAHP are considered non-productive, the genes URAD and ALLC are productive, meaning they produce protein-coding transcripts. In this study, literature data and public databases were analyzed to gather information on the products (RNA and protein) and expression profiles of the UOX, URAHP, URAD, and ALLC regions. This was followed by RNA-seq data analysis from case-control studies involving patients with conditions associated with hyperuricemia, to assess whether the expression of transcripts from these gene regions could contribute to the regulation of purine catabolism and the pathogenesis of diseases associated with uric acid accumulation. Since the initial hypothesis of the study was null and considering the potential presence of protein products from URAD and ALLC, evaluation was conducted on the possibility of producing active enzymes. Thus, alignments of the primary structure were performed, indicating that the protein product of URAD seemed more promising. Subsequently, modeling and evaluation of the tertiary structure were carried out, followed by expression in a heterologous system of recombinant *H. sapiens* OHCU decarboxylase. Upon obtaining the recombinant protein, enzymatic assays were performed, alongside characterization experiments using mass spectrometry, circular dichroism, and nuclear magnetic resonance, with the overall result indicating production of an active enzyme whose structure is consistent with other decarboxylases reported in the literature.

Key words: uric acid, hyperuricemia, RNA-seq, OHCU decarboxylase

LISTA DE FIGURAS

Figura 1.	Formação de adenilato e guanilato a partir do PRPP.	13
Figura 2.	Formação do ácido úrico a partir do adenilato e do guanilato.	14
Figura 3.	Via de degradação do ácido úrico	16
Figura 4.	Resultado da duplicação do gene da HIUase ancestral.	17
Figura 5.	Estrutura da alantoína sintase	18
Figura 6.	Rede de interação do URAD e ALLC com genes codificadores de proteínas.	30
Figura 7.	Expressão mediana dos genes do catabolismo do ácido úrico em tecidos humanos.	32
Figura 8.	Perfis de expressão dos genes do catabolismo do ácido úrico em tecidos humanos por sistema.	33
Figura 9.	Perfis de expressão de transcritos da região do gene URAHP em tecidos humanos por sistema	35
Figura 10.	PCA dos casos e controles dos projetos selecionados do SRA.	43
Figura 11.	PCA das variáveis de confusão dos projetos selecionados do SRA.	44
Figura 12.	Expressão diferencial a nível de gene dos projetos selecionados do SRA.	47
Figura 13.	Expressão diferencial a nível de transcrito dos projetos selecionados do SRA.	48
Figura 14.	Expressão diferencial dos genes e transcritos da via de degradação das purinas.	51
Figura 15.	Expressão diferencial dos genes e transcritos envolvidos na excreção de ácido úrico.	52
Figura 16.	Alinhamentos da estrutura primária do produto proteico da URAD com sequências de OHCU descarboxilase.	70
Figura 17.	Rede de correlação para resíduos de duas comunidades da OHCU descarboxilase	73
Figura 18.	Alinhamento da estrutura primária do produto proteico da ALLC com sequência de alantoicase e domínios presentes na sequência.	74
Figura 19.	Estruturas de OHCU descarboxilase depositadas no PDB.	77

Figura 20.	Modelos construídos para o produto proteico da URAD.	80
Figura 21.	Sítio de interação da OHCU descarboxilase de <i>D. rerio</i> com a alantoína.	81
Figura 22.	Digestão do plasmídeo	82
Figura 23.	Expressão da OHCU descarboxilase de <i>H. sapiens</i> utilizando a <i>E. coli</i> BL21(DE3)pLys.	83
Figura 24.	Avaliação da proteína contaminante de <i>E. coli</i> por espectrometria de massa MALDI-TOF.	84
Figura 25.	Perfis de absorção no UV do ácido úrico e da alantoína	86
Figura 26.	Ensaio enzimático.	87
Figura 27.	Expressão da OHCU descarboxilase de <i>H. sapiens</i> utilizando a <i>E. coli</i> ArcticExpress (DE3)	88
Figura 28.	Avaliação da OHCU descarboxilase de <i>H. sapiens</i> por espectrometria de massa MALDI-TOF.	89
Figura 29.	Ensaio enzimático e avaliação da interação com alantoína.	91
Figura 30.	Experimentos de dicroísmo circular com OHCU descarboxilase de <i>H. sapiens</i> .	93
Figura 31.	Avaliação da OHCU descarboxilase de <i>H. sapiens</i> por ressonância magnética nuclear.	94

LISTA DE TABELAS

Tabela 1 .	Regiões gênicas do catabolismo de ácido úrico no genoma humano	21
Tabela 2 .	Transcritos das regiões gênicas do catabolismo de ácido úrico no genoma humano	22
Tabela 3 .	Resultado do DisGeNET para os genes associados a hiperuricemia	40
Tabela 4 .	Projetos do SRA selecionados para a análise de transcriptoma	42
Tabela 5 .	Total de contagens das reads e valores de log ₂ FC e p-valor obtidos na análise da expressão diferencial para UOX, URAHP, URAD e ALLC	49
Tabela 6 .	Resíduos conservados presentes no produto proteico da ALLC	75
Tabela 7 .	Íons teóricos e experimentais obtidos na digestão da OHCU descarboxilase de <i>H. sapiens</i>	89

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 Metabolismo de purinas.....	13
1.2 Degradação do ácido úrico.....	15
1.3 Acúmulo de ácido úrico.....	18
1.4 Regiões gênicas das enzimas envolvidas na degradação do ácido úrico.....	20
2. PRIMEIRA PROPOSTA.....	25
3. MÉTODO.....	25
3.1 Avaliação dos produtos de expressão das regiões UOX, URAHP, URAD e ALLC.....	25
3.2 Avaliação dos perfis de expressão das regiões UOX, URAHP, URAD e ALLC.....	26
4. RESULTADOS E DISCUSSÃO.....	28
4.1 Produtos de expressão das regiões UOX, URAHP, URAD e ALLC.....	28
4.2 GTE _x : Expressão Genótipo-Tecido da UOX, URAHP, URAD e ALLC.....	31
5. SEGUNDA PROPOSTA.....	36
6. MÉTODO.....	37
6.1 DisGeNET: Doenças e Genes Associados a Hiperuricemia.....	37
6.2 RNA-seq: expressão diferencial da UOX, URAHP, URAD e ALLC.....	37
7. RESULTADOS E DISCUSSÃO.....	39
7.1 DisGeNET: Doenças e Genes Associados a Hiperuricemia.....	39
7.2 RNA-Seq: expressão diferencial da UOX, URAHP, URAD e ALLC em patologias associadas a hiperuricemia.....	41
8. TERCEIRA PROPOSTA.....	52
9. MATERIAIS E MÉTODOS.....	53
9.1 Avaliação da estrutura primária dos produtos proteicos da URAD e da ALLC.....	53
9.2 Predição e análise da estrutura terciária do produto proteico da URAD.....	54
9.3 Expressão e purificação do produto proteico da URAD.....	56
9.3.1 Multiplicação do plasmídeo em DH5 α e verificação da presença do inserto..	57
9.3.2 Transformação de E. coli BL21(DE3)pLys e E. coli ArcticExpress (DE3) para expressão da proteína recombinante.....	58
9.3.3 Expressão e purificação da proteína recombinante.....	59
9.3.3.1 E. coli BL21(DE3)pLys.....	59
9.3.3.2 E. coli ArcticExpress (DE3).....	61
9.4 Verificação da expressão da proteína recombinante por meio de Western Blot.....	62
9.5 Verificação da estrutura primária da proteína por meio de Espectrometria de Massa MALDI-TOF-MS/MS.....	63
9.6 Avaliação da atividade descarboxilase utilizando ensaios enzimáticos.....	66
9.7 Caracterização da estrutura secundária e da estabilidade térmica da proteína por meio de Dicroísmo Circular.....	67
9.8 Caracterização do enovelamento proteico com Ressonância Magnética Nuclear....	68
10. RESULTADOS E DISCUSSÃO.....	69
10.1 Presença de resíduos críticos conservados no produto da URAD indicam funções	

catalíticas essenciais e alinhamento da alantoicase de <i>S. cerevisiae</i> e ALLC indicam diferenças evolutivas significativas.....	69
10.2 O produto proteico da URAD possui conservação no enovelamento com outros organismos modelos e conservação espacial dos resíduos críticos para a interação com a alantoína.....	76
10.3 Avaliação da presença do gene URAD no vetor plasmidial.....	82
10.4 OHCU descarboxilase de <i>H. sapiens</i> demonstrou atividade catalítica, realizando a descarboxilação do OHCU para gerar (S)-alantoína.....	83
10.5 Expressão da OHCU descarboxilase de <i>H. sapiens</i> utilizando a <i>E. coli</i> ArcticExpress (DE3) reduz a degradação da enzima.....	87
10.6 OHCU descarboxilase de <i>H. sapiens</i> assume um padrão de estrutura secundária predominantemente em α -hélice.....	92
11. CONCLUSÃO E PERSPECTIVAS.....	96
12. REFERÊNCIAS.....	97
13. ANEXOS.....	104
Anexo I: Organização dos éxons e íntrons dos transcritos da UOX, URAHP, URAD e ALLC.....	104
Anexo II: Dados completos do Ensembl versão 109 - Fevereiro de 2023.....	106
Anexo III: Dados completos do Ensembl versão 111 - Janeiro de 2024.....	107
Anexo IV: Sequências das proteínas avaliadas.....	108
Anexo V: Scripts para modelagem por homologia utilizando o Modeller.....	109
Anexo VI: Alinhamento das cinco estruturas geradas pelo ColabFold e seus valores de pLDDT.....	110
Anexo VII: Alinhamento das cinco estruturas de maior e menor energia geradas pelo Modeller e seus valores de DOPE.....	111
Anexo VIII: Validação da OHCU descarboxilase de <i>D. rerio</i> (PDB 2O70).....	112
Anexo IX: Validação do modelo construído utilizando o ColabFold.....	114
Anexo X: Validação do modelo construído utilizando o Modeller.....	116
Anexo XI: Validação do modelo construído utilizando o iTASSER.....	118
Anexo XII: Espectros dos ensaios enzimáticos individuais.....	120

1. INTRODUÇÃO

1.1 Metabolismo de purinas

Os nucleotídeos púricos, fosfato de guanosina e fosfato de adenosina, são moléculas que apresentam um heterociclo composto por dois anéis, um pirimidínico e outro imidazólico, ligado ao carbono 1' de uma ribose, a qual também apresenta um ou mais grupos fosfato ligados ao carbono 3'. Além de precursores de ácidos nucleicos (DNA e RNA), estes compostos também apresentam função de carreadores de energia química (ATP e GTP), são componentes de cofatores (NAD, FAD e coenzima A) e podem atuar como segundos mensageiros nos processos de sinalização celular (cAMP e cGMP) (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

A biossíntese das purinas pode ocorrer pela via de salvação, na qual as bases nitrogenadas livres (guanina e adenina) e os nucleosídeos (ribose ligada a base nitrogenada) liberados na degradação de ácidos nucleicos são utilizados para formar novos nucleotídeos, ou pela via *de novo*. Neste último caso, a síntese se inicia com o 5-fosforribosil-1-pirofosfato (PRPP), formado a partir da ribose-5-fosfato, e prossegue com a incorporação de átomos de carbono, nitrogênio e oxigênio provenientes, principalmente, de aminoácidos e folato, o que leva, em última instância, à construção do anel púrico, conforme mostrado na **Figura 1** (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

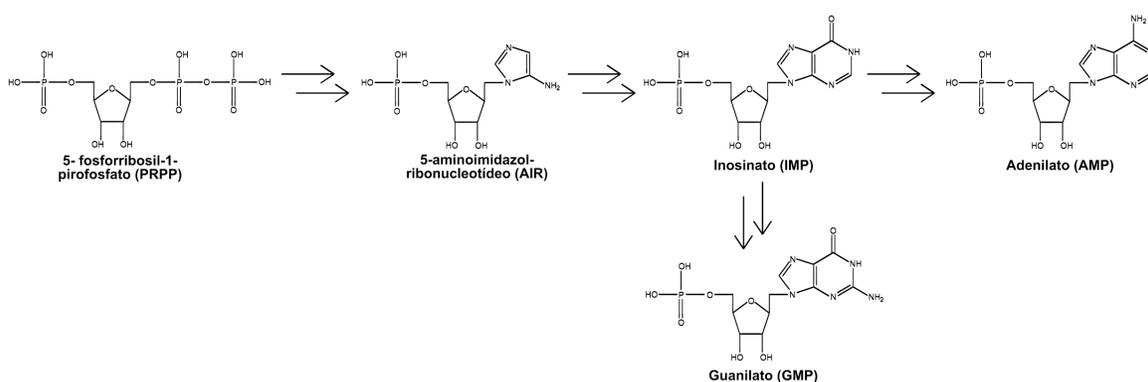


Figura 1. Formação de adenilato e guanilato a partir do PRPP. Esquema simplificado indicando o composto inicial (5-fosforribosil-1-pirofosfato), além de dois intermediários (5-aminoimidazol-ribonucleotídeo e Inosinato), da via de síntese de nucleotídeos púricos.

Estudos indicam que as enzimas que atuam na via *de novo* de síntese de purinas estão presentes nas células como complexos multienzimáticos, que permitem o direcionamento

dos intermediários das reações. Além disso, em alguns organismos também há atuação de enzimas multifuncionais, como a enzima bifuncional de *Homo sapiens* fosforribosil - aminoimidazol carboxilase / Fosforribosil - aminoimidazol succino - carboxamida sintetase (PAICS), um octâmero cujas subunidades apresentam domínios de AIR-carboxilase e SAICAR sintetase (LI *et al.*, 2007).

Já durante o processo de degradação, mostrado na **Figura 2**, os nucleotídeos púricos perdem o grupamento fosfato, o que leva à formação dos nucleosídeos que, em seguida, perdem a ribose. As bases nitrogenadas geradas passam por algumas reações, sendo gerado, então, o ácido úrico (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

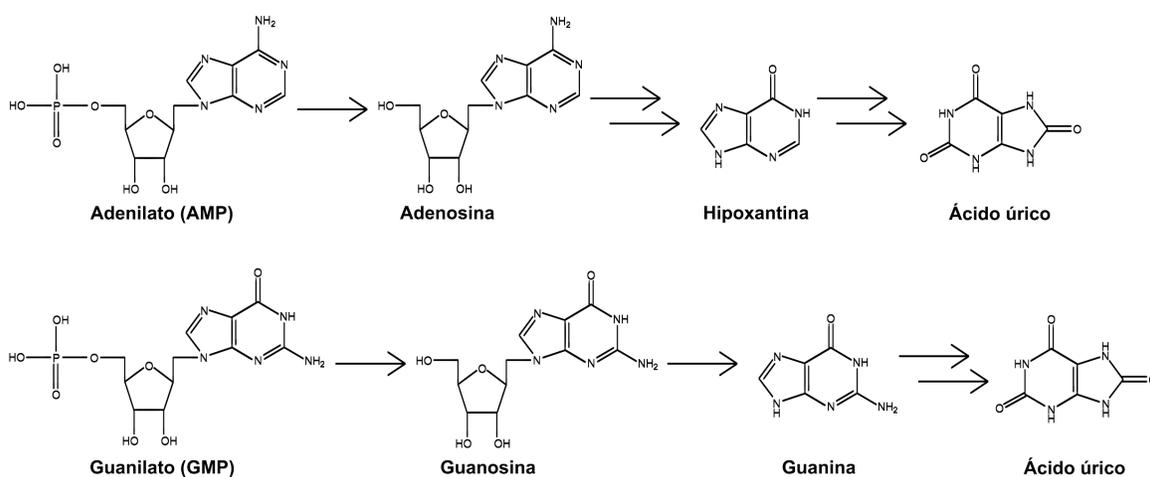


Figura 2. Formação do ácido úrico a partir do adenilato e do guanilato. Esquema simplificado indicando a via de degradação dos nucleotídeos púricos, sendo mostrados os intermediários após perda do grupamento fosfato (adenosina e guanosina) e da pentose (hipoxantina e guanina) e o produto final, o ácido úrico.

Como mencionado anteriormente, as bases nitrogenadas livres e os nucleosídeos liberados na degradação de purinas podem ser utilizados para formar nucleotídeos pela via de salvação. Um exemplo seria a reação catalisada pela hipoxantina guanina fosforribosiltransferase (HPRT), durante a qual ocorre a conversão de hipoxantina a IMP e guanina a GMP por meio da transferência da ribose-5-fosfato do PRPP para a base nitrogenada (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

Vale destacar que a deficiência de HPRT resulta na síndrome de Lesch-Nyhan, que se manifesta nos primeiros anos das crianças acometidas. Esta síndrome é caracterizada por deficiência intelectual e agressividade, manifestações decorrentes da dependência do cérebro das vias de salvação, além de presença de lesões teciduais semelhantes à gota, que

aparecem devido ao acúmulo de PRPP que, por sua vez, leva ao aumento da biossíntese de purinas pela via *de novo* e, conseqüentemente, a elevação nos níveis séricos de ácido úrico (TORRES; PUIG, 2007).

1.2 Degradação do ácido úrico

Em hominídeos, que inclui os gêneros *Homo*, *Pan*, *Gorilla* e *Pongo*, assim como em aves, répteis e insetos, o metabólito final no catabolismo de purinas é o ácido úrico, composto excretado pelo sistema renal e pelo trato gastrointestinal (NELSON; COX; LEHNINGER, 2017; RAMAZZINA *et al.*, 2006).

Já em outros mamíferos placentários, além anfíbios e peixes, esta via continua. Primeiro a urato-oxidase (uricase) catalisa a oxidação do urato a 5-hidroxiisourato (HIU), em seguida a HIU hidrolase (HIUase) catalisa a hidrólise do HIU a 2-oxo-4-hidroxi-4-carboxi-5-ureidoimidazolina (OHCU), e, por fim, a OHCU descarboxilase catalisa a descarboxilação do OHCU para gerar a (S)-alantoína, como demonstrado na **Figura 3** (ALVAREZ-LARIO; MACARRON-VICENTE, 2010; KEEBAUGH; THOMAS, 2010).

No caso dos anfíbios e peixes ocorrem, ainda, outras reações de degradação. A (S)-alantoína gerada passa por uma reação de hidrólise catalisada pela alantoinase, o que leva a formação de alantoato e, em seguida, a alantoicase catalisa a hidrólise do alantoato, produzindo (S)-ureidoglicolato e ureia, e a hidrólise do (S)-ureidoglicolato, que gera glioxilato e ureia (ALVAREZ-LARIO; MACARRON-VICENTE, 2010; KEEBAUGH; THOMAS, 2010). Além disso, conforme representado na **Figura 3**, em invertebrados marinhos, a ureia formada ainda sofre uma reação catalisada pela urease, sendo gerado amônio (ALVAREZ-LARIO; MACARRON-VICENTE, 2010).

Essas reações de degradação do ácido úrico e dos metabolitos formados em reações subsequentes favorecem a excreção dos produtos da via de catabolismo das purinas, pois há um aumento considerável na hidrossolubilidade dos compostos formados ao longo desta via. Enquanto a solubilidade do ácido úrico em água é 60 mg/L (20 °C), a alantoína apresenta uma solubilidade de 5,7 g/L (25 °C) e a ureia de 545 g/L (25 °C) (KIM *et al.*, 2023).

Por muito tempo, a HIUase e a OHCU descarboxilase não receberam tanta atenção quanto a uricase. Primeiro, porque a avaliação dos genes que codificam as enzimas envolvidas na degradação do ácido úrico é relativamente recente. Em estudos realizados no início dos anos 2000, foi discutido o papel do cluster de genes *pucABCDEFGHIJKLM*, de *Bacillus subtilis*, na via do catabolismo das purinas (SCHULTZ; NYGAARD; SAXILD, 2001). Em segundo lugar, porque o HIU é espontaneamente degradado em alantoína. Entretanto, essa reação espontânea produz uma mistura racêmica, composta pelos enantiômeros (S) e (R)-alantoína (**Figura 3**), e estudos indicaram que a formação de (R)-alantoína está associada ao aumento da proliferação celular e da oncogênese (STEVENSON *et al.*, 2010).

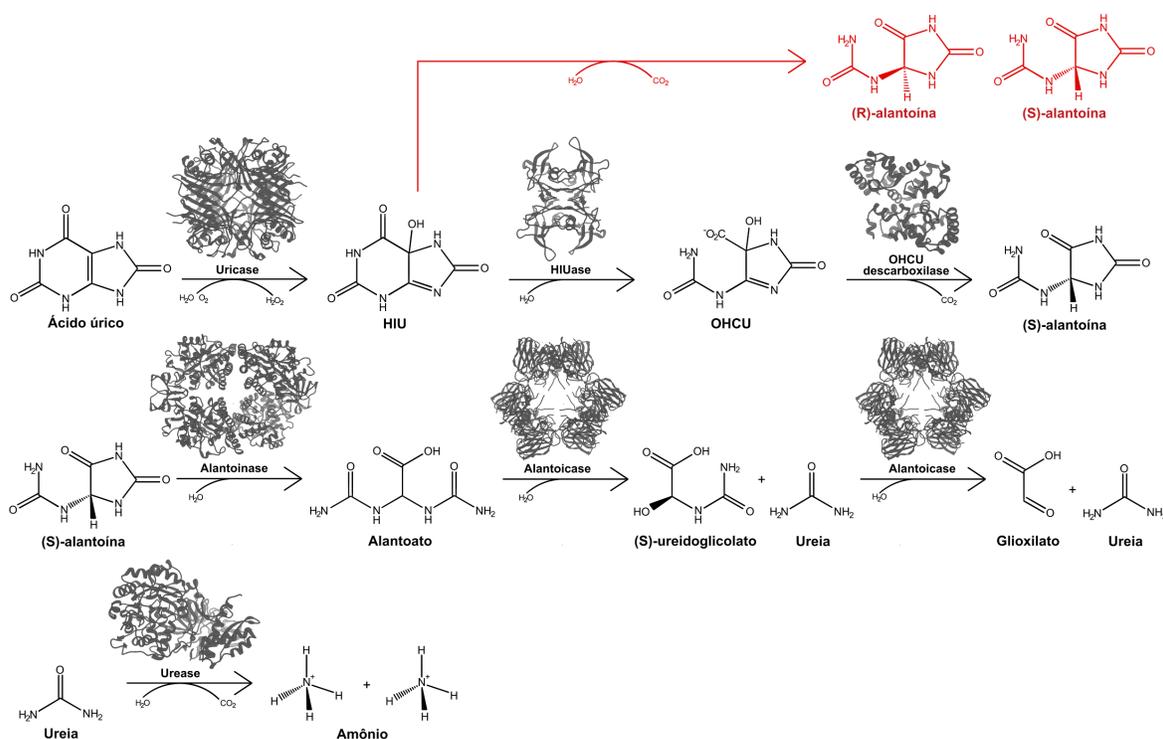


Figura 3. Via de degradação do ácido úrico. Esquema indicando os compostos intermediários formados no catabolismo do ácido úrico até amônio. As estruturas das enzimas que catalisam cada uma das reações (PDB 5LL1, 2H1X, 2O70, 3E74, 1SG3 e 2KAU), assim como a reação de degradação espontânea do HIU em (R) e (S)-alantoína (em vermelho), estão indicadas.

Ademais, nos primeiros estudos realizados com a HIUase a enzima foi denominada TRP/TLP (*transthyretin related/like proteins*) devido a sua semelhança estrutural com a transtirretina (TTR), responsável pelo transporte dos hormônios tireoidianos (RICHARDSON, 2015). Esta semelhança estrutural, indicada na **Figura 4**, diz respeito à sua estrutura tridimensional, sendo ambas homotetrâmeros, cujas quatro subunidades

apresentam predomínio de β -folhas (CARRIJO DE OLIVEIRA *et al.*, 2021; RICHARDSON, 2015). A análise da estrutura primária pelo alinhamento das sequências polipeptídicas da HIUase de *Danio rerio* e a TTR de *H. sapiens* indica uma identidade de 33,05% (SIEVERS *et al.*, 2011).

Ao avaliar os genes que codificam a TTR e a HIUase em vertebrados, notou-se que a estrutura de introns e exons dos genes era idêntica, o que levou a hipótese de que estas proteínas possuem um ancestral comum, tendo ocorrido um evento de duplicação no gene de uma HIUase ancestral, conforme ilustrado na **Figura 4** (CARRIJO DE OLIVEIRA *et al.*, 2021; PRAPUNPOJ *et al.*, 2000; ZANOTTI *et al.*, 2006). Ademais, em alguns organismos ocorreram outras duplicações no gene da HIUase, o que levou a presença de diferentes variantes da proteína. Um exemplo seria a presença de três variantes em *Arabidopsis thaliana*, sendo duas isoformas citoplasmáticas e a outra encontrada no peroxissomo (PESSOA *et al.*, 2010).

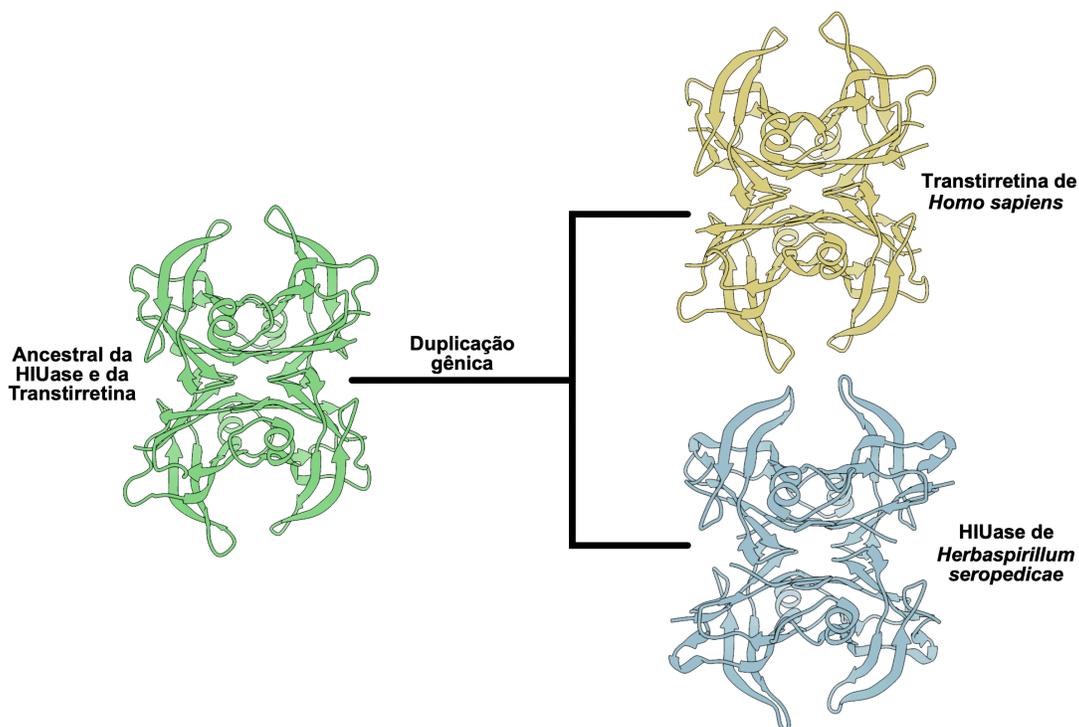


Figura 4. Resultado da duplicação do gene da HIUase ancestral. Esquema indicando a estrutura da HIUase ancestral (PDB 7KJJ), cujo gene sofreu um evento de duplicação. Enquanto uma das cópias manteve a HIUase (PDB 8UW4) como produto de expressão, a outra cópia, por meio do acúmulo de diversas mutações, passou a expressar a transtirretina (PDB 1F41).

Um fato interessante é que ao avaliar a atividade da HIUase de *A. thaliana*, Pessoa *et al.*, 2010, afirmaram que a enzima seria bifuncional, apresentando as funções de HIUase e o

OHCU descarboxilase (PESSOA *et al.*, 2010). Alguns anos depois, Juntaek Oh *et al* (2018) observaram que plantas e diatomáceas, algas unicelulares, apresentavam uma região gênica que correspondia a uma fusão dos genes destas duas enzimas (OH *et al.*, 2018). Ao resolver a estrutura da proteína de *Phaeodactylum tricornutum*, constataram que a enzima produzida, denominada alantoína sintase, correspondia a uma fusão das estruturas da HIUase e da OHCU descarboxilase, como pode ser visto na **Figura 5** (OH *et al.*, 2018).

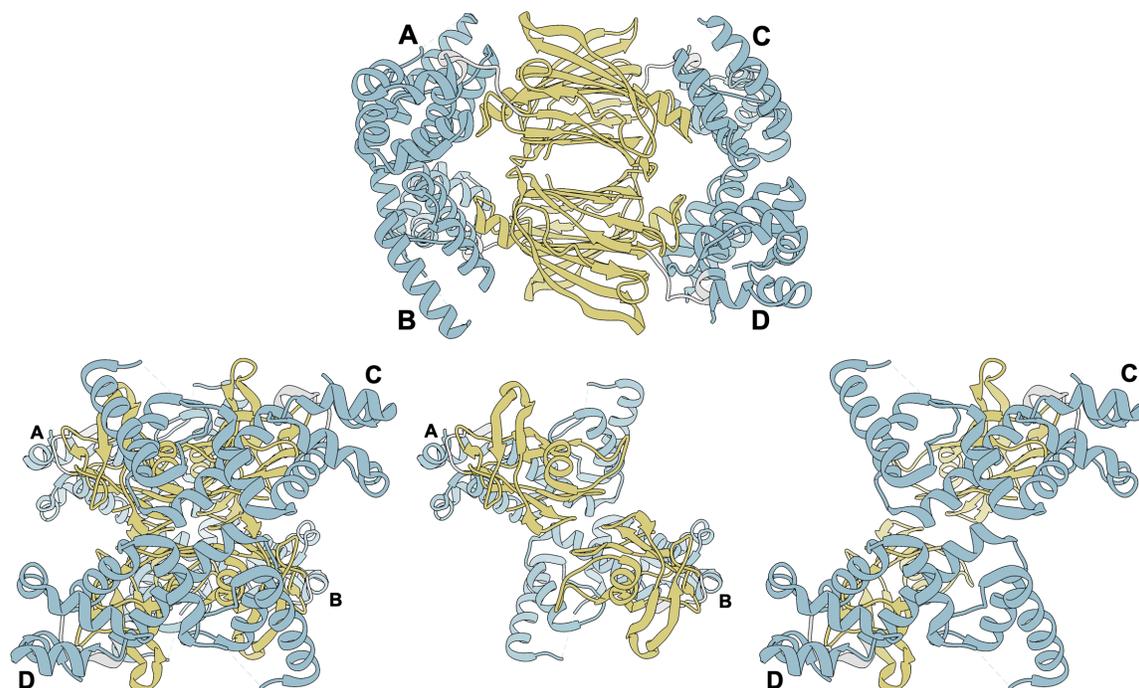


Figura 5. Estrutura da alantoína sintase. Visão frontal e lateral da estrutura da alantoína sintase (PDB 5Z5M), sendo identificadas as cadeias A, B, C e B, cada uma apresentando um domínio HIUase (amarelo) e OHCU descarboxilase (azul) unidas por uma curta sequência de ligação (branco).

Em *H. sapiens*, estudos envolvendo as enzimas da via de catabolismo das purinas indicaram que a ocorrência de mutações no gene da uricase levaram a uma redução na pressão seletiva para manter a integridade das regiões gênicas das outras enzimas da via, como a HIUase e a OHCU descarboxilase (KEEBAUGH; THOMAS, 2010; RAMAZZINA *et al.*, 2006).

1.3 Acúmulo de ácido úrico

A ausência da expressão destas enzimas traz consequências negativas em quadros de acúmulo de ácido úrico. Além da sua associação a gota, afecção de caráter inflamatório caracterizada pela deposição de cristais de urato monossódico (do inglês *monosodium urate*

– MSU) nas articulações, a hiperuricemia tem sido considerada um fator de risco para o desenvolvimento e/ou progressão de doenças cardiovasculares e diabetes (ROCK; KATAOKA; LAI, 2013; XIONG; LIU; XU, 2019), assim como um preditor da severidade do quadro de pré-eclâmpsia em gestantes (LÜSCHER *et al.*, 2022; WU *et al.*, 2012).

Essas associações são explicadas por diversos mecanismos. Os cristais de MSU, formados devido ao aumento dos níveis séricos de ácido úrico, são considerados um padrão molecular associado a dano (do inglês *damage-associated molecular pattern* – DAMP), ou seja, um composto que sinaliza para a ativação do sistema imune e, conseqüentemente, para iniciação do processo inflamatório (ROCK; KATAOKA; LAI, 2013; XIONG; LIU; XU, 2019).

No caso da gota, os cristais de MSU formados devido ao aumento dos níveis séricos de ácido úrico são capazes de promover o quadro inflamatório, em grande parte, pela ativação de monócitos e macrófagos, em um mecanismo dependente da atuação das caspases, proteases de cisteína ativadas pelo inflamassoma NLRP3 e responsáveis pela clivagem da pró-IL-1 β em IL-1 β , uma citocina pró-inflamatória (STAMP *et al.*, 2021).

Ademais, este processo inflamatório desencadeado pelo acúmulo de ácido úrico parece provocar diversas outras conseqüências, o que tem levado a associação da hiperuricemia com doenças cardiovasculares e síndromes metabólicas:

1) um aumento nas espécies reativas de oxigênio (do inglês *reactive oxygen species* – ROS), o que, além de ativar respostas inflamatórias, contribui para disfunção endotelial (XIONG; LIU; XU, 2019);

2) uma redução na biodisponibilidade de óxido nítrico (NO) no endotélio, o que também contribui para disfunção endotelial (XIONG; LIU; XU, 2019);

3) uma inibição das vias de sinalização da insulina, o que pode levar a um prejuízo no metabolismo e armazenamento de macromoléculas como carboidratos e lipídeos (PETERSEN; SHULMAN, 2018; XIONG; LIU; XU, 2019).

Outro quadro patológico que parece estar associado à hiperuricemia é a doença renal crônica (do inglês *chronic kidney disease* – CKD). Estudos indicam que a hiperuricosúria, que corresponde ao aumento da excreção urinária de ácido úrico, leva a lesão renal devido a

deposição de MSU nos néfrons, com conseqüente obstrução das estruturas tubulares. Além disso, a presença dos cristais nos rins parece induzir a inflamação do tecido renal, levando, em última análise, ao desenvolvimento e progressão da doença renal crônica (MALLAT *et al.*, 2016; STAMP *et al.*, 2021).

Por fim, a disfunção endotelial induzida pelo aumento dos níveis séricos de ácido úrico parece estar envolvida na vasoconstrição e remodelação estrutural do tecido pulmonar, que, por sua vez, favorecem o estabelecimento e o agravamento de quadros como a hipertensão arterial pulmonar (do inglês *pulmonary arterial hypertension* – PAH) e a doença pulmonar obstrutiva crônica (DPOC) (LIU *et al.*, 2022).

Outro ponto muito importante é que um número considerável de medicamentos amplamente utilizados pela população apresenta como efeito adverso o aumento nos níveis de ácido úrico (BEN SALEM *et al.*, 2016).

O uso de diuréticos de alça e tiazídicos, por exemplo, parece levar ao quadro de hiperuricemia devido a alteração na secreção e na reabsorção de ácido úrico provocada pela interação destes fármacos com transportadores de ânions orgânicos, como OAT1, OAT3 e OAT4, e transportadores de efluxo, como NPT4 e MRP4, localizados nos néfrons, as unidades funcionais básicas dos rins (BEN SALEM *et al.*, 2016).

Imunossuppressores inibidores da calcineurina, em especial a ciclosporina, fármacos utilizados no tratamento da tuberculose, como a pirazinamida, e a aspirina, um anti-inflamatório não esteroideal utilizado em baixas doses para prevenção secundária de eventos isquêmicos, também parecem afetar a atividade destes transportadores renais, levando ao quadro de hiperuricemia (BEN SALEM *et al.*, 2016).

1.4 Regiões gênicas das enzimas envolvidas na degradação do ácido úrico

Apesar da ausência de expressão gênicas que codificariam para as enzimas envolvidas na degradação do ácido úrico, durante a busca por essas enzimas no genoma humano utilizando o Ensembl, uma ferramenta que utiliza anotações do transcriptoma humano, são identificadas, conforme explicitado na **Tabela 1**, as regiões gênicas UOX, URAHP, URAD e ALLC, que correspondem à uricase, à HIUase, à OHCU decarboxilase e à alantoicase, respectivamente (CUNNINGHAM *et al.*, 2022).

Tabela 1: Regiões gênicas do catabolismo de ácido úrico no genoma humano

Enzima	Gene	Ensembl 109	Ensembl 111
Urato oxidase	UOX	ENSG00000240520	ENSG00000240520
5-hidroxiisourato hidrolase	URAHP	ENSG00000222019	ENSG00000222019 ENSG00000290738
2-oxo-4-hidroxi-4-carboxi- 5-ureidoimidazolina descarboxilase	URAD	ENSG00000183463	ENSG00000183463
Alantoinase	NA	NA	NA
Alantoicase	ALLC	ENSG00000151360	ENSG00000151360

Embora a análise ao nível do gene forneça várias informações, é importante lembrar que os RNAs gerados a partir da transcrição do gene podem passar por processamento, incluindo *splicing*. O *splicing* gera múltiplos transcritos, tanto codificadores de proteínas, cuja tradução leva à síntese da cadeia polipeptídica, quanto não codificadores de proteínas, podendo cada uma das isoformas de transcrito exercer um papel diferente (ULE; BLENCOWE, 2019).

Um transcrito é classificado como codificador de proteínas ou não codificador com base na presença ou ausência de regiões abertas de leitura (do inglês *Open Reading Frames* – ORF) ou sequência codificadora (do inglês *coding sequence* – CDS), e apesar de muita atenção ser dada aos transcritos codificadores de proteínas, menos de 30% das regiões de genes anotados são classificadas como genes codificadores de proteínas (FRANKISH *et al.*, 2021).

Além disso, os RNAs não codificadores (ncRNAs) podem exibir um maior número de isoformas de *splicing* alternativo do que os genes codificadores de proteínas. Em seu estudo, Deveson *et al.*, 2017, propõem que a presença de um número quase ilimitado de isoformas de RNA não codificadores é devido a ausência de restrições que governam o processo de *splicing* alternativo de sequências codificantes, como a necessidade de manutenção de uma ORF (DEVESON *et al.*, 2018).

Reforçando estes achados, Melé *et al.*, 2017, mostraram que os RNAs longos não codificadores intergênicos (do inglês *long intergenic non-coding RNA* – lincRNAs) se

distinguem dos mRNA por sua menor eficiência de *splicing*, o que está relacionado a menor ligação aos fatores de *splicing* e à presença de motivos relacionados ao *splicing* mais fracos, e expressão, devido ao fato de seus promotores estarem isentos da maioria das marcas de histonas e fatores de transcrição (MELÉ *et al.*, 2017).

Assim, a análise ao nível do gene não abrange toda a diversidade de transcritos codificadores e não codificadores gerados a partir da transcrição do gene, com a análise ao nível do transcrito sendo capaz de gerar uma quantidade muito maior de informações. Portanto, para avaliar os papéis potenciais de UOX, URAHP, URAD e ALLC, durante a busca utilizando o Ensembl (CUNNINGHAM *et al.*, 2022), também foram avaliados os transcritos destas regiões gênicas, indicados na **Tabela 2**.

Tabela 2: Transcritos das regiões gênicas do catabolismo de ácido úrico no genoma humano

Gene	Ensembl 109	Tipo de transcrito	Ensembl 111	Tipo de transcrito
UOX	ENST00000471089	pseudogene unitário transcrito	ENST00000471089	pseudogene unitário transcrito
	ENST00000483236	transcrito processado		
	ENST00000638373	transcrito processado		
	ENST00000638856	transcrito processado		
	ENST00000640983	transcrito processado		
URAHP	ENST00000610227	transcrito não processado	ENST00000610227	transcrito não processado
	ENST00000409768	lncRNA	ENST00000409768	lncRNA
	ENST00000409873	lncRNA	ENST00000409873	lncRNA
	ENST00000517889	lncRNA	ENST00000517889	lncRNA
	ENST00000521551	lncRNA	ENST00000521551	lncRNA
URAD	ENST00000332715	codificante de proteína	ENST00000332715	codificante de proteína
			ENST00000713555	codificante de proteína, CDS não definida
ALLC	ENST00000252505	codificante de proteína	ENST00000252505	codificante de proteína
	ENST00000471711	codificante de proteína, CDS não definida	ENST00000471711	codificante de proteína, CDS não definida
	ENST00000476389	retenção de íntron	ENST00000476389	retenção de íntron

Como podemos ver nas **Tabela 2**, para os genes URAD e ALLC há transcritos indicados como sendo codificantes de proteínas, contudo existe uma grande diversidade de transcritos não produtivos para UOX, URAHP, URAD e ALLC.

Alguns transcritos são classificados como pseudogenes, os quais têm origem em genes codificadores de proteínas ancestrais que sofreram mutações, incluindo códons de parada em fase e inserções/deleções de quadros de leitura (CHEETHAM; FAULKNER; DINGER, 2020; FRANKISH *et al.*, 2019). Eles são categorizados em pseudogenes processados, resultantes da retrotransposição de mRNAs processados; pseudogenes não processados, originados de eventos de duplicação; e pseudogenes unitários, formados por interrupções como mutações inativadoras em um único gene ancestral (CHEETHAM; FAULKNER; DINGER, 2020; FRANKISH *et al.*, 2019). Além disso, no caso de ALLC, há um transcrito, ALLC-203, caracterizado por 'retenção de íntrons', indicando a presença de sequências intrônicas (FRANKISH *et al.*, 2019).

Embora os pseudogenes tradicionalmente recebam menos atenção do que os genes codificadores de proteínas, um crescente corpo de pesquisa revela um aumento no número de pseudogenes funcionais sendo identificados e investigados, sugerindo que desempenham papéis biológicos significativos (CHEETHAM; FAULKNER; DINGER, 2020).

Outro tipo de transcrito classificado como não codificador de proteínas de forma canônica mostrado na Tabela 2 são os RNAs longos não codificadores (do inglês *long non-coding RNA* - lncRNAs). Atualmente, sabe-se que estas macromoléculas, normalmente compostas por mais de 200 nucleotídeos, apresentam diversos papéis regulatórios (STATELLO *et al.*, 2021). Eles são capazes de alterar o grau de compactação da cromatina por meio do recrutamento de complexos de remodelação da cromatina, podem interagir com fatores de transcrição e com regiões promotoras no DNA e, além disso, podem apresentar regiões de complementariedade com RNA mensageiros (do inglês *messenger RNA* – mRNA) (MATTICK *et al.*, 2023; STATELLO *et al.*, 2021). Desta forma, os lncRNA podem alterar a expressão gênica por diversos mecanismos.

Estudos recentes também já reconheceram o potencial dos lncRNA de serem traduzidos por vias alternativas, levando a formação de peptídeos (RION; RÜEGG, 2017; XING *et al.*, 2021). As vias alternativas incluem 1) mecanismo que envolvem elementos de tradução independentes de CAP5' (do inglês *CAP-independent translation elements sites* – CITE); 2) reconhecimento de sítios de entrada internos do ribossomo (do inglês *internal ribosome entry sites* – IRES); 3) modificações pós-transcricionais de RNA, como a

metilação de nucleotídeos de adenina; e 4) escape do ribossomo (KWAN; THOMPSON, 2019).

Assim como os próprios lncRNA, os peptídeos formados a partir da tradução destas moléculas de RNA parecem desempenhar diversas funções regulatórias. Um exemplo é o polipeptídeo SPAR, que atua no controle do processo de reciclagem de aminoácidos no lisossomo por bloqueio da ação de ATPases, o que reduz a ativação do alvo mamífero do complexo 1 de rapamicina (do inglês *mammalian target of rapamycin complex 1* – mTORC1) e altera o processo de acidificação do lisossomo (XING *et al.*, 2021).

Ademais, as sequências polipeptídicas formadas a partir da expressão de lncRNA parecem exercer papéis importantes na patogênese de certas doenças, podendo provocar alterações na proliferação celular em certos tipos de câncer e desregular o processo de apresentação de antígenos em quadros autoimunes (XING *et al.*, 2021).

Apesar destes peptídeos apresentarem funções importantes, é possível que grande parte destas sequências seja direcionada para a via de degradação nos peroxissomos (KESNER *et al.*, 2023). Kesner *et al.*, 2023, avaliaram diversos peptídeos sintetizados a partir de regiões não codificantes e notaram que estas macromoléculas apresentam uma cauda hidrofóbica na região C-terminal, a qual interage com o complexo proteico BAG6 e, conseqüentemente, sinaliza para a degradação destes peptídeos nos proteassomos (KESNER *et al.*, 2023).

Contudo, também foi demonstrado que estes peptídeos podem escapar deste mecanismo de degradação e ser expostos na membrana plasmática ou na membrana das organelas (KESNER *et al.*, 2023). A exposição de peptídeos sintetizados a partir de regiões não codificantes em membranas biológicas, juntamente com a formação de um número praticamente ilimitado de isoformas de *splicing* alternativo, contribui com o entendimento do processo de evolução de genes e da aquisição de novas funções por uma região gênica (DEVESON *et al.*, 2018; KESNER *et al.*, 2023).

Por serem capazes de atuar na expressão gênica, os lncRNA têm sido alvo de estudos que visam elucidar os mecanismos envolvidos na patogênese de certas doenças. Na própria hiperuricemia, estudos indicam que o lncRNA ANRIL (HU *et al.*, 2019) e o lncRNA HOTAIR (CHI *et al.*, 2021) podem estar envolvidos no estabelecimento e progressão do

processo inflamatório. No estudo de Hu *et al.*, 2019, foi indicado que o aumento na expressão do lncRNA ANRIL parece induzir o aumento na expressão das citocinas pró-inflamatórias IL-6 e IL-8 (HU *et al.*, 2019). No caso do lncRNA HOTAIR, este parece ser capaz de promover a ativação do inflamassoma NLRP3 ao ligar-se competitivamente ao microRNA miR-22 (CHI *et al.*, 2021).

2. PRIMEIRA PROPOSTA

Considerando todas as informações trazidas anteriormente, formulamos a hipótese de que a expressão dos transcritos das regiões gênicas UOX, URAHP, URAD e ALLC poderia contribuir para a regulação do catabolismo das purinas e a patogênese de doenças associadas à hiperuricemia. Para testar esta hipótese, analisamos dados da literatura, assim como de base de dados de acesso aberto, para levantar informações sobre os produtos (RNA e proteína) e perfis de expressão dessas regiões gênicas.

3. MÉTODO

3.1 Avaliação dos produtos de expressão das regiões UOX, URAHP, URAD e ALLC

Primeiro, para investigar possíveis papéis regulatórios das regiões UOX, URAHP, URAD e ALLC, foi realizada uma busca na literatura por evidências da expressão diferencial destas regiões. Para as que apresentam transcritos produtivos, também foi avaliada a possibilidade de encontrar os produtos proteicos correspondentes no UniProt (*Universal Protein Resource*) (THE UNIPROT CONSORTIUM *et al.*, 2023), um consórcio que fornece uma base de dados abrangente e de alta qualidade sobre sequências e informações funcionais de proteínas, assim como no PaxDb (HUANG *et al.*, 2023), um banco de dados de proteoma.

A partir do PaxDb (HUANG *et al.*, 2023) também foram obtidas informações provenientes do String (SZKLARCZYK *et al.*, 2023), ferramenta que facilita a exploração de associações funcionais entre proteínas, abrangendo tanto interações diretas quanto indiretas, as quais indicam que as proteínas compartilham papéis biológicos. Essas associações são categorizadas em sete canais de evidência distintos.

O canal 'Experimentos' compila evidências de experimentos de laboratório especificamente projetados para elucidar associações proteína-proteína. 'Banco de dados' importa informações de recursos bem estabelecidos como KEGG e Complexos de Ontologia Gênica, consolidando o conhecimento sobre complexos proteicos, vias e outras conexões funcionais (SZKLARCZYK *et al.*, 2023).

Os canais 'Vizinhança, Fusão e Co-ocorrência' utilizam contexto genômico para inferir associações exclusivamente a partir de sequências genômicas. Eles detectam restrições evolutivas indicativas de parcerias funcionais entre genes, sendo especialmente adequadas para genomas procarióticos (SZKLARCZYK *et al.*, 2023).

O canal 'Mineração de Texto' identifica, por meio da análise de artigos completos, resumos do PubMed e textos resumidos de recursos como OMIM e SGD, co-ocorrências de pares de proteínas. Essas menções são avaliadas em relação às frequências de menções individuais de proteínas, oferecendo insights sobre possíveis associações. 'Co-expressão' analisa medidas de genômica funcional (transcritos ou proteínas) em diversas condições, buscando evidências de regulação de expressão compartilhada (SZKLARCZYK *et al.*, 2023).

3.2 Avaliação dos perfis de expressão das regiões UOX, URAHP, URAD e ALLC

Para avaliar o padrão de expressão das regiões gênicas de interesse foi feita uma busca pelas regiões gênicas citadas acima no portal do GTEx (*Genotype-Tissue Expression*), sendo utilizado o dbGaP (*database of Genotypes and Phenotypes*), número de acesso phs000424.v8.p2. O portal, construído a partir de um projeto realizado com amostras de 54 tecidos distintos de quase 1000 indivíduos, fornece acesso aberto a dados de expressão gênica obtidos por RNA-Seq. Desta forma, seria possível identificar se a expressão das regiões gênicas de interesse seria tecido-específica e, caso seja, quais tecidos apresentam um maior nível de expressão, sendo estes alvos promissores para nossa análise.

Dentre os arquivos de acesso aberto, foi feito o *download* de:

- GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt: arquivo texto com as anotações de todas as amostras de acesso aberto. Todas as amostras apresentam um identificador (SAMPID) único e estão incluídas informações como o tecido do qual a amostra foi obtida (SMTS e SMTSD);

- `GTEX_Analysis_v8_Annotations_SampleAttributesDD.xlsx`: planilha que descreve cada coluna presente no `GTEX_Analysis_v8_Annotations_SampleAttributesDS.txt`, informando o tipo de informação está contido em cada coluna do arquivo de anotação e que é representado por uma abreviação ou sigla (SAMPID, SMTS, SMTSD, entre outros);

- `GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct`: arquivo que contém a quantidade relativa de transcritos de cada gene em cada amostra em em transcritos por milhão (do inglês *transcripts per million* – TPM), ou seja, levando em consideração tanto o comprimento do transcrito quanto o número total de *reads* mapeadas na amostra;

- `GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct`: arquivo que contém os valores da mediana da quantidade relativa de transcritos, expressa em TPM, de cada gene de cada região gênica em cada um dos tecidos, ou seja, uma sumarização dos dados de expressão das amostras considerando o tecido do qual foram obtidas;

- `GTEX_Analysis_2017-06-05_v8_RSEMv1.3.0_transcript_tpm.gct`: arquivo que contém a quantidade relativa de cada isoforma de transcrito, expressa em TPM, para todas as regiões gênicas em cada uma das amostras.

Primeiro, foi realizada a seleção das regiões gênicas de interesse (UOX, URAHP, URAD e ALLC) nos arquivos contendo dados de expressão. Como os arquivos mais recentes do GTEx foram gerados em 2017, os identificadores das regiões gênicas seguiam a versão 109 do Ensembl, sendo eles ENSG00000240520 (UOX), ENSG00000222019 (URAHP), ENSG00000183463 (URAD) e ENSG00000151360 (ALLC) (CUNNINGHAM *et al.*, 2022). Além disso, como o objetivo desta análise era avaliar a expressão das regiões gênicas de acordo com o tecido, no arquivo de anotação foi feita a seleção apenas dos atributos SAMPID, SMTS e SMTSD.

Como para no arquivo de anotação cada amostra estava representada em uma linha e nos arquivos contendo os valores de TPM para os genes e para os transcritos cada amostra estava representada em uma coluna, foi feita a inversão das linhas e colunas dos arquivos contendo os valores de TPM para possibilitar a combinação das informações constantes nos arquivos.

O arquivo texto gerado foi, então, importado para o RStudio (RStudio Team, 2020), onde as amostras foram agrupadas por tecido e, em seguida, de acordo com o sistema ao qual o tecido pertencia. Os dados de mediana foram plotados em gráficos de dispersão e os demais, referentes aos TPM dos genes e transcritos, foram plotados em gráficos de violino, em ambos os casos tendo sido utilizado o pacote ggplot2 (WICKHAM, 2016). O gráfico de violino é utilizado para visualizar a distribuição de dados numéricos, sendo representadas tanto as estatísticas resumidas quanto a densidade de cada variável.

4. RESULTADOS E DISCUSSÃO

4.1 Produtos de expressão das regiões UOX, URAHP, URAD e ALLC

Ao realizar a busca pelas regiões gênicas de interesse na literatura foi possível encontrar o trabalho de Li e Li, 2020, no qual os autores evidenciaram que a transcrição do URAHP é aumentada em grávidas com pré-eclâmpsia (LI; LI, 2020). Já no estudo de Griffen *et al.*, 2021, a expressão do URAHP apresentou uma associação positiva com a ocorrência de recidiva em pacientes diagnosticados com leucemia linfocítica crônica (LLC) (GRIFFEN *et al.*, 2021).

Como alguns estudos indicam que há uma associação entre a hiperuricemia e a pré-eclâmpsia (LÜSCHER *et al.*, 2022; WU *et al.*, 2012), poderia ser considerada a possibilidade de no estudo de Li e Li (LI; LI, 2020) a regulação exercida pela URAHP sobre a patogênese pré-eclâmpsia estar de alguma forma vinculada ao catabolismo de purinas. Contudo, este não parece ser o caso, visto que neste estudo o aumento da expressão do URAHP estava associado, de alguma forma, a expressão de regiões gênicas que participam de vias envolvidas na regulação do crescimento e diferenciação celular, como o PPP5C (do inglês *Protein Phosphatase 5 Catalytic Subunit*) (LI; LI, 2020).

Além disso, foi encontrado o trabalho de Giaccherini *et al.*, 2023, que, ao estudar variantes na região codificadora do genoma humano, identificou uma associação significativa entre o SNP 13q12.2-rs9581957, uma variante intrônica do URAD, e o aumento do risco de adenocarcinoma pancreático (GIACCHERINI *et al.*, 2023). Eles discutem que, embora o papel da URAD em humanos seja incerto, evidências recentes mostram que sua expressão foi detectada em vários órgãos do sistema digestivo e, além disso, sugerem sua atividade em peroxissomos. Os autores hipotetizaram que a presença do

SNP 13q12.2-rs9581957, que implica na alteração do resíduo de arginina (Arg) 144 por uma serina (Ser), poderia levar a produção de um produto proteico da URAD defeituoso e, então, prejudicar o processo de desintoxicação nos peroxissomos, potencialmente aumentando o risco de adenocarcinoma pancreático (GIACCHERINI *et al.*, 2023).

Prosseguindo com as análises dos produtos proteicos dos transcritos produtivos, ao buscar a proteína sintetizada a partir da expressão do URAD no Uniprot (THE UNIPROT CONSORTIUM *et al.*, 2023) encontramos a entrada A6NGE7, na qual é indicado que a presença da OHCU descarboxilase na célula é incerta. No caso da alantoicase, que seria codificada pelo ALLC, a entrada no Uniprot (THE UNIPROT CONSORTIUM *et al.*, 2023) é Q8N6M5 e é indicado que existe evidência de sua presença apenas a nível de transcrito.

Para ambas entradas é possível notar que não há estrutura resolvida ou calculada experimentalmente, havendo o modelo gerado pelo AlphaFold (JUMPER *et al.*, 2021), ferramenta capaz de prever a estrutura de uma proteína utilizando a aprendizagem profunda, cuja capacidade de leitura de padrões foi obtida por meio da análise de estruturas depositadas em bancos de dados, como o PDB (BERMAN, 2000).

Ao avaliar os dados de proteoma no PaxDb (HUANG *et al.*, 2023) foi possível encontrar os produtos proteicos tanto da URAD quanto da ALLC. No caso da URAD, sua proteína foi encontrada em abundâncias que variam entre 0,02 e 12,5 ppm a depender do tipo celular ou tecido avaliado (HUANG *et al.*, 2023). Já para a ALLC, os valores variam entre 0,02 e 106 ppm (HUANG *et al.*, 2023). Esta potencial presença dos produtos proteicos da URAD e da ALLC em bancos de dados de proteoma sugere que houve alguma pressão para manter a integridade dessas regiões gênicas, indicando que estas proteínas podem desempenhar algum papel no organismo.

Os resultados do String (SZKLARCZYK *et al.*, 2023), que indicam as associações funcionais destas proteínas e que podem, desta forma, auxiliar no entendimento do papel dos produtos proteicos da URAD e da ALLC, podem ser vistos na **Figura 6**.

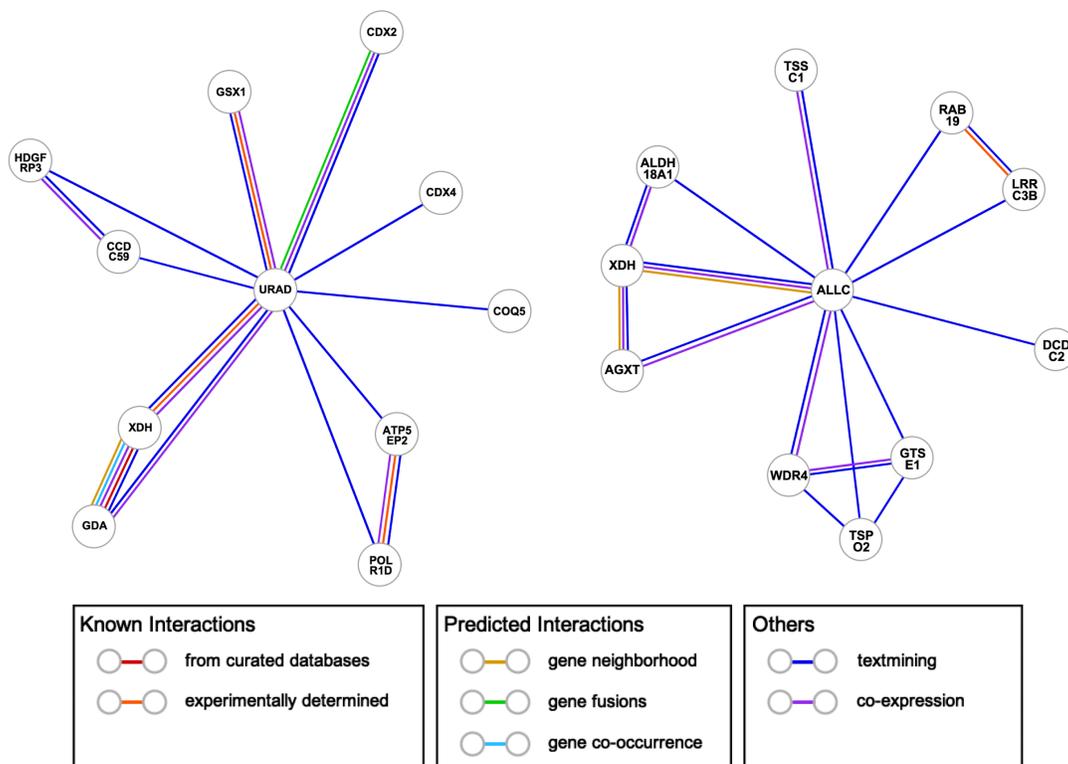


Figura 6. Rede de interação do URAD e ALLC com genes codificadores de proteínas. A figura mostra as redes de interação geradas a partir do banco de dados STRING que ilustram as conexões entre o gene de interesse (URAD - à esquerda e ALLC - à direita) e genes codificadores de proteínas. As interações podem surgir da vizinhança dos genes, fusão de genes, co-ocorrência de genes e co-expressão, entre outros.

Entre as interações da URAD é possível notar uma associação com a guanina desaminase (GDA) e a xantina oxidase (XDH), categorizadas nos canais 'mineração de texto' e 'co-expressão' (SZKLARCZYK *et al.*, 2023). Além disso, a associação URAD-XDH também se enquadra no canal 'experimentos' (SZKLARCZYK *et al.*, 2023). Ambas as enzimas desempenham um papel na degradação de purinas, com a GDA catalisando a desaminação da guanina para xantina, e a XDH, como mencionado anteriormente, catalisando a oxidação da hipoxantina para xantina e, em seguida, para ácido úrico (THE UNIPROT CONSORTIUM *et al.*, 2023).

Quanto ao ALLC, as interações são observadas principalmente com a xantina oxidase (XDH), categorizadas nos canais 'mineração de texto', 'co-expressão' e 'vizinhança' (SZKLARCZYK *et al.*, 2023). Além disso, associações com a alanina-glioxilato aminotransferase (AGXT) são identificadas nos canais 'mineração de texto' e 'co-expressão' (SZKLARCZYK *et al.*, 2023). A AGXT é responsável por catalisar a reação na qual o

glioxilato, produto da enzima ALLC, e a alanina são utilizados para gerar piruvato e glicina (THE UNIPROT CONSORTIUM *et al.*, 2023).

Assim, os dados do STRING indicam que a expressão de URAD e ALLC poderia de fato estar associada à via catabólica das purinas, pois há uma aparente co-expressão dessas regiões com os genes da guanina desaminase (GDA), da xantina oxidase (XDH) e da alanina-glioxilato aminotransferase (AGXT) (SZKLARCZYK *et al.*, 2023).

No entanto, a relação de URAD também é observada com fatores de transcrição (CDX2, CDX4 e GSX1), componentes da RNA polimerase (POLR1D) e enzimas envolvidas na cadeia de transporte de elétrons (COQ5 e ATP5EP2) (SZKLARCZYK *et al.*, 2023). Da mesma forma, ALLC está associado a enzimas de edição de tRNA (WDR4), proteínas ou fragmentos de peptídeos que parecem atuar como supressores de tumores (LRRC3B e TSSC1), proteínas envolvidas no transporte intracelular de colesterol (TSPO2), entre outros (SZKLARCZYK *et al.*, 2023). Portanto, conforme indicado na **Figura 6**, as análises do STRING também associam URAD e ALLC a outras proteínas-chave envolvidas em diferentes processos celulares, indicando funções regulatórias mais amplas.

4.2 GTEx: Expressão Genótipo-Tecido da UOX, URAHP, URAD e ALLC

Ao avaliar os gráficos de dispersão construídos utilizando os dados, obtidos nos arquivos do GTEx, da mediana de TPM em cada tecido para os genes de interesse é possível notar que na maioria dos tecidos ocorre a expressão da URAHP, enquanto para as outras regiões, UOX, URAD e ALLC, há poucos tecidos nos quais elas são expressas, como pode ser visto na **Figura 7**.

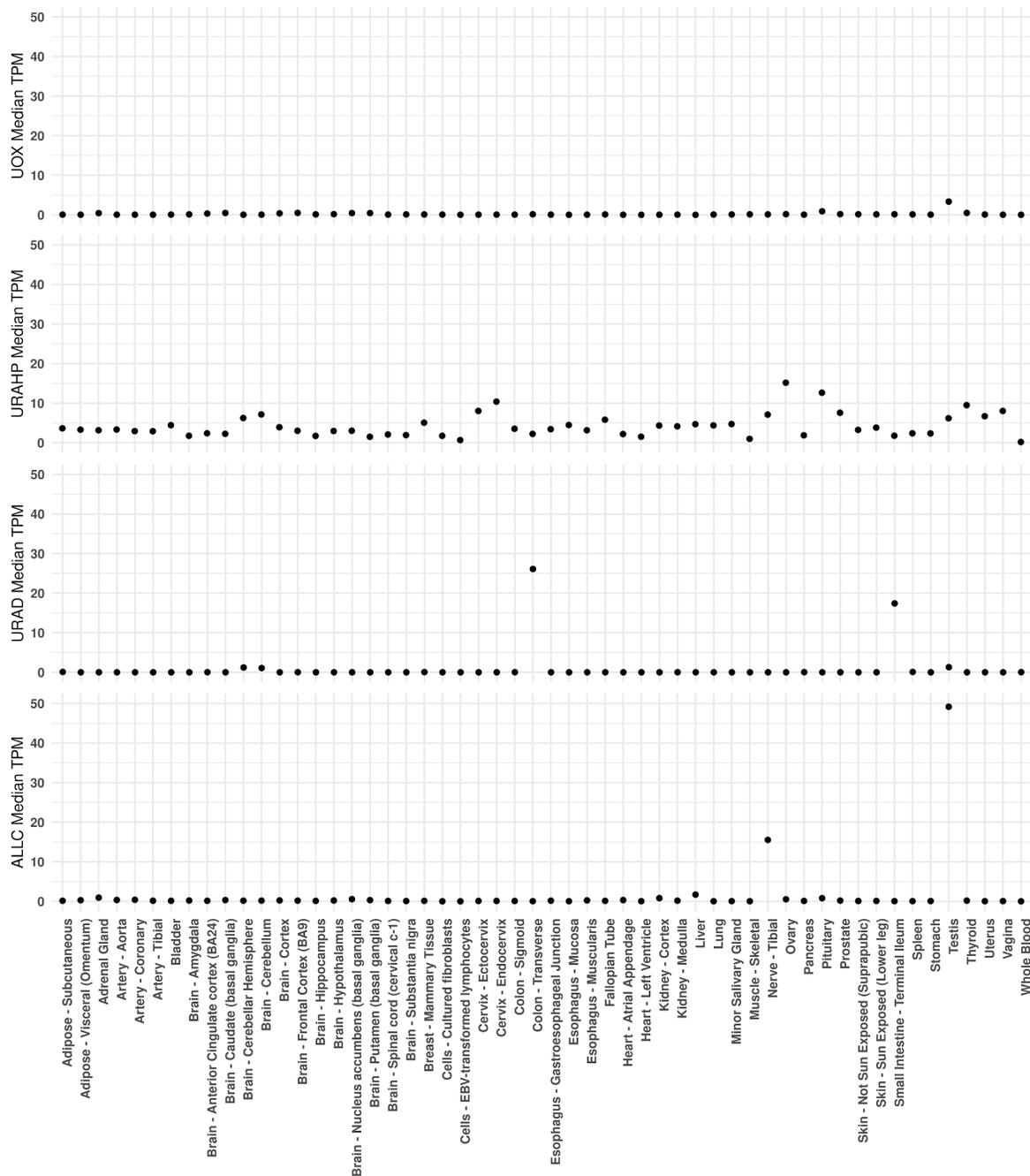


Figura 7. Expressão mediana dos genes do catabolismo do ácido úrico em tecidos humanos. Gráfico indicando a expressão mediana, em TPM, de cada uma das regiões gênicas UOX, URAHP, URAD e ALLC de acordo com diferentes tecidos analisados no projeto GTE_x.

Para melhor entender os padrões de expressão destas regiões gênicas também foi feita a avaliação dos gráficos de violino construídos a partir dos dados de expressão de todas as amostras para os genes de interesse. Para melhor visualização e análise dos dados, estes foram divididos de acordo com os sistemas (circulatório, renal, nervoso) e, considerando que os dados de expressão gênica podem variar em várias ordens de magnitude, foi

utilizada a transformação em logaritmo de base 2 das quantidades relativas de transcritos de cada gene.

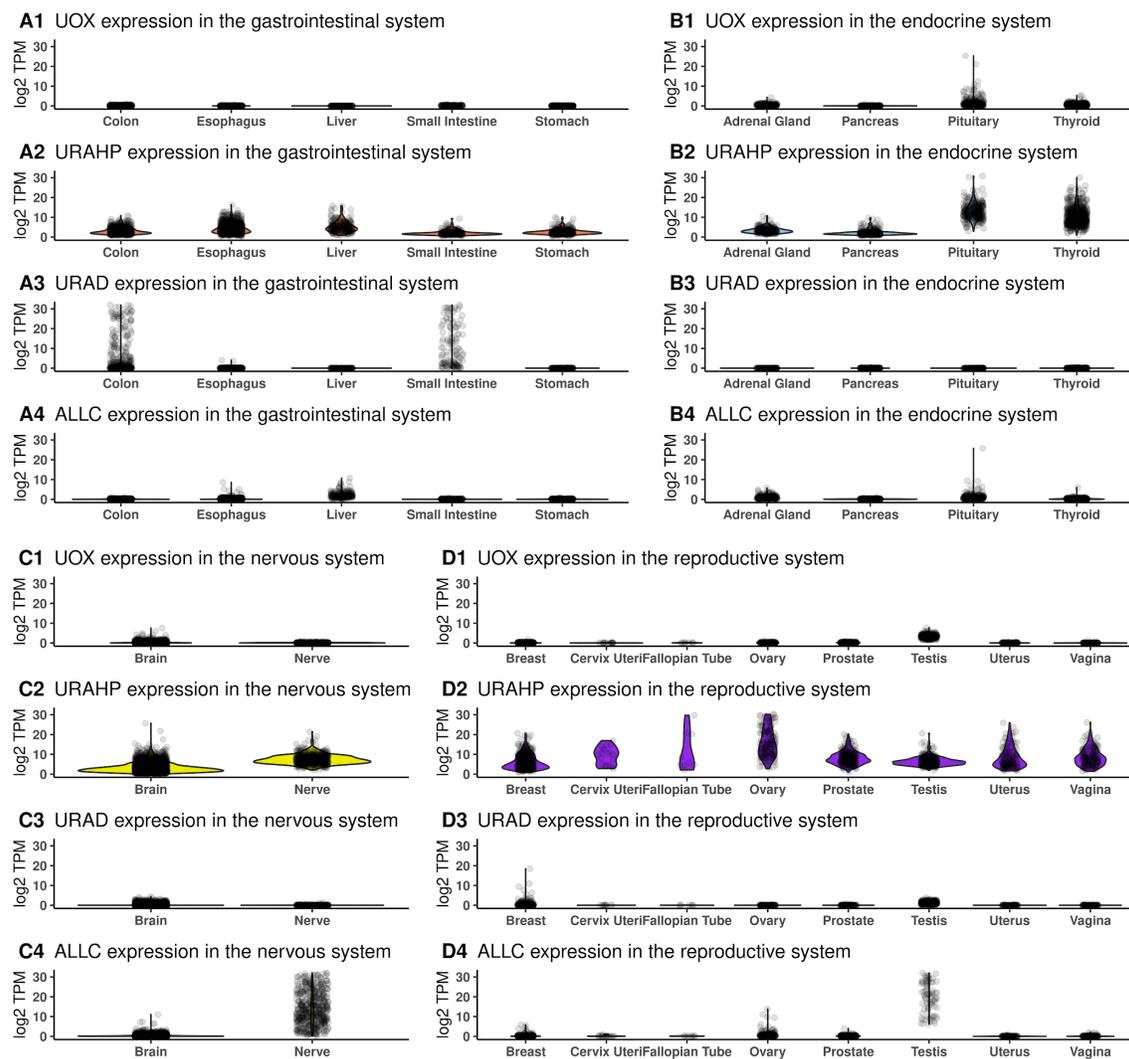


Figura 8. Perfis de expressão dos genes do catabolismo do ácido úrico em tecidos humanos por sistema. Os gráficos de violino mostram a expressão das regiões UOX, URAHP, URAD e ALLC em diferentes tecidos humanos, agrupados por sistemas fisiológicos (A - gastrointestinal, B - endócrino, C - nervoso, D - reprodutivo). O eixo x representa os diferentes tecidos, e o eixo y mostra os valores de transcritos por milhão (TPM).

Os dados apresentados na **Figura 8** revelam diferenças notáveis nos padrões de expressão entre as regiões dos genes UOX, URAD e ALLC em diversos tecidos humanos, categorizados por sistemas fisiológicos, quando comparados com a URAHP. A expressão mais pronunciada de URAHP em grande parte dos tecidos, em comparação com as outras regiões gênicas, sugere um papel potencialmente significativo em processos celulares.

O nível de expressão de URAHP parece ser fortemente dependente do tecido, como demonstrado nas **Figuras 7 e 8**. Notavelmente, tecidos do sistema reprodutivo feminino, incluindo os ovários e o colo do útero, exibem níveis de expressão de URAHP significativamente elevados. Da mesma forma, tecidos endócrinos, como a glândula pituitária, demonstram uma expressão aumentada de URAHP. Por outro lado, em certos tecidos como o intestino delgado e o estômago, a expressão de URAHP parece ser comparativamente mais baixa.

A análise de expressão da UOX em diferentes tecidos revela que no sistema reprodutor masculino (testículos), sistema nervoso central (cérebro) e sistema endócrino (glândulas pituitária e tireoide) ocorre uma maior expressão desta região gênica (**Figuras 7 e 8**). No entanto, vale ressaltar que os níveis de TPM são consideravelmente mais baixos do que os de URAHP quando avaliados nos mesmos tecidos.

A URAD exibe perfis de expressão distintos, mostrando uma expressão aumentada no trato gastrointestinal, especificamente no cólon e no intestino delgado. Além disso, níveis moderados de expressão são observados no cerebelo e nos testículos. Em contraste, ALLC parece ser mais expresso no nervo tibial e nos testículos, com uma expressão modesta sendo detectada no fígado (**Figuras 7 e 8**).

Essas observações destacam os mecanismos regulatórios tecido-específicos que governam a expressão de UOX, URAHP, URAD e ALLC, lançando luz sobre seus papéis potenciais dentro de diferentes contextos fisiológicos. Além disso, a distribuição da expressão gênica na maioria dos tecidos revela que, enquanto a maioria das amostras se agrupa em torno do valor de expressão mediano, há exceções notáveis onde amostras individuais exibem desvios consideráveis. Este é o caso da expressão de URAD no cólon e no intestino delgado e da expressão de ALLC no nervo tibial e nos testículos, que, conforme ilustrado na **Figura 8**, demonstram variabilidade significativa entre as amostras, resultando na ausência de uma forma de violino distinta nos gráficos. Essa variabilidade coloca em destaque a natureza dinâmica da regulação gênica dentro desses tecidos e ressalta a importância de considerar perfis de expressão individuais ao representar e interpretar os dados.

Considerando o maior nível de expressão da URAHP, essa região gênica foi selecionada para a análise a nível transcrito. Assim, quando o mesmo procedimento foi

aplicado para avaliar a dispersão das quantidades relativas de cada isoforma de transcrito da região gênica URAHP, foram obtidos os gráficos da **Figura 9**.

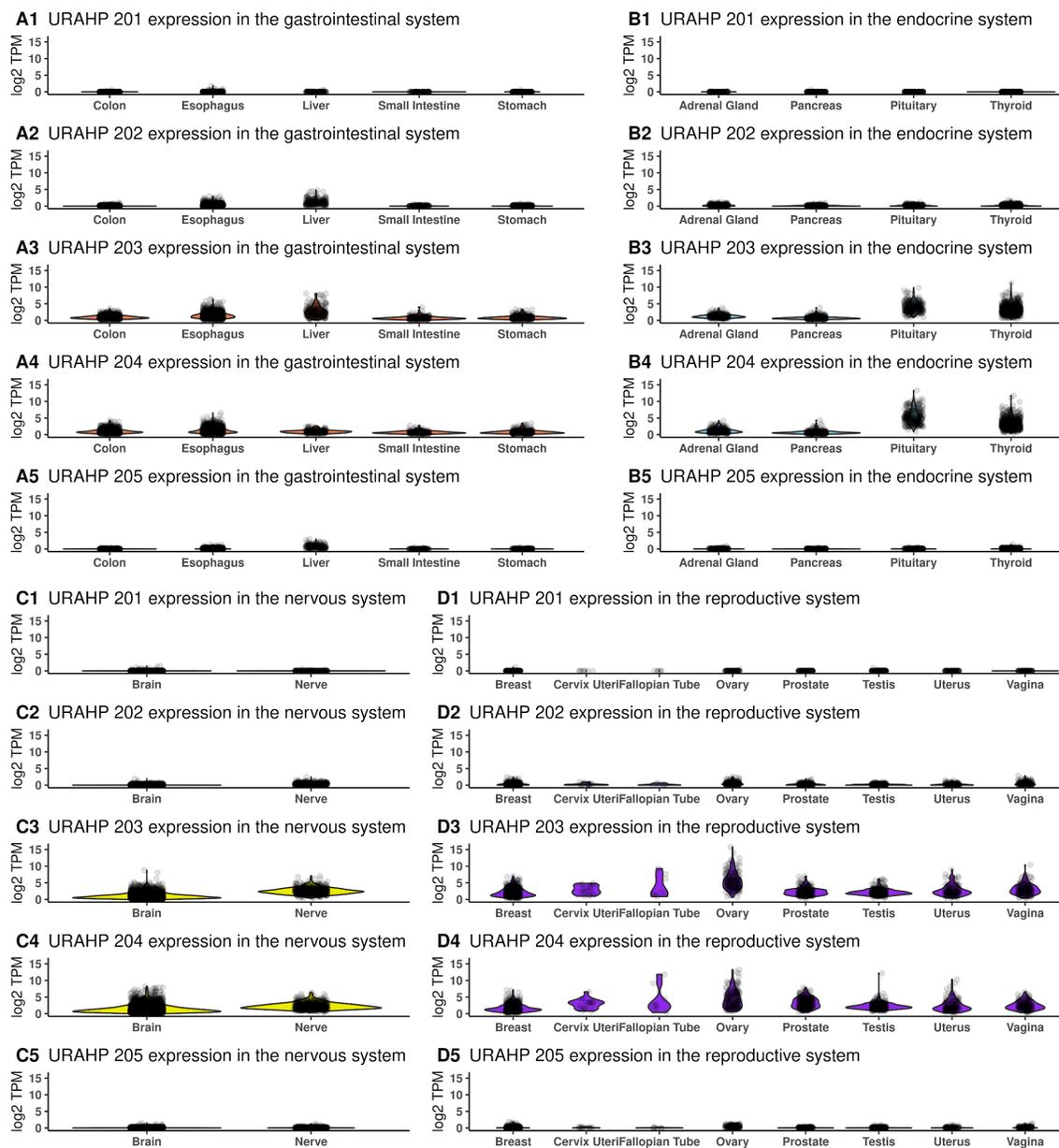


Figura 9. Perfis de expressão de transcritos da região do gene URAHP em tecidos humanos por sistema. O gráfico em forma de violino exibe os perfis de expressão de cinco transcritos (URAHP 201, URAHP 202, URAHP 203, URAHP 204 e URAHP 205) produzidos pela região do gene URAHP em vários tecidos humanos, agrupados por sistemas fisiológicos (A - Gastrointestinal, B - Endócrino, C - Nervoso, D - Reprodutivo). O eixo x representa diferentes tecidos, e o eixo y mostra os valores de Transcritos Por Milhão (TPM).

Notavelmente, as isoformas URAHP-203 e URAHP-204 exibiram um padrão de expressão distinto quando comparadas às isoformas URAHP-201, URAHP-202 e URAHP-205. O nível de expressão comparativamente mais alto dos transcritos

URAHP-203 e URAHP-204 (**Figura 9**) indica que mecanismos regulatórios diferenciais governam sua atividade transcricional. Assim, as diferentes isoformas de URAHP podem desempenhar papéis específicos em processos celulares.

Considerando a hipótese de que a expressão da UOX, URAHP, URAD e ALLC poderia desempenhar um papel na regulação da patogênese de doenças associadas à hiperuricemia em humanos, seria esperado que os transcritos dessas regiões gênicas fossem mais expressos na mucosa intestinal e no tecido hepático, onde ocorre a degradação de purinas e a formação de ácido úrico (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011). Além disso, a expressão também seria esperada nos tecidos endoteliais e renais, uma vez que o ácido úrico é transportado pelo sangue para os rins, onde sua excreção ocorre predominantemente (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

Contrariamente às expectativas, os dados do GTEx, representados nas **Figuras 7 e 8**, mostram diferentes padrões de expressão para UOX, URAHP, URAD e ALLC, indicando que estas regiões gênicas podem exercer papéis diferentes em tecidos não relacionados ao catabolismo de purinas ou à excreção de ácido úrico. Apesar de URAD apresentar uma expressão mais elevada no trato gastrointestinal, UOX é mais expresso em tecidos relacionados ao sistema reprodutivo masculino e ao sistema endócrino, URAHP é mais expresso em tecidos do sistema reprodutivo feminino e do sistema endócrino, e ALLC é mais expresso em tecidos do sistema nervoso periférico e do sistema reprodutivo masculino.

5. SEGUNDA PROPOSTA

Para confrontar os resultados obtidos nas análises anteriores e, então, confirmar ou rejeitar a hipótese de que a expressão dos transcritos das regiões gênicas UOX, URAHP, URAD e ALLC poderia contribuir para a regulação do catabolismo das purinas e a patogênese de doenças associadas à hiperuricemia, prosseguiu-se com análises de dados de sequenciamento de RNA (RNA-seq) provenientes de estudos caso-controle com pacientes acometidos com afecções associadas à hiperuricemia.

6. MÉTODO

6.1 DisGeNET: Doenças e Genes Associados a Hiperuricemia

Primeiro, para melhor analisar as associações da hiperuricemia encontradas durante as buscas na literatura (LIU *et al.*, 2022; LÜSCHER *et al.*, 2022; MALLAT *et al.*, 2016; ROCK; KATAOKA; LAI, 2013; STAMP *et al.*, 2021; WU *et al.*, 2012; XIONG; LIU; XU, 2019), foi feita uma busca por hiperuricemia na plataforma DisGeNET (PIÑERO *et al.*, 2019) para avaliação das afecções associadas ao aumento dos níveis séricos de ácido úrico (DDAs). Em seguida, foi realizada a avaliação das associações gene-doença (GDAs) utilizando a mesma plataforma, de forma a avaliar genes que poderiam estar diferencialmente expressos em pacientes acometidos com afecções associadas à hiperuricemia.

6.2 RNA-seq: expressão diferencial da UOX, URAHP, URAD e ALLC

Após a escolha dos objetos de estudo, foi realizada a pesquisa no *Sequence Read Archive* (SRA) (LEINONEN *et al.*, 2011), banco de dados de sequências do *National Center for Biotechnology Information* (NCBI). No SRA são depositados dados brutos de sequenciamento com as informações de metadados.

As buscas foram limitadas por bibliotecas *paired-end*, nas quais a leitura da sequência de um fragmento é feita em uma certa direção até atingir um comprimento de leitura pré-definido e, em seguida, é iniciada outra rodada de leitura na extremidade oposta do fragmento. Apesar de consumir mais tempo, esta técnica permite uma melhor identificação da posição relativa das *reads* no genoma de referência, assim como um número maior de detecções quando comparada a bibliotecas *single-end* (FREEDMAN; GASPAR; SACKTON, 2020).

Além disso, foi estabelecido que deveria haver, no mínimo, 5 indivíduos no grupo caso e 5 indivíduos no grupo controle e, considerando que seria feita a análise a nível de transcrito e as regiões de interesse parecem apresentar um baixo nível de expressão, o tamanho das *reads* (≥ 100 pares de base) e a alta profundidade das bibliotecas também foram critérios de seleção.

Além dos dados brutos de RNA-seq, para realizar a quantificação a nível de transcritos é necessária a utilização de um transcriptoma de referência para identificar a região no genoma a partir da qual o fragmento de RNA é transcrito. Assim, foi realizado o download tanto dos arquivos fastq, utilizando a ferramenta *fasterq-dump* do SRA Toolkit, quanto do transcriptoma humano do GENCODE (FRANKISH *et al.*, 2021), sendo utilizada a versão 43.

Tendo os dados brutos de RNA-seq e o transcriptoma de referência, a quantificação foi realizada utilizando o Salmon (PATRO *et al.*, 2017). Este programa realiza o pseudo-alinhamento das *reads* obtidas no RNA-seq no transcriptoma de referência, sendo obtida a quantificação dos transcritos sem a geração de um arquivo de alinhamento (PATRO *et al.*, 2017). Este método permite que a quantificação seja realizada de forma mais rápida e sem um alto consumo de memória. A partir do uso do Salmon são obtidos arquivos contendo comprimentos de transcrição, abundância relativa (TPM) e contagens estimadas (número de *reads*) para cada um dos transcritos (PATRO *et al.*, 2017).

Para realizar as análises estatísticas é necessária a importação dos dados de quantificação para o RStudio (RStudio Team, 2020) e, então, a construção de matrizes contendo as contagens a nível de gene a partir destes dados, o que pode ser realizado utilizando o pacote *Tximport* (SONESON; LOVE; ROBINSON, 2015). Para gerar as matrizes, o *Tximport* precisa receber como entrada, além dos arquivos de quantificação, um arquivo contendo duas colunas, uma contendo os ID dos transcritos e outra os ID dos genes associados a cada transcrito (SONESON; LOVE; ROBINSON, 2015).

O resultado obtido será utilizado juntamente com um arquivo contendo os ID e descrições de cada amostra para realizar as análises dos dados de RNA-seq usando o DESeq2 (LOVE; HUBER; ANDERS, 2014). O DESeq2 parte de um modelo de distribuição binomial negativa para realizar a avaliação da expressão diferencial dos genes entre as amostras de interesse (LOVE; HUBER; ANDERS, 2014). Os dados acima mencionados são armazenados por um objeto da classe *DESeqDataSet* contendo as contagens das *reads* e as abundâncias relativas (LOVE; HUBER; ANDERS, 2014).

A partir daí é realizada uma pré-filtragem para remover as linhas com contagens muito baixas e reduzir o tamanho do objeto da classe *DESeqDataSet*, o que permite um aumento da velocidade do processamento das funções do DESeq2 (LOVE; HUBER;

ANDERS, 2014; ZHU; IBRAHIM; LOVE, 2019). A partir dos objetos criados é possível representar graficamente as variações na expressão dos genes entre as amostras para interpretar os resultados das quantificações.

Além da avaliação a nível de gene, foi realizada a análise a nível de transcrito. Neste caso, a importação dos dados de quantificação e a construção de matrizes contendo as contagens a nível de transcrito foi realizada por meio do pacote Tximeta (LOVE *et al.*, 2020). Neste caso não é necessário fornecer o arquivo contendo os ID dos transcritos e dos genes.

A partir das matrizes geradas foi utilizada a função Swish do pacote Fishpond (ZHU *et al.*, 2019) para ser obtido o objeto no qual estariam armazenados os dados de expressão diferencial dos transcritos. A função Swish, que parte de uma estatística não paramétrica, apresenta um maior controle da taxa de falsos positivos no processo de identificação dos transcritos diferencialmente expressos (ZHU *et al.*, 2019). Neste caso também foi aplicada uma pré-filtragem para remover as linhas com contagens muito baixas (ZHU *et al.*, 2019).

7. RESULTADOS E DISCUSSÃO

7.1 DisGeNET: Doenças e Genes Associados a Hiperuricemia

Ao realizar a busca por hiperuricemia (*Hyperuricemia*) no portal DisGeNET (PIÑERO *et al.*, 2019), os resultados das associações doença-doença (DDAs) considerando apenas fontes curadas são consistentes com os dados da literatura abordados anteriormente. Podem ser observadas as afecções mais comumente associadas a hiperuricemia, como a gota e a síndrome de Lesch-Nyhan, a associação da hiperuricemia com reação adversa a fármacos, assim como à:

- Quadros cardiovasculares, como hipertensão, isquemia e insuficiência cardíaca;
- Quadros renais, como doença renal crônica;
- Quadros pulmonares, como embolismo e tromboembolismo pulmonar.

Além disso, os resultados também indicam a associação com doenças hepáticas, doenças do sistema nervoso periférico e central, assim como diversos tipos de neoplasias.

Já no caso dos resultados das associações gene-doença (GDAs) também considerando apenas fontes curadas, foram indicados, conforme mostrado na **Tabela 3**, alguns genes envolvidos no catabolismo de purinas e outros envolvidos no transporte renal de ácido úrico, ou seja, genes envolvidos na formação e excreção do ácido úrico.

Tabela 3: Resultado do DisGeNET para os genes associados a hiperuricemia

Gene	Entrez ID	UniProt	Gene Full Name	Protein Class
XDH	7498	P47989	xanthine dehydrogenase	Enzyme
HPRT1	3251	P00492	hypoxanthine phosphoribosyltransferase 1	Enzyme
SLC2A9	56606	Q9NRM0	solute carrier family 2 member 9	Transporter
ABCG2	9429	Q9UNQ0	ATP binding cassette subfamily G member 2 (Junior blood group)	Transporter
ALDH16A1	126133	Q8IZ83	aldehyde dehydrogenase 16 family member A1	Enzyme
APLN	8862	Q9ULZ1	apelin	Signaling
SLC5A8	160728	Q8N695	solute carrier family 5 member 8	Transporter

Como mencionado anteriormente a xantina oxidase (XDH) atua no catabolismo das purinas. Em relação à aldeído desidrogenase, família 16 membro A1 (ALDH16A1), esta enzima parece não exercer atividade catalítica em mamíferos e peixes. Contudo, ela possivelmente interage com a hipoxantina guanina fosforribosiltransferase (HPRT1), enzima que também atua na via de metabolismo das purinas. Um ponto importante a ser mencionado é que um estudo conduzido na Islândia indicou uma associação entre uma mutação no gene da ALDH16A1 e níveis elevados de ácido úrico e ocorrência de gota (VASILIOU *et al.*, 2013).

O cassete de ligação a ATP, membro G da subfamília 2 (ABCG2), é uma proteína que atua na excreção renal e extrarrenal de urato (THE UNIPROT CONSORTIUM *et al.*, 2023). O transportador de soluto, família 2 membro 9 (SLC2A9), realiza o transporte de

urato, podendo ter um papel na reabsorção de urato nos túbulos proximais (THE UNIPROT CONSORTIUM *et al.*, 2023).

Os outros genes, aparentemente, não estão diretamente relacionados com a formação e a excreção do ácido úrico. O transportador de soluto, família 5 membro 8 (SLC5A8), atua no transporte de ácidos graxos de cadeia curta, fármacos e corpos cetônicos dependente de sódio (Na⁺) e cloreto (Cl⁻) (THE UNIPROT CONSORTIUM *et al.*, 2023). Já o gene APLN codifica uma pré-pró-proteína que é clivada em fragmentos peptídicos biologicamente ativos, que são ligantes endógenos do receptor de apelinina acoplado à proteína G (APLNR), presente em vários tecidos, capazes de regular diversas funções biológicas, incluindo homeostase de fluidos, função cardiovascular e secreção de insulina (THE UNIPROT CONSORTIUM *et al.*, 2023).

7.2 RNA-Seq: expressão diferencial da UOX, URAHP, URAD e ALLC em patologias associadas a hiperuricemia

Considerando os resultados do DisGeNET (PIÑERO *et al.*, 2019), assim como o estudo de Li e Li (LI; LI, 2020), após a análise de diversos projetos nos quais foram realizados estudos caso-controle com pacientes acometidos com doenças pulmonares e renais, pré-eclâmpsia, assim como outras condições cardiovasculares, foram selecionados quatro projetos, indicados na **Tabela 4**.

Na busca por trabalhos envolvendo pré-eclâmpsia foi encontrado o projeto do SRA PRJNA623568 de Yang *et al.*, 2020, que avaliou o transcriptoma do tecido placentário de pacientes com pré-eclâmpsia grave de início precoce com intuito de determinar se há alguma associação entre processos de edição de RNA e a ocorrência de complicações na gravidez (YANG *et al.*, 2020).

Outro projeto do SRA avaliado foi o PRJEB25002, no qual os pesquisadores buscaram compreender melhor a relação entre o nível de expressão gênica e a tradução de proteínas na placenta humana, avaliando como a modificação N6-metiladenosina, um mecanismo de regulação pós-transcricional do RNA, afeta a expressão gênica na placenta de mulheres com pré-eclâmpsia.

Já ao buscar por trabalhos de hipertensão pulmonar, foram encontrados estudos como o de Reyes-Palomares *et al.*, 2020, no qual foi analisado o transcriptoma de células

endoteliais das artérias pulmonares de pacientes com hipertensão pulmonar para melhor entender as alterações que poderiam contribuir para a disfunção endotelial observada neste quadro patológico (REYES-PALOMARES *et al.*, 2020). Os dados de RNA-Seq deste estudo estão disponíveis no SRA no projeto PRJNA521416.

Por fim, ao buscar os trabalhos de doença renal crônica foi encontrado o trabalho de Brightwell *et al.*, 2021. Os autores avaliaram as associações da taxa estimada de filtração glomerular, utilizada para medir a função renal, com medidas de fibrose no tecido musculoesquelético (BRIGHTWELL *et al.*, 2021). Para isso, os autores selecionaram indivíduos com doença renal crônica, casos, e controles saudáveis e avaliaram as alterações no tecido musculoesquelético utilizando análises transcriptômicas. Os dados de RNA-Seq deste estudo estão disponíveis no SRA no projeto PRJNA662524.

Tabela 4: *Projetos do SRA selecionados para a análise de transcriptoma*

Projeto	Condição	Bibliotecas (casos x controles)	Profundidade	Tamanho das leituras	Plataforma de sequenciamento
PRJNA623568	Pré-eclâmpsia	43 (9 x 32)	~25M	125 bp	Illumina
PRJEB25002	Pré-eclâmpsia	38 (6 x 28)	~30M	100 bp	Illumina
PRJNA521416	Hipertensão pulmonar	89 (7 x 9)	~15M	100 bp	Illumina
PRJNA662524	Doença renal crônica	18 (7 x 7)	~28M	150 bp	Illumina

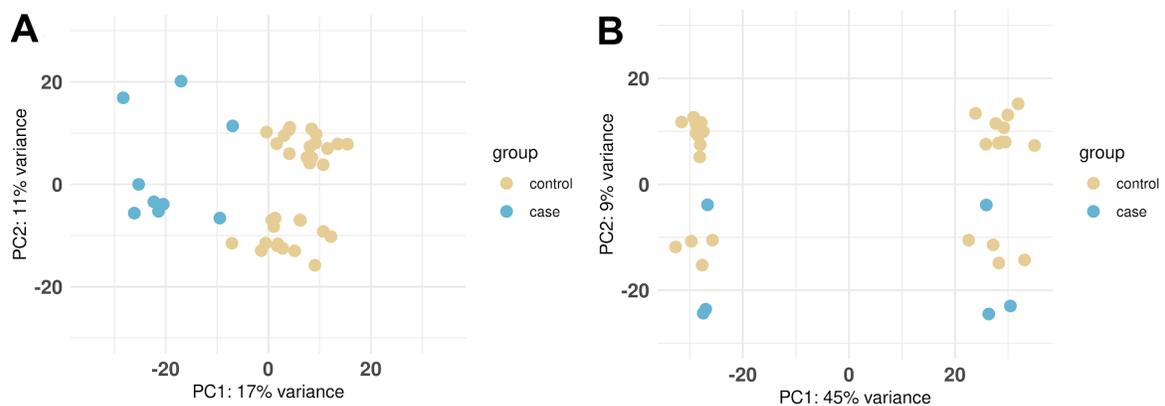
Algumas bibliotecas de cada projeto não foram incluídas nas análises. Nos projetos de pré-eclâmpsia foram utilizadas para análise apenas as bibliotecas construídas com amostra de placenta, não sendo consideradas as amostras de tipos celulares e tecidos diferentes. No projeto de hipertensão pulmonar não foram incluídos dois casos que apresentavam uma mutação que está associada ao quadro de hipertensão pulmonar familiar (REYES-PALOMARES *et al.*, 2020), sendo analisados apenas os casos com hipertensão pulmonar idiopática. E, por fim, na biblioteca de doença renal não foram consideradas as bibliotecas dos casos após serem submetidos a diálise.

Além disso, vale ressaltar que durante o processamento dos dados para obtenção dos genes e transcritos diferencialmente expressos foram aplicados processos de filtragem que incluíam a remoção de regiões gênicas que apresentavam contagens de *reads* mapeadas

muito baixas (< 10 reads). Esse processo auxilia a minimizar a interferência de ruídos experimentais e, conseqüentemente, reduzir a taxa de falsos positivos (LOVE; HUBER; ANDERS, 2014; ZHU *et al.*, 2019; ZHU; IBRAHIM; LOVE, 2019).

Primeiro, foram construídos gráficos PCA, que reduzem o número de dimensões de um banco de dados. No caso da visualização de dados bulk RNA-Seq, parte-se de uma matriz com N amostras e M valores de expressão para cada gene ou isoforma de transcrito. São, então, definidos valores de componente principal 1 (PC1) e 2 (PC2), novas variáveis que representam as direções de maior variabilidade dos dados. Em seguida, o PC1 e o PC2 de todos os genes ou isoformas de transcrito são calculados. Esses valores são utilizados para determinar os PC1 e PC2 globais de cada amostra, permitindo que cada amostra seja representada como um ponto em um gráfico bidimensional (MIGENDA; MÖLLER; SCHENCK, 2021).

Considerando que as bibliotecas foram construídas utilizando tecidos de pacientes e que os bancos de dados originais, cuja dimensionalidade foi reduzida, incluíam dados de expressão de milhares de regiões gênicas, nem sempre haverá uma boa separação entre os grupos caso e controle no PCA, como pode ser observado na **Figura 10**. Notavelmente, para as amostras dos projetos PRJNA623568 e PRJNA662524 é possível visualizar um agrupamento mais claro das amostras dos grupos caso e controle.



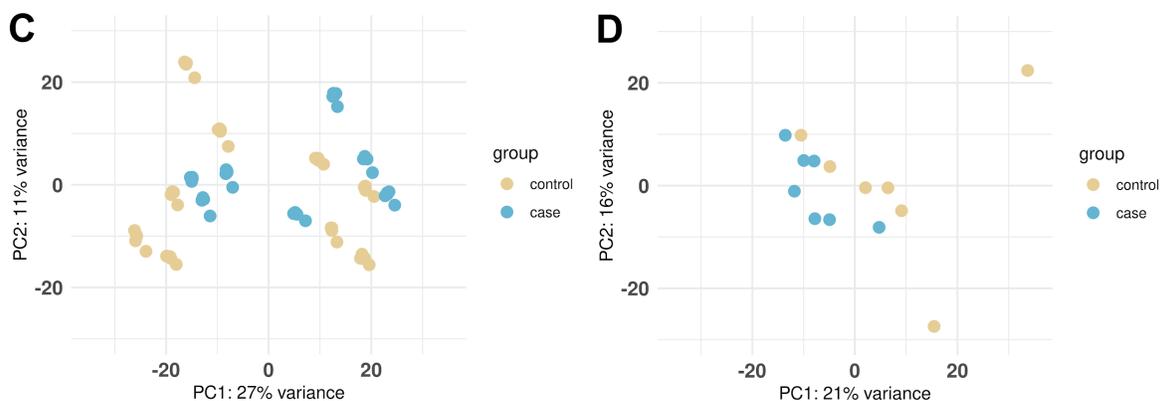
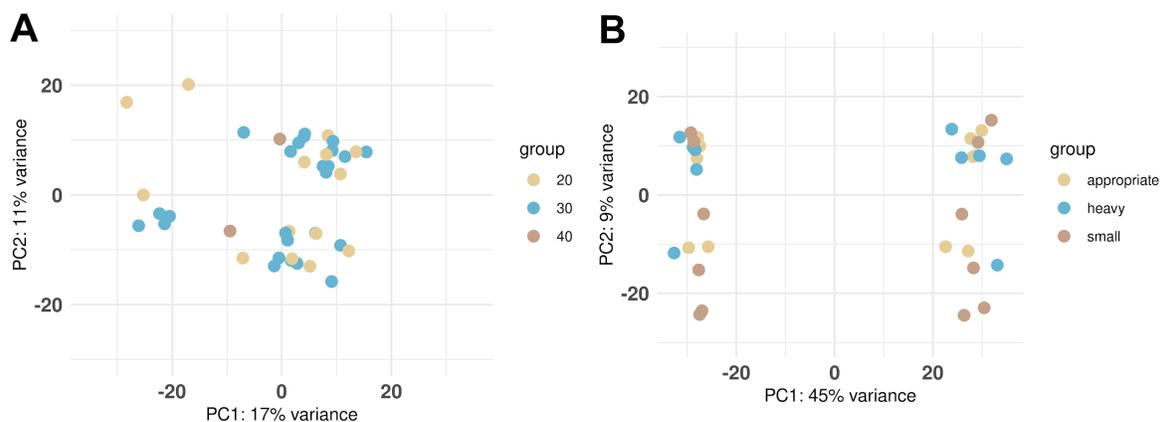


Figura 10. PCA dos casos e controles dos projetos selecionados do SRA. Gráficos de análise de componentes principais para os projetos A) PRJNA623568, B) PRJEB25002, C) PRJNA521416 e D) PRJNA662524, sendo indicados os indivíduos do grupo caso (azul) e controle (amarelo).

Considerando que a prevalência e a severidade de diversas afecções sofrem grande influência de fatores como idade, gênero e etnia (CULLEN *et al.*, 2022; JIANG; HOLMES; MCVEAN, 2021; MONTAGNANA *et al.*, 2010), foram construídos PCA para avaliar se estes fatores estariam provocando a clusterização dos dados analisados, o que poderia levar a incorporação de um viés, ou seja, um erro sistemático, no estudo.

Como esta análise depende do fornecimento, por parte dos executores dos projetos do SRA, de arquivos de metadados, contendo as informações pertinentes de cada paciente que cedeu a amostra para construção da biblioteca, não foi possível obter as informações completas. Contudo, ainda assim foi possível construir os PCA para alguns destes fatores e, ao avaliar os gráficos resultantes, mostrados na **Figura 11**, não foi observada a clusterização das amostras de acordo com as variáveis de confusão.



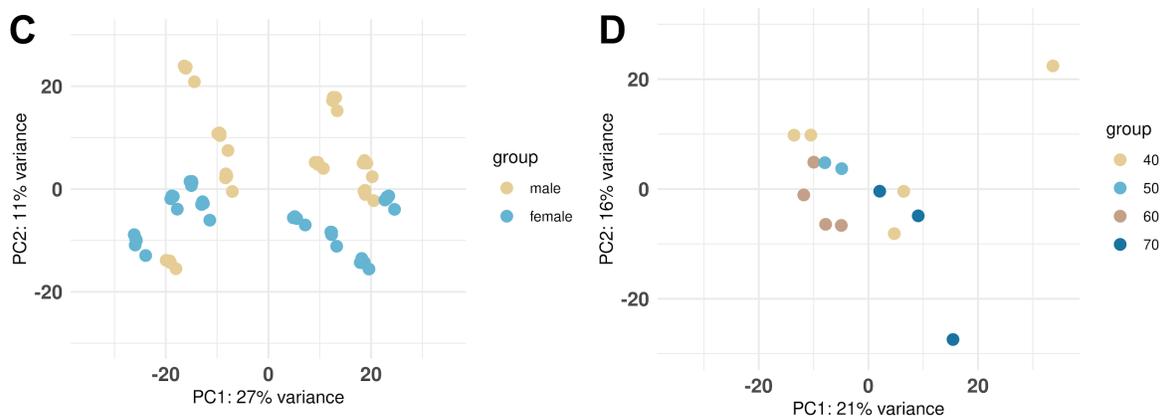


Figura 11. PCA das variáveis de confusão dos projetos selecionados do SRA. Gráficos de análise de componentes principais para os projetos A) PRJNA623568, B) PRJEB25002, C) PRJNA521416, e D) PRJNA662524, sendo indicados os indivíduos de acordo com a idade, o tamanho do feto, o gênero e novamente a idade, respectivamente.

Vale ressaltar que para todos os pacientes do grupo caso do projeto PRJEB25002, ao avaliar o crescimento fetal, foi observado que os fetos apresentavam um tamanho reduzido em relação ao que é esperado, seguindo as curvas de crescimento fetal. Este resultado já era esperado, uma vez que a pré-eclâmpsia está fortemente associada à restrição de crescimento fetal (FGR, do inglês *Fetal Growth Restriction*) DE SOUZA RUGOLO; BENTLIN; TRINDADE, 2011; TAKAHASHI *et al.*, 2021).

Para a avaliação da expressão diferencial foram definidos pontos de corte para o valor-p (SAMUELS; WITMER; SCHAFFNER, 2016) e para o log2 fold change (log2FC) (JIANG *et al.*, 2024). O valor-p é um dos componentes de um teste de hipóteses (SAMUELS; WITMER; SCHAFFNER, 2016), neste cenário podemos considerar as seguintes hipóteses:

- H_0 : os níveis de expressão da região gênica URAHP são iguais entre os grupos caso e controle, ou seja, não há expressão diferencial da URAHP;

- H_1 : os níveis de expressão da região gênica URAHP são diferentes entre os grupos caso e controle, ou seja, há expressão diferencial da URAHP.

O valor-p representa a probabilidade de ser obtido um valor de estatística de teste tão extremo quanto o que seria obtido com os dados amostrais caso H_0 seja verdadeira (não há expressão diferencial da URAHP) (SAMUELS; WITMER; SCHAFFNER, 2016). Para aceitar ou rejeitar H_0 , o valor p é comparado com o nível de significância (α) estabelecido,

sendo este o valor máximo tolerado para taxa de falsos positivos (SAMUELS; WITMER; SCHAFFNER, 2016). Assim, se o valor-p for inferior a α haverá evidências suficientes para rejeitar H_0 (há expressão diferencial da URAHP) (SAMUELS; WITMER; SCHAFFNER, 2016).

O log2 fold change (log2FC) é uma métrica muito utilizada em estudos de RNA-Seq, principalmente quando busca-se quantificar a mudança na expressão gênica entre duas condições diferentes (JIANG *et al.*, 2024). O fold change (FC) é uma razão que indica quanto a expressão de um gene mudou entre duas condições (JIANG *et al.*, 2024). Se E1 é a expressão da URAHP no grupo caso e E2 é a expressão da URAHP no grupo controle, o FC é calculado como $FC = E1 / E2$. A aplicação do logaritmo de base 2 ao fold change é realizada, pois, como dito anteriormente, os dados de expressão gênica podem variar em várias ordens de magnitude, assim a transformação logarítmica ajuda a tornar os dados mais comparáveis (JIANG *et al.*, 2024).

Para este estudo o α foi estabelecido em 0,05, ou seja, o valor máximo de falsos positivos aceito foi 5% e, conseqüentemente, a expressão diferencial de um gene ou transcrito era considerada significativa com p-valor $< 0,05$. Já para o log2FC foi estabelecido um ponto de corte em log2FC absoluto $> 0,5$, sendo, então, selecionados os genes e transcritos que apresentaram um aumento relativo de 0,5 ou uma redução relativa de -0,5 na expressão ao comparar os casos com os controles.

Neste caso foi aplicado um valor de log2FC absoluto inferior ao que geralmente é observado na literatura, pois transcritos não produtivos, como lncRNA, possuem um nível de expressão mais baixo se comparado a genes codificadores de proteínas. Além disso, como visto nos resultados do GTEx, mesmo para as regiões gênicas URAD e ALLC, que expressam transcritos produtivos, o nível de expressão parece ser bem reduzido.

Considerando estes pontos de corte observados 1162, 909, 313 e 357 genes superexpressos nos casos, em relação aos controles, nos projetos PRJNA623568, PRJEB25002, PRJNA521416 e PRJNA662524, respectivamente. No caso dos genes subexpressos foram observados 435, 246, 236 e 386 genes nos projetos PRJNA623568, PRJEB25002, PRJNA521416 e PRJNA662524, respectivamente. Este resultado está representado, como gráficos de barra, na **Figura 12**.

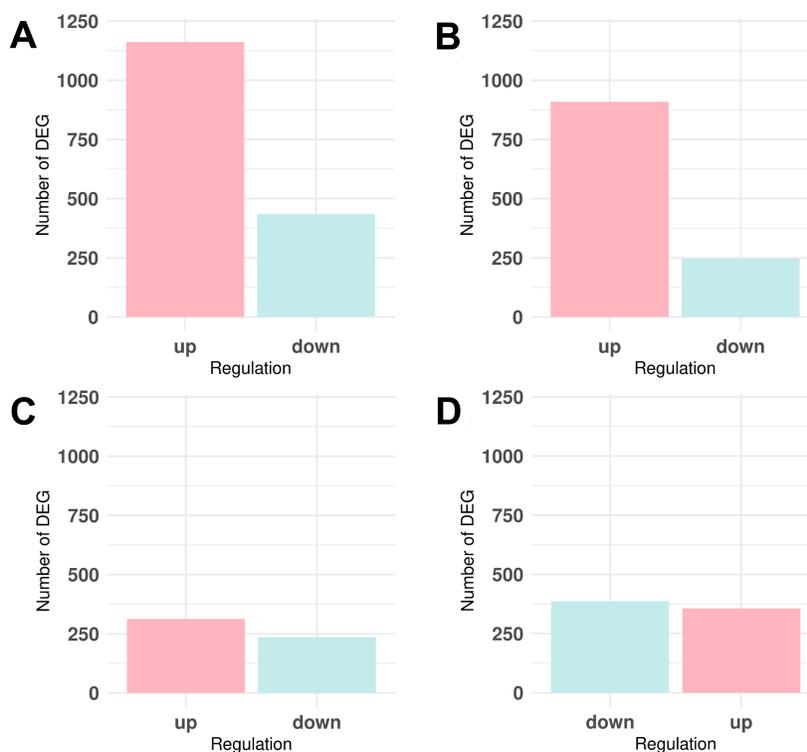


Figura 12. *Expressão diferencial a nível de gene dos projetos selecionados do SRA. Gráficos de barras indicando o número de genes superexpressos (up) e subexpressos (down) nos projetos A) PRJNA623568, B) PRJEB25002, C) PRJNA521416, e D) PRJNA662524. Neste caso estão incluídos os genes que passaram pelos pontos de corte p -valor $< 0,05$ e \log_2FC absoluto $> 0,5$.*

Já a nível de transcrito, foram observados 8222, 4462, 1308 e 6489 transcritos diferencialmente expressos ao comparar os casos com controles nos projetos PRJNA623568, PRJEB25002, PRJNA521416 e PRJNA662524, respectivamente.

Para todas as bibliotecas foi possível notar que entre os transcritos não produtivos, os que apresentam um maior nível de expressão diferencial são lncRNA, íntron retido, codificante de proteína com CDS não definida e decaimento mediado por nonsense, sendo esses últimos uma classe de transcritos que apresenta um códon de parada prematuro. O perfil de transcritos superexpressos e subexpressos, de acordo com o tipo de transcrito das classes mencionadas, está indicado na **Figura 13**.

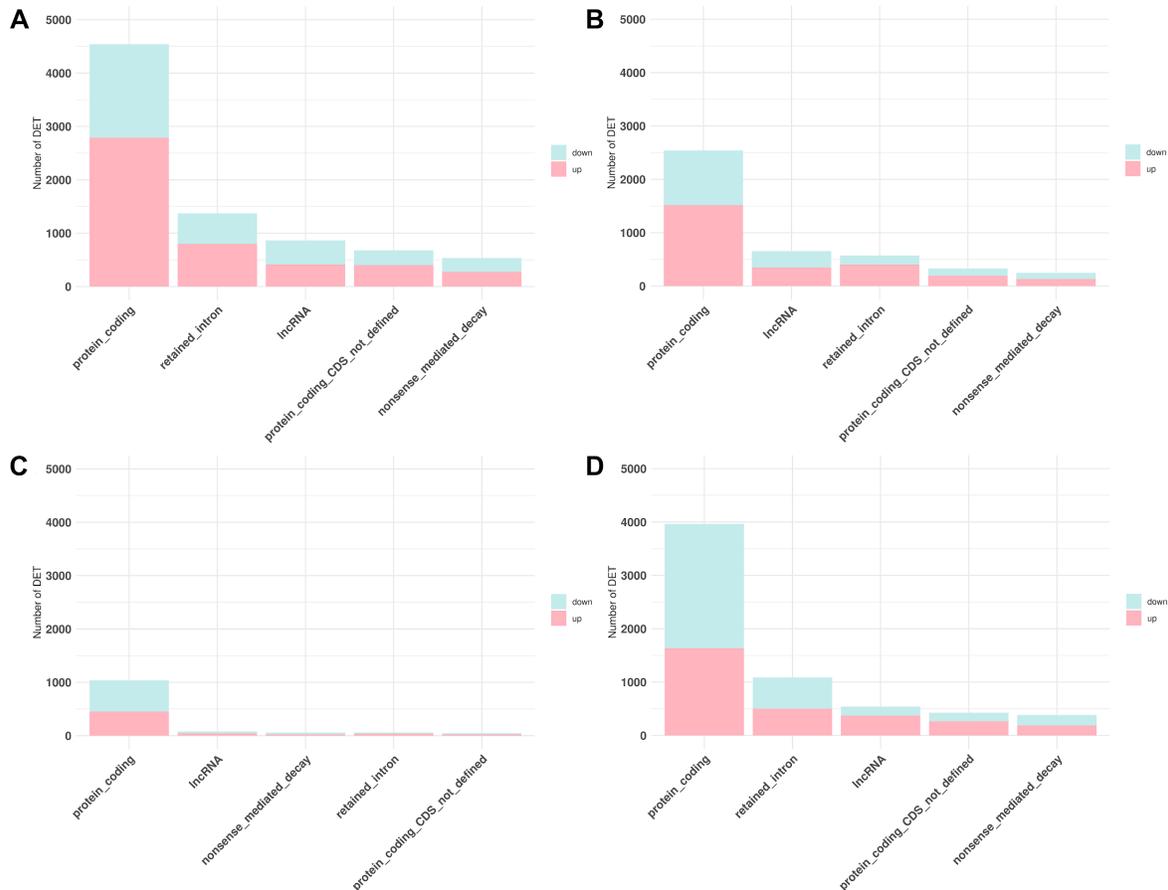


Figura 13. Expressão diferencial a nível de transcrito dos projetos selecionados do SRA. Gráficos de barras indicando o número de transcrito superexpressos (up) e subexpressos (down) nos projetos A) PRJNA623568, B) PRJEB25002, C) PRJNA521416, e D) PRJNA662524 para os principais tipos de transcrito diferencialmente expressos (codificante de proteína, lncRNA, íntron retido, codificante de proteína com CDS não definida e decaimento mediado por nonsense). Neste caso estão incluídos os transcritos que passaram pelos pontos de corte p -valor $< 0,05$ e \log_2FC absoluto $> 0,5$.

É possível notar que para as amostras do experimento caso controle envolvendo hipertensão pulmonar (PRJNA521416) foi observado um número mais reduzido de transcritos diferencialmente expressos, o que é devido, em parte, ao fato de apresentar uma profundidade mais reduzida (~15M) quando comparado com os outros experimentos.

Apesar de no trabalho de Li e Li ter sido observado um aumento significativo na expressão do URAHP em gestantes com pré-eclâmpsia (LI; LI, 2020), o mesmo resultado não foi obtido na análise dos dois projetos do SRA cuja afecção de interesse era a pré-eclâmpsia, PRJNA623568 e PRJEB25002. Além disso, também não foi observada alteração na expressão das outras regiões gênicas, UOX, URAD e ALLC, ao comparar o grupo caso com o grupo controle.

O mesmo resultado foi obtido para os projetos envolvendo hipertensão pulmonar (PRJNA521416) e doença renal crônica (PRJNA662524), não tendo sido observada alteração significativa na expressão das regiões UOX, URAHP, URAD e ALLC. Um dos fatores que influencia esse resultado é o número de contagens das *reads* mapeadas nestas regiões, o qual está abaixo, ou muito próximo, do ponto de corte de 10 *reads* utilizado para remoção de ruídos. Na **Tabela 5** estão indicadas as contagens totais, ou seja, considerando todas as amostras, para cada uma das regiões gênicas.

Tabela 5: Total de contagens das *reads* e valores de \log_2FC e *p*-valor obtidos na análise da expressão diferencial para UOX, URAHP, URAD e ALLC

PRJNA623568				
Ensembl ID	Gene	Total de contagens	\log_2FC	pvalue
ENSG00000151360.10	ALLC	46.632063	1.030013e-06	0.4871923
ENSG00000183463.6	URAD	3.949386	-	-
ENSG00000222019.8	URAHP	0.000000	-	-
ENSG00000290738.1	URAHP	430.824782	6.021042e-06	0.1312368
ENSG00000240520.7	UOX	18.373776	-4.480376e-07	0.535796
PRJEB25002				
Ensembl ID	Gene	Total de contagens	\log_2FC	pvalue
ENSG00000151360.10	ALLC	52.889564	-2.344626e-06	0.005873798
ENSG00000183463.6	URAD	6.968304	-	-
ENSG00000222019.8	URAHP	0.000000	-	-
ENSG00000290738.1	URAHP	52.285512	6.87446e-08	0.9554222
ENSG00000240520.7	UOX	5.578734	-	-
PRJNA521416				
Ensembl ID	Gene	Total de contagens	\log_2FC	pvalue
ENSG00000151360.10	ALLC	1.799642	-	-
ENSG00000183463.6	URAD	0.000000	-	-
ENSG00000222019.8	URAHP	8.978051	-	-
ENSG00000290738.1	URAHP	246.217673	-0.01757187	0.4600766

ENSG00000240520.7	UOX	209.165494	-0.009012967	0.6422127
PRJNA662524				
Ensembl ID	Gene	Total de contagens	log2FC	pvalue
ENSG00000151360.10	ALLC	35.3171624	-7.848799e-07	0.3188277
ENSG00000183463.6	URAD	0.9956786	-	-
ENSG00000222019.8	URAHP	0.000000	-	-
ENSG00000290738.1	URAHP	221.0117750	1.582583e-06	0.5531198
ENSG00000240520.7	UOX	4.7571163	-	-

Desta forma, utilizando estes bancos de dados, não foi observada uma expressão diferencial das regiões gênicas de interesse. Contudo, não é possível afirmar com certeza que a UOX, URAHP, URAD e ALLC não exercem papel regulatório em afecções associadas ao acúmulo de ácido úrico, pois este resultado pode ser influenciado pelo fato das bibliotecas terem sido construídas a partir de tecidos, nos quais há diversos tipos celulares distintos, assim como pela profundidade das bibliotecas.

Apesar da hipótese ter se provado nula, ainda foi realizada uma busca pelos genes e transcritos da via de degradação das purinas, além dos genes e transcritos de transportadores envolvidos na excreção do ácido úrico para investigar um possível papel regulatório destes nas afecções associadas a hiperuricemia. Foram considerados alguns dos genes associados a hiperuricemia resultantes da pesquisa no DisGeNET (PIÑERO *et al.*, 2019), mas também foram consideradas outras enzimas da via de degradação de purinas.

No caso dos genes e transcritos envolvidos no catabolismo de purinas, como guanina desaminase (GDA), adenina deaminase (ADA) e 5' nucleotidases (NT5E, NT5C2), não é possível observar um padrão claro que indicaria que esta via está sendo regulada. Como ilustrado na **Figura 14**, alguns transcritos estão sub-expressos, outros superexpressos e para muitos deles a expressão diferencial não é significativa, ou seja, não passa dos pontos de corte de p-valor ($< 0,05$) e/ou log2FC ($> 0,5$ ou $< -0,5$).

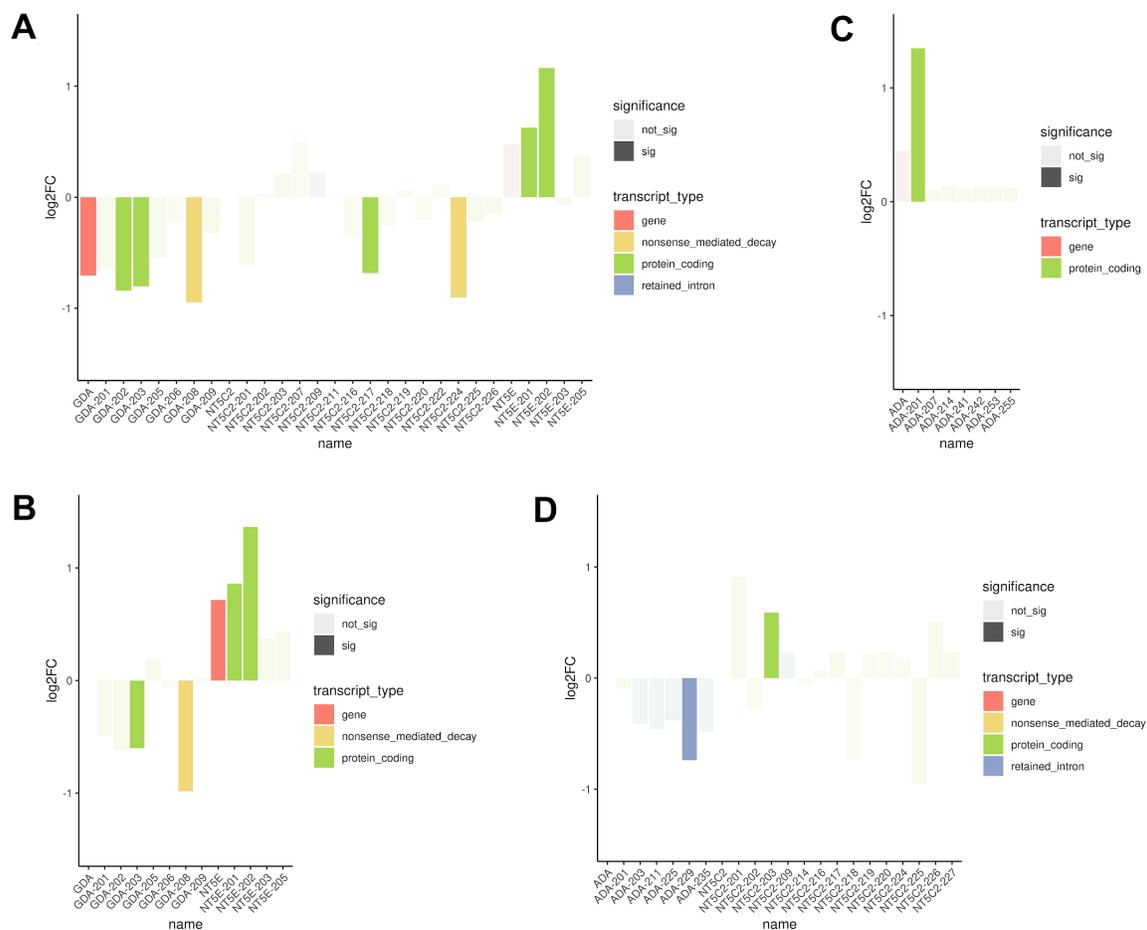


Figura 14. Expressão diferencial dos genes e transcritos da via de degradação das purinas. Gráfico de barras indicando a expressão diferencial em \log_2FC de alguns genes via de degradação de purinas e seus transcritos nos projetos A) PRJNA623568, B) PRJEB25002, C) PRJNA521416, e D) PRJNA662524. No eixo X estão indicadas as regiões gênicas (ADA - adenina deaminase, GDA - guanina deaminase e NTC - 5' nucleotidase) e seus transcritos. O preenchimento das barras corresponde ao tipo de transcrito e a opacidade da cor depende da expressão diferencial ser ou não significativa (p -valor $< 0,05$ e \log_2FC absoluto $> 0,5$).

Para os genes e transcritos envolvidos na excreção do ácido úrico, como o cassette de ligação a ATP, membro G da subfamília 2 (ABCG2) e o transportador de soluto, família 2 membro 9 (SLC2A9), é obtido um resultado similar, com alguns transcritos do ABCG2 estando sub-expressos, alguns transcritos do SLC2A9 estando superexpressos e, novamente, para certos transcritos a expressão diferencial não é significativa, como demonstrado na **Figura 15**.

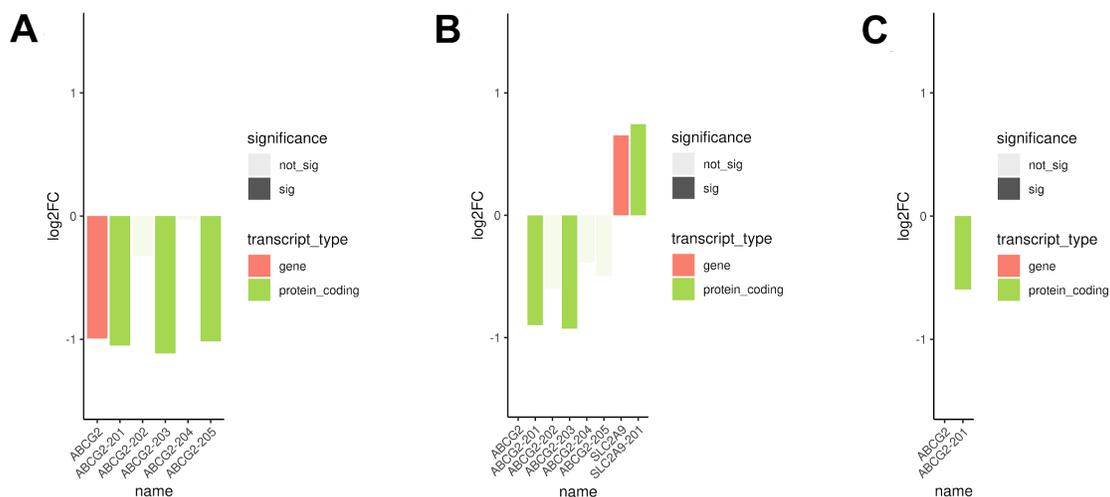


Figura 15. Expressão diferencial dos genes e transcritos envolvidos na excreção de ácido úrico. Gráfico de barras indicando a expressão diferencial em \log_2FC de alguns genes envolvidos na excreção de ácido úrico e seus transcritos nos projetos A) PRJNA623568, B) PRJEB25002 e C) PRJNA662524. No eixo X estão indicadas as regiões gênicas (ABCG2 - cassete de ligação a ATP e SLC2A9 - transportador de soluto) e seus transcritos. O preenchimento das barras corresponde ao tipo de transcrito e a opacidade da cor depende da expressão diferencial ser ou não significativa (p -valor $< 0,05$ e \log_2FC absoluto $> 0,5$).

Assim, utilizando estes bancos de dados, não foi observado uma alteração clara na expressão de genes e transcritos envolvidos no catabolismo de purinas e no transporte renal de ácido úrico, não sendo possível, assim, validar a hipótese de que regiões gênicas envolvidos na formação e excreção do ácido úrico poderiam estar de alguma forma exercendo um papel regulatório nas afecções associadas ao acúmulo de ácido úrico.

8. TERCEIRA PROPOSTA

Tendo a hipótese da análise anterior se mostrado nula e considerando a potencial presença dos produtos proteicos da URAD e da ALLC em bancos de dados de proteoma, que sugere que houve alguma pressão para manter a integridade dessas regiões gênicas, partiu-se para a investigação dos transcritos codificantes de proteína. Assim, por meio da análise da estrutura primária, da modelagem e avaliação da estrutura terciária e da expressão da proteína recombinante, buscou-se caracterizar os produtos proteicos da URAD e da ALLC.

9. MATERIAIS E MÉTODOS

9.1 Avaliação da estrutura primária dos produtos proteicos da URAD e da ALLC

Para os transcritos produtivos foi realizada a avaliação da estrutura primária, ou seja, da sequência de resíduos de aminoácidos das cadeias polipeptídicas que seriam geradas a partir da tradução dos mesmos.

Primeiro, foi feito o alinhamento da sequência de aminoácidos das proteínas hipotéticas de *Homo sapiens* com a sequência de proteínas que foram resolvidas ou calculadas experimentalmente. Esta análise foi conduzida utilizando o T-Coffee (ClustalW - Clustal2.1), uma ferramenta de alinhamento de sequências múltiplas (MSA, do inglês Multiple Sequence Alignments) (NOTREDAME; HIGGINS; HERINGA, 2000). Ao inserir as sequências das proteínas de interesse nas ferramentas de MSA, obtém-se como resultado o alinhamento das cadeias polipeptídicas, sendo que os resíduos alinhados em uma determinada coluna são estruturalmente e/ou funcionalmente homólogos (NOTREDAME; HIGGINS; HERINGA, 2000).

Assim, dependendo da variabilidade de resíduos em uma determinada coluna, pode-se inferir as posições que são: 1) invariantes, ou seja, que apresentam um resíduo de aminoácido cujas propriedades químicas e estéricas são essenciais para estabilidade da proteína; 2) conservadas, podendo haver uma variação do resíduo de aminoácido presente nesta posição contanto que o novo resíduo seja similar ao que se fazia presente (por exemplo, a substituição de um aspartato por um glutamato); e hipervariáveis, ou seja, posições cujos requerimentos químicos e estéricos são inespecíficos, podendo comportar resíduos de aminoácidos com características distintas entre si (NELSON; COX; LEHNINGER, 2017; VOET; VOET, 2011).

Para as primeiras análises foram selecionadas sequências de aminoácido das enzimas de OHCU descarboxilase e de alantoicase cujas estruturas já haviam sido resolvidas ou calculadas, por meio de técnicas experimentais, para realizar os alinhamentos com os produtos proteicos da URAD e da ALLC, respectivamente. Assim, foi feita uma busca no PDB (BERMAN, 2000) pela estrutura primária e terciária destas enzimas.

Em seguida, foi realizada uma análise usando PFstats (FONSECA-JÚNIOR *et al.*, 2018) para analisar a conservação posicional e redes de coevolução de resíduos, permitindo a identificação de grupos de resíduos estrutural e funcionalmente importantes. As sequências de aminoácido das enzimas OHCU descarboxilase e alantoicase utilizadas para o alinhamento foram obtidas por meio do Pfam Legacy (MISTRY *et al.*, 2021). Para minimizar alguns problemas, como a inclusão de fragmentos e de cadeias polipeptídicas repetidas, as sequências passaram por uma filtragem, de forma que apenas a que apresentassem uma cobertura mínima de 80% e uma identidade máxima de 80% com as sequências selecionadas como referência, fossem selecionadas (FONSECA-JÚNIOR *et al.*, 2018).

Além dessa filtragem, para realizar as análises no PFstats foi necessário definir as sequências que seriam utilizadas como referência e a frequência mínima que um resíduo deve aparecer para que seja considerado conservado, tendo sido utilizado o valor de 70% (FONSECA-JÚNIOR *et al.*, 2018). Ademais, é definido o valor de sub-alinhamento representativo mínimo (MRsA), que indica o número mínimo do sub-alinhamento no qual é obtido um resultado representativo do original, utilizando o alinhamento completo (FONSECA-JÚNIOR *et al.*, 2018). Este valor, juntamente com outros parâmetros, é utilizado para construir a rede de correlação.

9.2 Predição e análise da estrutura terciária do produto proteico da URAD

Em seguida, foi realizada a modelagem do produto proteico da URAD, ou seja, a OHCU descarboxilase de *H. sapiens*. Primeiramente foi utilizado o AlphaFold por meio do Google Colab (MIRDITA *et al.*, 2022), o ColabFold, que foi desenvolvido com intuito de democratizar a modelagem de estruturas proteicas, visto que permite que usuários executem, por meio de códigos em Python e diretamente no navegador, as ferramentas avançadas de previsão de estrutura criadas pela DeepMind (MIRDITA *et al.*, 2022).

Para gerar as estruturas pelo ColabFold foi utilizada a metodologia MMseqs2 (*Many-against-Many sequence searching*), um programa que realiza a busca e a comparação rápida de sequências de proteínas em grandes bancos de dados, facilitando a identificação de sequências similares e homólogas, o que auxilia no refinamento das previsões das estruturas (MIRDITA; STEINEGGER; SÖDING, 2019). Foram gerados cinco modelos, os quais foram minimizados utilizando o campo de força AMBER99SB. Os

modelos gerados foram classificados por seus valores de pLDDT (*predicted Local Distance Difference Test*), uma métrica que fornece uma estimativa da confiabilidade de cada resíduo de aminoácido na estrutura prevista, sendo o valor calculado com base na predição da distância entre átomos na estrutura da proteína (JUMPER *et al.*, 2021).

Em seguida, foi utilizada a modelagem por homologia no Modeller, por meio da qual são obtidas as coordenadas 3D de uma proteína de interesse cuja sequência de aminoácidos é conhecida, mas a estrutura tridimensional é desconhecida baseando-se no alinhamento da sequência desta proteína de interesse com estruturas resolvidas ou calculadas experimentalmente, sendo utilizadas, geralmente, estruturas de proteínas homólogas (ŠALI; BLUNDELL, 1993). Os *scripts* do Modeller, utilizados para alinhar as sequências e gerar as estruturas, foram rodados no terminal do NMRBox¹ (MACIEJEWSKI *et al.*, 2017).

Foram gerados 3000 modelos, sendo eles organizados de acordo com sua pontuação DOPE (*Discrete Optimized Protein Energy*). A pontuação DOPE corresponde à soma de energias de pares de átomos, onde cada termo da soma é baseado na distância entre os átomos e a probabilidade de encontrar as distâncias observadas, considerando os dados da estrutura conhecida (SHEN; SALI, 2006). Assim, valores mais baixos de DOPE indicam modelos de proteína com maior probabilidade de serem estruturalmente corretos e semelhantes à estrutura nativa (SHEN; SALI, 2006). Foi selecionado o modelo de menor energia, sendo este minimizado de acordo com sua energia global no programa Chimera (PETTERSEN *et al.*, 2004), utilizando o campo de força AMBER ff14SB. Os parâmetros foram: 200 passos de descida com tamanho 0,02 Å e 20 passos de gradiente conjugado com tamanho 0,02 Å.

Além disso, foi realizada a modelagem por *threading* por meio do iTASSER (*Iterative Threading ASSEmbly Refinement*) (YANG *et al.*, 2015). Nesta metodologia, o iTASSER começa procurando, em bancos de dados como o PDB (BERMAN, 2000), por estruturas de proteínas que sejam semelhantes à sequência da proteína de interesse. A sequência de interesse é alinhada com as sequências das proteínas do banco de dados usando métodos de *threading*, que avalia se a sequência da proteína de interesse pode se ajustar a uma estrutura tridimensional conhecida (YANG *et al.*, 2015). Usando estes

¹ NMRbox é um recurso para programas de RMN (Ressonância Magnética Nuclear) biomolecular. Ele fornece ferramentas para encontrar o programa necessário, documentação e tutoriais para aproveitar ao máximo cada ferramenta, além de máquinas virtuais baseadas em nuvem para a execução dos programas em questão.

alinhamentos de threading, o iTASSER constroi modelos iniciais baseados nas coordenadas atômicas das proteínas do banco e, em seguida, utiliza uma abordagem iterativa, com fragmentos da estrutura das proteínas, para otimizar a conformação local e global da proteína (YANG *et al.*, 2015). O iTASSER gera múltiplos modelos, sendo selecionados os 5 melhores, os quais são avaliados usando pontuações baseadas em energia e em similaridade com estruturas conhecidas, incluindo o TM-score (do inglês, *Template Modeling score*) e C-score (do inglês, *Confidence score*), que indicam a confiabilidade e a qualidade dos modelos gerados. (YANG *et al.*, 2015).

As estruturas de OHCU descarboxilase de *H. sapiens* que apresentaram uma melhor pontuação foram analisadas e validadas usando o servidor SAVES (*Structural Analysis and Verification Server*), que fornece uma plataforma com ferramentas que podem ser usadas para avaliar a qualidade de modelos de proteínas. Neste caso, foi utilizado o ERRAT (COLOVOS; YEATES, 1993), que analisa a estrutura tridimensional de uma proteína para identificar regiões que possivelmente contêm erros nas suas interações não covalentes, e o PROCHECK (LASKOWSKI *et al.*, 1993), por meio do qual é possível avaliar parâmetros estereoquímicos utilizando, principalmente, o gráfico de Ramachandran, que indica se os ângulos diédricos ϕ (phi) e ψ (psi) dos resíduos de aminoácidos estão em regiões favoráveis (RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963). Como forma de comparação também foram avaliados os resultados do PROCHECK e do gráfico de Ramachandran para uma estrutura de OHCU descarboxilase cuja estrutura já está depositada no PDB.

9.3 Expressão e purificação do produto proteico da URAD

Para testar a hipótese de que a partir da expressão da região gênica URAD poderia ser produzida uma enzima funcional, foi realizada a aquisição do plasmídeo (FastBio) no qual o gene de interesse foi clonado no vetor pET-28a(+)-TEV usando os sítios de restrição *NdeI* e *XhoI*, de modo que a proteína expressa apresentasse, no N-terminal uma His-Tag e um sítio de clivagem da protease TEV (MGHHHHHHENLYFQGH). O vetor pET-28a(+)-TEV apresenta o gene de resistência à canamicina, o que permite o uso do antibiótico em meio de cultivo para selecionar as células que foram transformadas.

9.3.1 Multiplicação do plasmídeo em DH5 α e verificação da presença do inserto

Para aumentar a quantidade de plasmídeo, *Escherichia coli* DH5 α foi transformada por meio de choque térmico. Um pré-inóculo foi preparado utilizando uma colônia isolada de DH5 α , que foi inoculada em 10 mL de meio Luria Bertani (LB). O pré-inóculo foi incubado com agitação a 37°C durante a noite (16 horas, aproximadamente).

No dia seguinte, a inoculação foi realizada adicionando 1 mL de pré-inóculo a 50 mL de meio LB. O inóculo foi incubado com agitação a 37°C até atingir uma densidade óptica (DO) de 0,3. As células foram, então, resfriadas em gelo por 1 hora e, em seguida, foi realizada uma centrifugação a 3000 \times g por 15 minutos a 4°C. Após a centrifugação, as células foram ressuspensas em 10 mL de solução estéril de cloreto de cálcio (CaCl₂) 50 mM. Este processo de centrifugação e ressuspensão em 10 mL de CaCl₂ 50 mM foi repetido uma vez, seguido por outra centrifugação a 3000 \times g por 15 minutos a 4°C. Posteriormente, as células foram ressuspensas em 1 mL de CaCl₂ 50 mM e mantidas em gelo por 1 hora. Após este período, 150 μ L de células foram incubadas com 200 ng de plasmídeos contendo o gene URAD, e o sistema foi submetido a um choque térmico em um bloco de calor a 42°C por 3 minutos. Em seguida, 1 mL de meio LB foi adicionado às células transformadas, que foram incubadas com agitação a 37°C por 1 hora.

As células transformadas foram então usadas para preparar um pré-inóculo utilizando 10 mL de meio LB contendo 1 mM de canamicina. No dia seguinte 8 mL do pré-inóculo foram centrifugados a 6000 \times g por 3 minutos a 8°C. O sobrenadante foi descartado e com o pellet prosseguiu-se com o protocolo de Mini Prep. Primeiro foi adicionada ao pellet 200 μ L da solução I, um tampão contendo Tris e EDTA para ressuspender as células. Após ressuspender as células foram adicionados 200 μ L da solução II, que apresenta NaOH e SDS, agentes que irão favorecer a lise alcalina das células. Esse sistema foi misturado por inversão por 5 minutos e, em seguida, foram adicionados 350 μ L da solução III, uma solução neutralizante contendo ácido acético e acetato de potássio. O sistema foi misturado por inversão e, então, foi realizada a centrifugação a 10000 \times g por 15 minutos a 8°C.

Enquanto isso, foi realizado o preparo da coluna, sendo adicionados 500 μ L de solução de preparo de coluna e realizada a centrifugação a 12000 \times g por 1 minuto a 8°C. Com a coluna preparada, foi transferido o sobrenadante obtido após a lise das células e o sistema foi centrifugado a 6000 \times g por 1 minuto a 8°C. O *flow-through* foi descartado e

foram adicionados 750 μL de solução de lavagem, sendo realizada uma nova centrifugação a 6000 g por 1 minuto a 8°C. O *flow-through* foi novamente descartado e, então, foram adicionados 40 μL de solução de eluição e o sistema foi novamente centrifugado a 6000 g por 1 minuto a 8°C. O processo de adicionar 40 μL de solução de eluição foi repetido mais uma vez e, após uma nova centrifugação a 6000 \times g por 1 minuto a 8°C, foi realizada a quantificação do plasmídeo presente nos 80 μL de solução de eluição. Para realizar a quantificação foi utilizado o Nanodrop.

Tendo sido obtida uma maior quantidade do plasmídeo, para verificar se o gene de interesse foi inserido de forma completa no vetor, foi realizada a digestão do plasmídeo utilizando as enzimas de restrição *XhoI*, *XbaI* e *NdeI* (Promega). Foram avaliadas cinco condições, sendo três digestões simples (*XhoI* ou *XbaI* ou *NdeI*) com apenas uma das enzimas de restrição e duas digestões duplas (*XhoI* e *XbaI* ou *XhoI* e *NdeI*).

Nas cinco condições de digestão foram utilizados 2 μL de tampão D 10x (6 mM Tris-HCl, 6 mM MgCl_2 , 150 mM NaCl, 1 mM DTT, pH 7,9), tendo o volume final sido ajustado para 20 μL utilizando água ultrapura. Nas digestões simples foram utilizados 0,5 μL de *XhoI* (10 u/ μL) ou 0,5 μL de *XbaI* (10 u/ μL) ou 0,5 μL de *NdeI* (10 u/ μL) e 100 ng do plasmídeo. Nas digestões duplas foram utilizados 0,5 μL de *XhoI* (10 u/ μL) e 0,5 μL de *XbaI* (10 u/ μL) ou 0,5 μL de *XhoI* (10 u/ μL), 0,5 μL de *NdeI* (10 u/ μL) e 200 ng do plasmídeo. Como controle foi preparada uma amostra contendo apenas 2 μL de tampão D 10x e 100 ng do plasmídeo, o volume final também foi ajustado para 20 μL utilizando água ultrapura.

Todas as amostras foram mantidas no banho-maria à 37°C por 1 hora e, após esse período foram utilizadas, juntamente com tampão de amostra 5x, para aplicação no gel TAE (40 mM Tris-acetato, 1 mM EDTA) agarose 1%. Foram aplicadas as amostras e o padrão (1kb DNA Ladder RTU).

9.3.2 Transformação de *E. coli* BL21(DE3)pLys e *E. coli* ArcticExpress (DE3) para expressão da proteína recombinante

Foi realizada a transformação de *E. coli* BL21(DE3)pLys e *E. coli* ArcticExpress (DE3). Dois pré-inóculos foram preparados utilizando uma colônia de BL21 pLys e uma colônia de Arctic, que foram inoculadas em 10 mL de meio Luria Bertani (LB) contendo

1mM de clorafenicol no caso da BL21 pLys e 1 mM de gentamicina (G418) no caso da Arctic. O pré-inóculo foi incubado com agitação a 37°C durante a noite.

No dia seguinte, a inoculação foi realizada adicionando 1 mL de pré-inóculo a 50 mL de meio LB contendo 1mM de clorafenicol no caso da BL21(DE3)pLys e 1mM de gentamicina (G418) no caso da ArcticExpress (DE3).

O inóculo foi incubado com agitação a 37°C até atingir uma densidade óptica (DO) de 0,3. As células foram, então, resfriadas em gelo por 1 hora e, em seguida, foi realizada uma centrifugação a 3000×g por 15 minutos a 4°C. Após a centrifugação, as células foram ressuspensas em 10 mL de solução estéril de CaCl₂ 50 mM. Este processo de centrifugação e ressuspensão em 10 mL de CaCl₂ 50 mM foi repetido uma vez, seguido por outra centrifugação a 3000×g por 15 minutos a 4°C.

Posteriormente, as células foram ressuspensas em 1 mL de CaCl₂ 50mM e mantidas em gelo por 1 hora. Após este período, 150 µL de células foram incubadas com 200 ng de plasmídeos contendo o gene URAD, e o sistema foi submetido a um choque térmico em um bloco de calor a 42°C por 3 minutos. Em seguida, 1 mL de meio LB foi adicionado às células transformadas, que foram incubadas com agitação a 37°C por 1 hora.

9.3.3 Expressão e purificação da proteína recombinante

9.3.3.1 *E. coli* BL21(DE3)pLys

As células BL21(DE3)pLys transformadas foram então usadas para preparar um pré-inóculo utilizando 10 mL de meio LB contendo 1 mM de cloranfenicol e 1 mM de canamicina. O pré-inóculo foi incubado com agitação a 37°C durante a noite. No dia seguinte, a inoculação foi realizada adicionando o pré-inóculo a 800 mL de meio LB contendo 1 mM de cloranfenicol e 1 mM de canamicina. O inóculo foi incubado com agitação a 37°C até atingir uma OD de 0,7, seguido pela adição de 0,5 mM de isopropil β-D-1-tiogalactopiranosídeo (IPTG). As células foram então incubadas com agitação a 18°C durante a noite.

Após este período, as células foram centrifugadas a 6000×g por 5 minutos a 4°C. O pellet foi ressuspendido em 20 mL de tampão fosfato A (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 50 mM de imidazol, pH 7,6) contendo um inibidor de protease

(SIGMAFAST™ Protease Inhibitor Cocktail Tablets), 10mM MgCl₂ e 10% glicerol. A lise celular foi realizada utilizando sonicação, aplicando pulsos de 3 segundos a 30% de amplitude (3s ligado, 3s desligado), com tempo total de 15 minutos. Subsequentemente, foi realizada centrifugação a 12000×g por 1 hora a 4°C.

O sobrenadante obtido foi submetido a purificação por afinidade utilizando uma coluna de polipropileno da BioRad (Econo-Pac®) contendo 1 mL de resina carregada com níquel. Para remoção de macromoléculas que não interagem com o níquel foi utilizado o tampão fosfato A (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 50 mM de imidazol, pH 7,6). A eluição da proteína de interesse foi realizada utilizando tampão fosfato B (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 500 mM de imidazol, pH 7,6). Durante o processo de purificação, a presença de proteínas foi avaliada utilizando o reagente Bradford (Thermo Scientific).

As frações obtidas na afinidade foram, então, colocadas em um sistema de diálise. A amostra foi colocada no interior de membranas de diálise, tendo estas sido fechadas em ambas extremidades. As membranas contendo a amostra foram colocadas em 4 litros de tampão fosfato C (50 mM de fosfato de sódio e 50 mM de cloreto de sódio, pH 7,6) e o sistema foi deixado sob agitação na câmara fria. Após a diálise foi realizada a concentração da proteína utilizando tubos Centricon de 30 mL (Merck) até que fosse obtido um volume de 5 mL.

Então, a cromatografia de exclusão por tamanho foi realizada no sistema de cromatografia ÄKTA Pure usando uma coluna HiTrap™ S75 superdex 16/60 (Cytiva). A eluição da proteína foi obtida utilizando tampão fosfato C. Foi, então, realizada a concentração da proteína utilizando tubos Centricon de 30 mL (Merck) até que fosse obtido um volume de, aproximadamente, 3 mL. A concentração final da proteína foi determinada utilizando a absorvância a 280 nm, considerando o coeficiente de extinção molar 8480 M⁻¹ cm⁻¹ e o caminho óptico de 1 cm.

Em todas as etapas da expressão e purificação foram coletadas amostras para serem adicionadas ao tampão de amostra 5x (SDS 10%, azul de bromofenol 0,25%, glicerol 50%, β-mercaptoetanol 500 mM, Tris-HCl 250 mM, pH = 6,8) e, então, utilizadas para a realização da eletroforese em gel de poliacrilamida (SDS-PAGE) 12%, conduzida a 30 mA, na presença de tampão de corrida (SDS 0,5%, Glicina 940 mM, Tris 125 mM). Para avaliar

a presença da proteína de interesse foi utilizado o padrão não pré-corado (Unstained Protein MW Marker - Thermo Scientific).

9.3.3.2 *E. coli* ArcticExpress (DE3)

As células Arctic transformadas foram então usadas para preparar um pré-inóculo utilizando 10 mL de meio LB contendo 1 mM de gentamicina (G418) e 1 mM de canamicina. O pré-inóculo foi incubado com agitação a 37°C durante a noite. No dia seguinte, a inoculação foi realizada adicionando o pré-inóculo a 800 mL de meio LB sem adição de antibióticos. O inóculo foi incubado com agitação a 24°C até atingir uma OD de 0,7, seguido pela adição de 0,5 mM de isopropil β -D-1-tiogalactopiranosídeo (IPTG). As células foram então incubadas com agitação a 15°C por 36 horas.

Após este período, as células foram centrifugadas a 6000×g por 5 minutos a 4°C. O pellet foi ressuscitado em 20 mL de tampão fosfato A (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 50 mM de imidazol, pH 7,6) contendo um inibidor de protease, 10mM MgCl₂ e 10% glicerol. A lise celular foi realizada utilizando sonicação, aplicando pulsos de 3 segundos a 30% de amplitude (3s ligado, 3s desligado), com tempo total de 15 minutos. Subsequentemente, foi realizada centrifugação a 12000×g por 1 hora a 4°C.

O sobrenadante obtido foi submetido a purificação por afinidade utilizando uma coluna de polipropileno da BioRad (Econo-Pac®) contendo 1mL de resina carregada com níquel. Para remoção de macromoléculas que não interagem com o níquel foi utilizado o tampão fosfato A (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 50 mM de imidazol, pH 7,6). A eluição da proteína de interesse foi realizada utilizando tampão fosfato B (20 mM de fosfato de sódio, 300 mM de cloreto de sódio e 500 mM de imidazol, pH 7,6). Durante o processo de purificação, a presença de proteínas foi avaliada utilizando o reagente Bradford (Thermo Scientific).

As frações obtidas na afinidade foram, então, colocadas em um sistema de diálise. A amostra foi colocada no interior de membranas de diálise, tendo estas sido fechadas em ambas extremidades. As membranas contendo a amostra foram colocadas em 4 litros de tampão fosfato C (50 mM de fosfato de sódio e 50 mM de cloreto de sódio, pH 7,6) e o sistema foi deixado sob agitação na câmara fria. Após a diálise foi realizada a concentração da proteína utilizando tubos Centricon de 30mL até que fosse obtido um volume de 5 mL.

Então, a cromatografia de exclusão por tamanho foi realizada no sistema de cromatografia ÄKTA Pure usando uma coluna HiTrap™ S200 sephacryl 16/60 (Cytiva). A eluição da proteína foi obtida utilizando tampão fosfato C. Foi, então, realizada a concentração da proteína utilizando tubos Centricon de 30mL (Merck) até que fosse obtido um volume de, aproximadamente, 3 mL. A concentração final da proteína foi determinada utilizando a absorvância a 280 nm, considerando o coeficiente de extinção molar 8480 M⁻¹ cm⁻¹ e o caminho óptico de 1 cm.

Novamente, em todas as etapas da expressão e purificação foram coletadas amostras para serem adicionadas ao tampão de amostra 5x e, então, utilizadas para a realização da eletroforese em gel de poliacrilamida (SDS-PAGE) 12%, conduzida a 30 mA, na presença de tampão de corrida. Para avaliar a presença da proteína de interesse foi utilizado o padrão não pré-corado (Unstained Protein MW Marker - Thermo Scientific).

9.4 Verificação da expressão da proteína recombinante por meio de Western Blot

Para verificar se a partir da expressão estava sendo obtida a proteína recombinante, foi utilizada a técnica de Western Blot, a partir da qual é possível confirmar a presença de proteínas em uma amostra utilizando um anticorpo primário, que interage especificamente com algum elemento estrutural presente na proteína de interesse, e um anticorpo secundário, que interage com o primário e, além disso, é conjugado com uma enzima ou uma substância fluorescente, o que permite sua detecção (YANG; MAHMOOD, 2012).

Para realização do western blot foram utilizadas as amostras, adicionadas ao tampão de amostra 5x, previamente preparadas. Foi realizada a eletroforese em gel de poliacrilamida (SDS-PAGE) 12%, conduzida a 30 mA, na presença de tampão de corrida utilizando os padrões pré-corado (Page Ruker™ Prestained Protein - Thermo Scientific) e não pré-corado (Unstained Protein MW Marker - Thermo Scientific).

As proteínas do SDS-PAGE foram então transferidas para uma membrana de fluoreto de polivinilideno (PVDF) Amersham Hybond de poro 0,45 µm (GE Healthcare) em um sistema de transferência semi-seco Trans-Blot (BioRad) por 60 min, conduzido a 300 mA. Para molhar a membrana e o gel foi utilizado o tampão de transferência (Tris 48 mM, Glicina 39 mM, SDS 0,037% e metanol 20%).

Após transferência a membrana foi temporariamente corada com Ponceau para verificar se a transferência havia ocorrido. A membrana foi, então, lavada e bloqueada em uma solução de 5% de leite em tampão TBS-T (Tris 20 mM, NaCl 150 mM, Tween 20 0,1%, pH = 7,6) por 2 horas. Após este período a membrana foi lavada utilizando TBS-T e, em seguida, incubada com o anticorpo primário anti-HisTag (Santa Cruz Biotechnology) na diluição 1:1000 em TBS-T contendo 2% de BSA. O sistema foi deixado sob agitação overnight à 4°C.

No dia seguinte foi retirada a solução contendo o anticorpo primário e a membrana foi lavada com TBS-T. Em seguida, a membrana foi incubada com o anticorpo secundário anti-mouse IgG conjugado com peroxidase na diluição 1:4000 em uma solução de 2% de leite em tampão TBS-T. O sistema foi deixado sob agitação por 1 hora e 30 minutos à temperatura ambiente.

Após esse período foi retirada a solução contendo o anticorpo secundário e a membrana foi lavada com TBS-T. As bandas foram reveladas utilizando reagente colorimétrico (Pierce™ 1-Step Ultra TMB Blotting Solution - Thermo Scientific).

Como controle, as mesmas amostras foram aplicadas em SDS-PAGE 12% na mesma quantidade, sendo a eletroforese conduzida também a 30 mA, na presença de tampão de corrida (SDS 0,5%, Glicina 940 mM, Tris 125 mM). Também foram utilizados os padrões pré-corado (Page Ruler™ Prestained Protein - Thermo Scientific) e não pré-corado (Unstained Protein MW Marker - Thermo Scientific). Neste caso, os géis foram corados utilizando corante black (H_3PO_4 25%, $(\text{NH}_4)_2\text{SO}_4$ 10%, Coomassie brilliant blue G250 0,12% e Etanol 20%).

9.5 Verificação da estrutura primária da proteína por meio de Espectrometria de Massa MALDI-TOF-MS/MS

Uma técnica amplamente utilizada para verificação da estrutura primária de peptídeos e proteínas é a espectrometria de massa MALDI-TOF-MS/MS (Matrix Assisted Laser Desorption Ionization – Time Of Flight) (LI *et al.*, 2022). Nesta técnica a amostra contendo as macromoléculas de interesse é misturada a uma matriz doadora de prótons e, então, cristalizada na placa de aço inoxidável a qual é inerte a ionização (LI *et al.*, 2022). Esta placa é levada ao equipamento, onde a amostra será incidida por um laser de 337 nm que

fornece energia suficiente para promover a sua dissociação e ionização (LI *et al.*, 2022). As moléculas ionizadas são então aceleradas por um campo elétrico para percorrer o tubo de voo, sendo os íons separados com base em sua razão m/z , visto que sua velocidade é inversamente proporcional a essa razão (LI *et al.*, 2022). Assim, íons menores chegam mais rapidamente ao detector, seguidos pelos íons maiores, que chegam mais lentamente (LI *et al.*, 2022).

Assim, procedeu-se com o protocolo de digestão da proteína no gel utilizando tripsina. Assim, as amostras obtidas após a diálise e concentração, adicionadas ao tampão de amostra 5x, foram aplicadas nos poços do gel de poliacrilamida (SDS-PAGE) 12% e procedeu-se com a eletroforese a 30 mA, na presença de tampão de corrida. O gel foi corado de forma branda utilizando o corante black e, em seguida, o máximo possível de corante foi retirado utilizando água destilada.

As bandas foram excisadas do gel utilizando um bisturi e cortadas em pequenos pedaços, os quais foram colocados em eppendorfs Lo Binding de 2 mL. Foi realizado, então, o seguinte procedimento de limpeza:

- Foram adicionados 600 μ L de água ultrapura e os eppendorfs foram colocados no vórtex por 15 minutos;

- A água foi retirada, foram adicionados 600 μ L de solução de acetonitrila 60% e os eppendorfs foram colocados no vórtex por 15 minutos;

- A solução de acetonitrila 60% foi retirada e, nos casos em que os pedaços de gel ainda estavam com o corante muito forte foram colocados mais 600 μ L de solução de acetonitrila 60% e os eppendorfs foram colocados no vórtex por mais 15 minutos;

- A solução de acetonitrila 60% foi retirada, foram adicionados 600 μ L de água ultrapura e os eppendorfs foram colocados no vórtex por 15 minutos;

- A água foi retirada, foram adicionados 600 μ L de solução de etanol 50% e os eppendorfs foram colocados no vórtex por 15 minutos;

- A solução de etanol 50% foi retirada, foram adicionados 600 μ L de etanol 100% e os eppendorfs foram colocados no vórtex por 15 minutos;

- Nos casos em que os pedaços de gel ainda não estavam opacos foram colocados mais 600 μ L de etanol 100% e os eppendorfs foram colocados no vórtex por mais 15 minutos.

O etanol 100% foi retirado e os eppendorfs contendo os pedaços de geis foram colocados para secar na evaporadora de solventes a vácuo (SpeedVac) por 1 hora a 45°C no modo vácuo aquoso (V.AQ). Após secos, os pedaços de gel foram submetidos a redução e alquilação por meio da reidratação em 300 μ L de uma solução contendo 10 mM de cloridrato de tris(2-carboxietil) fosfina (TCEP), 40 mM de brometo de tetraetilamônio (TEAB) e 40 mM de cloroacetamida. Os geis permaneceram na solução de TCEP, TEAB e cloroacetamida a temperatura ambiente por 2 horas.

Passadas as duas horas foi repetido o processo de limpeza utilizando água ultrapura, acetonitrila 60%, etanol 50% e etanol 100%. O etanol 100% foi retirado e os eppendorfs contendo os pedaços de geis foram colocados para secar na Speed Vac por 40 minutos a 45°C no modo vácuo aquoso (V.AQ). Após secos, os pedaços de gel foram submetidos à digestão por tripsina por meio da reidratação em 600 μ L de solução contendo 50 mM de ácido acético, 50 mM de TEAB e 0,2 μ g/ μ L de tripsina. Os geis permaneceram na solução de ácido acético, TEAB e tripsina a temperatura ambiente por 45 minutos e, em seguida, foram adicionados 250 μ L de uma solução contendo 50 mM de TEAB. O sistema foi deixado a 37°C por 16 horas.

O sobrenadante da digestão foi transferido para eppendorfs Lo Binding de 0,5 mL. Para extrair mais peptídeos foram adicionados 300 μ L de uma solução contendo 50% de acetonitrila e 5% de ácido fórmico aos pedaços de gel, os eppendorfs foram colocados no vórtex por 15 minutos e o sobrenadante foi transferido para eppendorfs Lo Binding de 0,5 mL.

Os sobrenadantes obtidos após a digestão e após a extração com 50% de acetonitrila e 5% de ácido fórmico foram colocados na Speed Vac por 30 minutos a 60°C no modo vácuo aquoso (V.AQ), sendo esse processo repetido até que os sobrenadantes secassem completamente. Após esta etapa, o material foi ressuspensão em 20 μ L de água ultrapura contendo 0,1% de ácido trifluoroacético (TFA).

Em seguida, procedeu-se com o protocolo para análise por espectrometria de massa MALDI-TOF. Para isso foram utilizadas ponteiras Zip-Tip com resina C₁₈ para eluição dos peptídeos. Primeiro, foram utilizadas soluções de acetonitrila 100% contendo 0,1% de TFA e água ultrapura contendo 0,1% TFA para ativar a coluna. Em seguida a amostra foi carregada na coluna, sendo utilizada água ultrapura contendo 0,1% TFA para limpar a coluna e, então, uma solução de acetonitrila 60% contendo 0,1% de TFA para eluir os peptídeos. Após este preparo, 2 µL da amostra foram misturados com 2 µL de solução contendo ácido α -ciano-4-hidroxicinâmico, a matriz doadora de prótons, e a mistura foi aplicada em duplicatas na placa de MALDI.

Os dados de espectrometria de massa MS e MS/MS foram obtidos utilizando um espectrômetro de massa MALDI-TOF/TOF (Bruker Daltonics AutoFlex III) equipado com um laser SmartBeam™ e controlado pelo programa FlexControl 3.0 (Laboratório Multiusuário de Proteômica, LMProt, ICB, UFMG). Os espectros foram adquiridos no modo de reflexão com calibração externa usando uma mistura de peptídeos. A estrutura primária foi interpretada manualmente através de sequenciamento *de novo*. Os espectros indicando as intensidades dos íons de acordo com sua relação m/z, os íons de série y e b, assim como o erro associado ao sequenciamento foram construídos utilizando a ferramenta *Interactive Peptide Spectral Annotator* (BRADEMAN *et al.*, 2019).

9.6 Avaliação da atividade descarboxilase utilizando ensaios enzimáticos

A fração obtida após a purificação por exclusão foi concentrada, e, então, utilizada para avaliar se o produto de expressão da URAD possuía atividade de OHCU descarboxilase. Para isso foram realizados ensaios enzimáticos utilizando o espectrofotômetro BioMate3S (Thermo Scientific), tendo as reações ocorrido em um tampão fosfato de potássio (pH 7,6), que também foi utilizado como branco. Além do produto de expressão da URAD, foram utilizadas a uricase recombinante de *Candida sp.* (Sigma Aldrich U0880) e a HIUase recombinante de *Herbaspirillum seropedicae* também produzida em nosso laboratório seguindo protocolo publicado (LEMOS *et al.*, 2024).

Diferentes condições foram testadas, incluindo: 30 µM de ácido úrico + 50 nM de uricase; 30 µM de ácido úrico + 50 nM de uricase + 50 nM de HIUase; 30 µM de ácido úrico + 50 nM de uricase + 50 nM de HIUase + 50nM de descarboxilase ; 30 µM de ácido úrico + 50 nM de uricase + 50 nM de HIUase + 100nM de OHCU descarboxilase.

Primeiro, foram medidos os espectros de soluções 30 μM e 300 μM de ácido úrico e de alantoína para verificar o intervalo de comprimentos de onda que deveria ser acompanhado. De porte desta informação, para os ensaios foi realizada uma varredura entre os comprimentos de onda 190 nm e 340 nm, tendo a aquisição sido realizada no modo rápido (FAST).

Os comprimentos de onda analisados foram selecionados de acordo com estudos anteriores com essas enzimas. Enquanto o desaparecimento do ácido úrico, que é um indicativo da atividade da uricase, é geralmente monitorado em 292 nm (NYBORG *et al.*, 2016), o desaparecimento do HIU pode ser usado para medir a atividade da HIUase em 302 nm (SARMA *et al.*, 1999). A conversão de OHCU para alantoína é avaliada na região de 240–260 nm, sendo, geralmente, monitorada a mudança na absorbância em 256 nm (FRENCH; EALICK, 2010).

9.7 Caracterização da estrutura secundária e da estabilidade térmica da proteína por meio de Dicroísmo Circular

Para verificar o padrão de estrutura secundária e a estabilidade térmica da proteína recombinante foi utilizada a técnica espectroscópica Dicroísmo Circular (CD). Esta técnica se baseia no princípio de que moléculas opticamente ativas, que apresentam pelo menos um centro quiral e, portanto, apresentam enantiômeros, interagem de maneira diferente com a luz polarizada circularmente, que é dividida em duas componentes, uma à esquerda e uma à direita (KELLY; JESS; PRICE, 2005). As moléculas opticamente ativas podem absorver essas duas componentes de forma diferente, resultando em um espectro que revela informações sobre a estrutura da molécula (KELLY; JESS; PRICE, 2005).

No caso das proteínas, que são macromoléculas opticamente ativas, os diferentes padrões de estrutura secundária, como α -hélices, fitas- β , assim como as regiões desordenadas, produzem espectros de dicroísmo circular distintos, que refletem a presença ou ausência de estrutura regular (KELLY; JESS; PRICE, 2005). Assim, esta técnica espectroscópica pode ser utilizada tanto para verificar o padrão de estrutura secundária na proteína em sua forma nativa, quanto para avaliar sua estabilidade térmica, pois ao submeter a proteína a valores crescentes de temperatura pode ser determinado o ponto no qual ocorre a desnaturação (KELLY; JESS; PRICE, 2005).

Nos experimentos de dicroísmo circular foi utilizada a fração concentrada obtida após a purificação por exclusão. Primeiro, foram coletados os espectros da amostra diluída em tampão fosfato C (1:1) à 20°C, sendo obtidos os valores de elipticidade entre os comprimentos de onda 190-260 nm, e de 20 à 90°C, sendo obtidos os valores de elipticidade nos comprimentos de onda 220 e 222 nm conforme a temperatura era acrescida em 5°C. Além disso, a amostra não diluída foi utilizada para coletar os espectros à 20°C, sendo, novamente, obtidos os valores de elipticidade entre os comprimentos de onda 190-260 nm, e de 20 à 100°C, sendo obtidos os valores de elipticidade entre os comprimentos de onda 210-260 nm conforme a temperatura era acrescida em 5°C.

Os espectros foram obtidos a partir da média de três leituras consecutivas e da subtração dos valores do sinal dicroico do espectro do tampão utilizado como branco. As elipticidades (mdeg) obtidas no ensaio foram convertidas em elipticidade molar ($[\theta]$) ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$) segundo a equação abaixo:

$$[\theta] = \frac{RMW \times \alpha}{10 \times C \times l}$$

Onde, RMW a massa molecular média por resíduo (Da), α é o sinal dicroico em mdeg, C é a concentração da amostra (mg/mL) e l é o trajeto óptico (cm) da cubeta utilizada.

Todos os espectros foram registrados utilizando um espectropolarímetro Jasco J-815 equipado com uma lâmpada, sendo utilizadas cubetas de quartzo de 0,1 mm (Laboratório Multiusuário de Proteômica, LMProt, ICB, UFMG).transferência de energia

9.8 Caracterização do enovelamento proteico com Ressonância Magnética Nuclear

Para confirmar se a proteína se encontrava de fato enovelada, foi utilizada a técnica espectroscópica Ressonância Magnética Nuclear (RMN), uma técnica espectroscópica de absorção baseada nas propriedades intrínsecas de *spin* nuclear de alguns núcleos atômicos (HU *et al.*, 2021).

O analito dentro do tubo de RMN é submetido a um campo magnético constante, o que faz com que o vetor de momento magnético precesse ao redor do eixo do campo. Quando um pulso de radiofrequência é aplicado, ocorre a absorção de energia e

transferência do nível base para um nível de energia mais alto e, quando o *spin* retorna ao seu nível base, ocorre um processo de relaxação, com transferência de magnetização de um átomo para outros que estejam próximos espacialmente (HU *et al.*, 2021). O sinal correspondente a essa transferência, que depende do ambiente químico no qual o átomo se encontra, é medido e processado para gerar um espectro de RMN para o núcleo analisado. Como em uma macromolécula os átomos se encontram em ambientes químicos distintos, tanto devido a interações covalentes quando as não covalentes, os diferentes átomos vão apresentar sinais de deslocamento distintos no espectro de RMN, o que permite a caracterização da macromolécula analisada (HU *et al.*, 2021).

Assim, a fração concentrada obtida após a purificação por exclusão foi utilizada para coleta do espectro 1D de ressonância magnética nuclear, sendo utilizados 450 µL de amostra, adicionados de 50 µL de óxido de deutério (10% D₂O). O espectro de 1H foi registrado à 25°C utilizando um espectrômetro Bruker Ascend Neo 600 equipado com uma sonda BBO de 5 mm de 2 canais, a 600,15 MHz para a frequência de 1H (Laboratório de Ressonância Magnética de Alta Resolução, LAREMAR, DQ, UFMG).

10. RESULTADOS E DISCUSSÃO

10.1 Presença de resíduos críticos conservados no produto da URAD indicam funções catalíticas essenciais e alinhamento da alantoicase de *S. cerevisiae* e ALLC indicam diferenças evolutivas significativas

Ao utilizar o T-Coffee (ClustalW - Clustal2.1) (NOTREDAME; HIGGINS; HERINGA, 2000) para alinhar a sequência de resíduos de aminoácidos do produto de expressão da URAD com sequências de OHCU descarboxilase cuja estrutura já foi resolvida ou calculada e depositada no PDB (BERMAN, 2000) foram encontrados os seguintes resultados:

- OHCU descarboxilase de *Danio rerio* (2O70): 52,02% de identidade;
- OHCU descarboxilase de *Agrobacterium tumefaciens* (PDB 2O8I) de 38,51%;
- Domínio descarboxilase da alantoína sintase de *P. tricornutum* (5Z5M): 31,48%;
- OHCU descarboxilase de *Klebsiella pneumoniae* (3O7H): 28,75%;

- OHCU descarboxilase de *A. thaliana* (2Q37): 24,67%.

Então, prosseguiu-se com a avaliação do sítio catalítico da OHCU descarboxilase, conforme proposto pelos autores dos artigos referentes a cada uma das estruturas. No caso enzima de *D. rerio*, Cendron, L. *et al.* (2007) propõem que o sítio catalítico seria composto pelos resíduos com cadeia lateral carregada glutamato (Glu) 23, histidina (His) 67, Glu 87 e arginina (Arg) 161, pelos resíduos com cadeia lateral polar serina (Ser) 84 e glutamina (Gln) 88; e leucina (Leu) 70, valina (Val) 120, alanina (Ala) 123 e isoleucina (Ile) 157 com cadeia lateral apolar (CENDRON *et al.*, 2007).

No trabalho de French e Ealick (2010) os autores propõem que sítio ativo da OHCU descarboxilase de *K. pneumoniae* seria composto por resíduos com cadeia lateral hidrofóbica prolina (Pro) 68, fenilalanina (Phe) 117, Ile 119 e Ala 121, assim como por resíduos com cadeia lateral hidrofílica His 67, Ser 84, Glu 87 e Gln 88 (FRENCH; EALICK, 2010).

Para a OHCU descarboxilase de *A. thaliana*, Kim, Park, e Rhee (2007) propõem que o sítio catalítico seria composto pelos resíduos com cadeia lateral hidrofóbica Ile 61, Ile 112, Ala 115 e Ile 149 e pelos resíduos com cadeia lateral hidrofílica His 58, Glu 80 e Arg 153 (KIM; PARK; RHEE, 2007). Já no artigo de Oh, J. *et al.* (2018) os autores afirmam no domínio OHCU descarboxilase de *P. tricornutum* é possível observar os mesmos resíduos com cadeia lateral hidrofílica que estão presentes na enzima de *A. thaliana*, sendo eles His 75, Glu 95 e Arg 169 na sequência da proteína de *P.tricornutum* (OH *et al.*, 2018).

Conforme indicado nos alinhamentos na **Figura 16**, a maior parte destes resíduos que integram o sítio catalítico da OHCU descarboxilase está conservada no produto proteico da URAD, tendo, em alguns casos, ocorrido substituições de resíduos com cadeia lateral hidrofóbica. Contudo, o resíduo substituído na cadeia polipeptídica do produto proteico da URAD também apresenta cadeia lateral hidrofóbica.



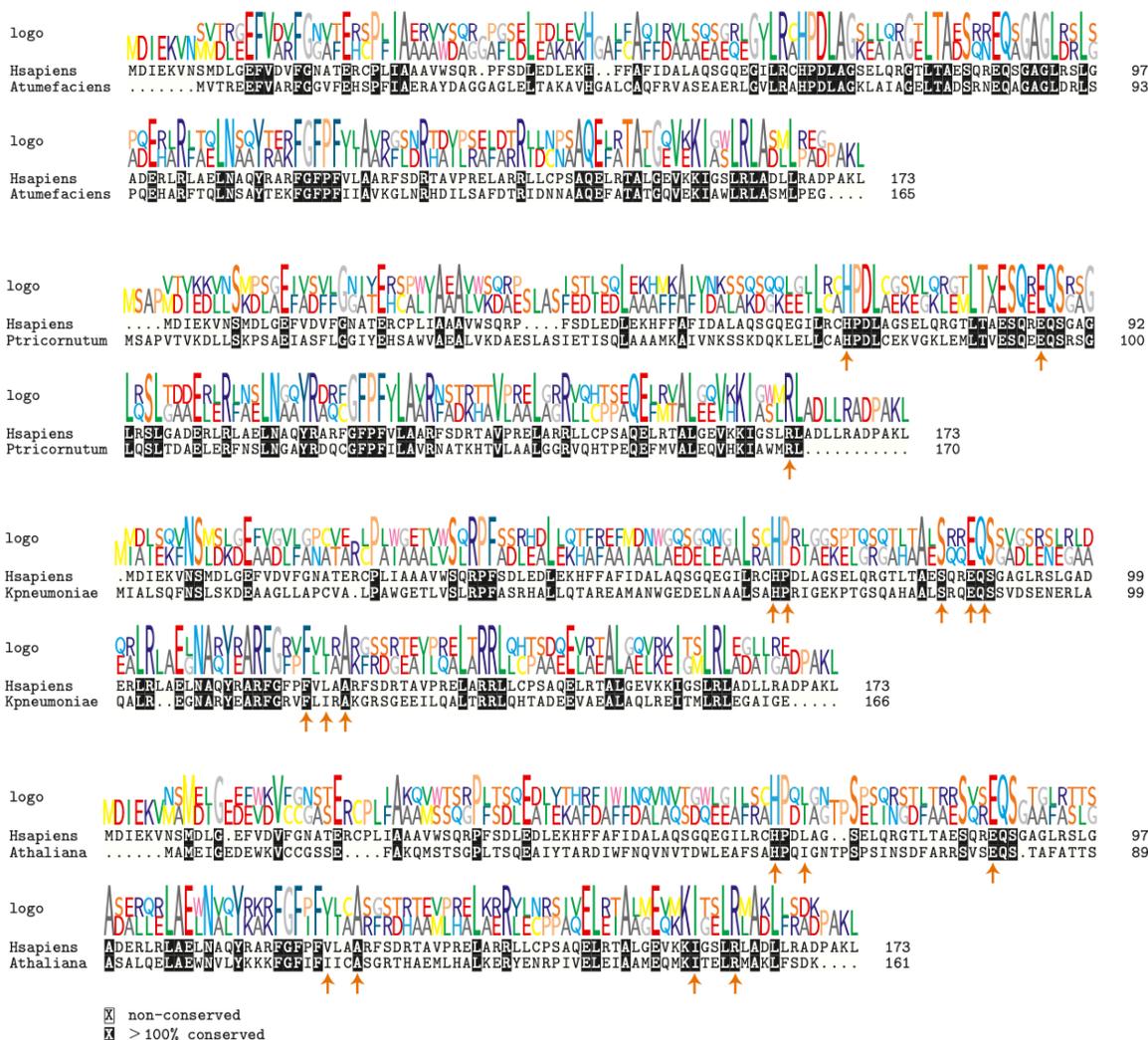


Figura 16. Alinhamentos da estrutura primária do produto proteico da URAD com seqüências de OHCU descarboxilase. Alinhamento da seqüência do produto proteico da URAD com seqüências de OHCU descarboxilase cuja estrutura está depositada no PDB (2070, 2081, 5Z5M, 307H e 2Q37). Os resíduos idênticos entre as seqüências são indicados pelas caixas pretas e os resíduos do sítio catalítico de cada estrutura depositada são indicados pelas setas laranjas.

Além disso, pode-se notar a presença da seqüência de direcionamento ao peroxissomo tipo 1 (PTS-1), definida pelo tripeptídeo (Ser/Ala)-(Lys/Arg/His)-Leu na região C-terminal da URAD hipotética. No trabalho de Oh, J. *et al.* (2018) os autores discutem que, enquanto em plantas geralmente é encontrada a seqüência de direcionamento ao peroxissomo tipo 2 (PTS-2) na região de ligação entre os domínios HIUase e OHCU descarboxilase da enzima bifuncional, em outros eucariotos a HIUase apresenta a PTS-2 na região N-terminal e a OHCU descarboxilase a PTS-1 na região C-terminal (OH *et al.*, 2018).

A presença do PTS-1 no produto proteico da URAD é consistente com o que foi levantado anteriormente em relação ao trabalho de Giaccherini *et al*, 2023, no qual os autores discutem que evidências recentes sugerem a atividade da URAD em peroxissomos (GIACCHERINI *et al.*, 2023).

Para avaliar se os resíduos mencionados acima são estrutural e funcionalmente importantes para a OHCÚ descarboxilase, foi realizada a análise utilizando o PFstats (FONSECA-JÚNIOR *et al.*, 2018). Tendo utilizado os filtros explicitados na metodologia, o número final de sequências foi 4207. Entre os resíduos do produto proteico da URAD que apresentaram uma conservação acima de 70% estão His 67, Pro 68, Ser 84, Glu 87, Gln 88, asparagina (Asn) 108, tirosina (Tyr) 111, Phe 115, Arg 138, Glu 146, Ile 157, Arg 161.

Na análise de correlação foram identificadas duas comunidades cujos resíduos formam uma rede complexa de interações, como mostrado na **Figura 17**. A comunidade 1 é composta pelos resíduos Ala 273, Pro 339, Glu 315, Phe 8, glicina (Gly) 274, Leu 220, Glu 18, Leu 322, valina (Val) 374 e Leu 275, correspondentes aos resíduos Ala 91, Pro 118, Glu 100, Phe 14, Gly 92, Leu 70, Glu 23, Leu 107, Ala 123 e Leu 93 no produto proteico da URAD, tendo ocorrido a substituição da Val 374 pela Ala 123. A comunidade 2 é composta pelos resíduos Glu 269, His 215, Asn 323, Pro 218, glutamina (Gln) 270 e Ser 265 correspondentes aos resíduos Glu 87, His 67, Asn 108, Pro 68, Gln 88 e Ser 84 no produto proteico da URAD.

Além disso, foram identificadas outras comunidades compostas por 2 a 4 resíduos de aminoácidos (**Figura 17**), o que pode ser um indicativo de interações intermoleculares que são estabelecidas no interior da proteína. Para algumas comunidades pode ser observada a conservação dos resíduos de aminoácidos no produto proteico da URAD, como é o caso da comunidade composta pela Gly 222, Ala 221 e treonina (Thr) 81. Em outras comunidades ocorreram substituições, mas preservando as propriedades físico-químicas dos resíduos, como na comunidade composta pelos resíduos de Leu 341 e Leu 527, que correspondem a Val 120 e Val 144 no produto proteico da URAD.

Diversos resíduos que apresentaram uma alta conservação, acima do ponto de corte de 70%, se fazem presentes na comunidade 2, indicando que estes resíduos juntos muito provavelmente desempenham um papel importante na estrutura e função da proteína. Além disso, muitos dos resíduos conservados e que fazem parte das duas comunidades fazem

parte, ou estão na mesma região, do sítio ativo proposto para a OHCU descarboxilase, o que fornece mais uma evidência da necessidade da manutenção destes resíduos na sequência da enzima.

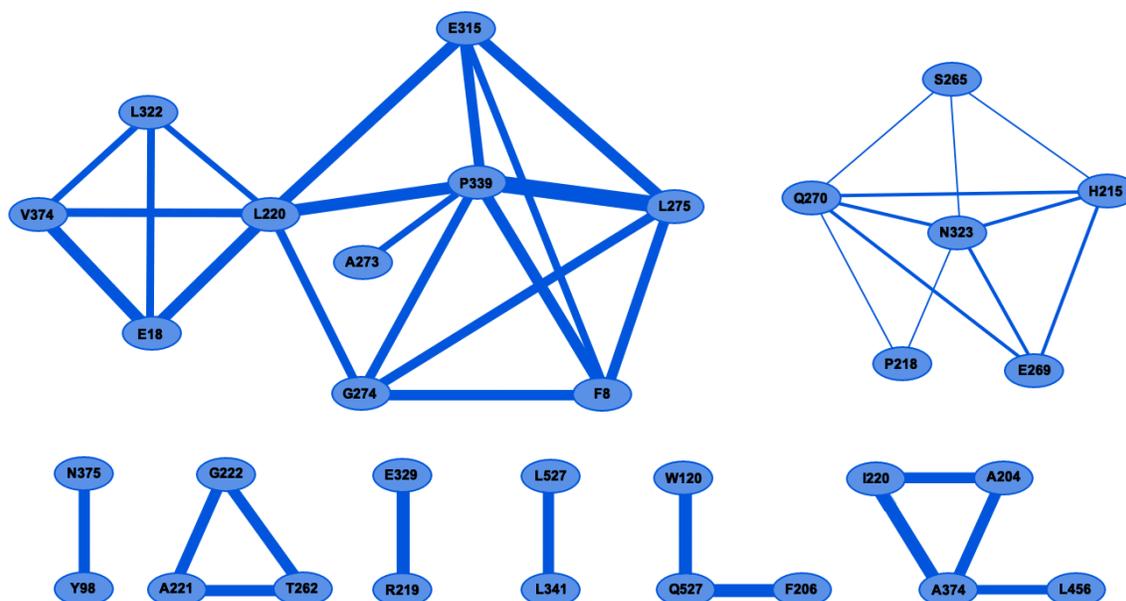


Figura 17. Redes de correlação para resíduos de duas comunidades da OHCU descarboxilase. Na parte superior estão representadas as redes de correlação das comunidades 1 (esquerda) e 2 (direita) obtidas a partir da análise do alinhamento das sequências de OHCU descarboxilase realizadas no PFstats. Na parte inferior estão representadas as redes de correlação das comunidades 3 a 8.

Em relação à alantoicase, a busca no PDB (BERMAN, 2000) indicou que havia sido resolvida apenas a estrutura da enzima de *Saccharomyces cerevisiae*. Assim, ao alinhar a sequência de resíduos de aminoácidos do produto proteico da ALLC com a sequência da alantoicase de *S. cerevisiae* (PDB 1O59), utilizando o T-Coffee (ClustalW - Clustal2.1) (NOTREDAME; HIGGINS; HERINGA, 2000), foi indicado que as sequências apresentam 35,13% de identidade.

Em seu estudo, Xu, Q. *et al* (2004) propõem que o sítio ativo da alantoicase é composto por dois clusters de resíduos hidrofílicos, o AR1, que compreende os resíduos Glu 72, Arg 75, aspartato (Asp) 82, Asn 108 e Asp 172, e o AR2 que compreende os resíduos Glu 235, Arg 238, Asp 246, Asn 272 e Asp 332 (XU *et al.*, 2004). Destes 10 resíduos, 8 estão conservados na ALLC hipotética, os outros dois resíduos, a Asn 108 e o Asp 246, foram substituídos, porém, por outros resíduos hidrofílicos, o Asp 85 e o Glu 260, respectivamente, conforme indicado no alinhamento na **Figura 18A**.

Ao avaliar a sequência da alantoicase de *S. cerevisiae* e o produto proteico da ALLC utilizando o InterPro (PAYSAN-LAFOSSE *et al.*, 2023), é possível notar que há dois domínios em ambas sequências, como ilustrado na **Figura 18B**, sendo que cada um dos clusters hidrofílicos está em um dos domínios. Assim, é possível que a enzima apresente dois sítios catalíticos independentes.

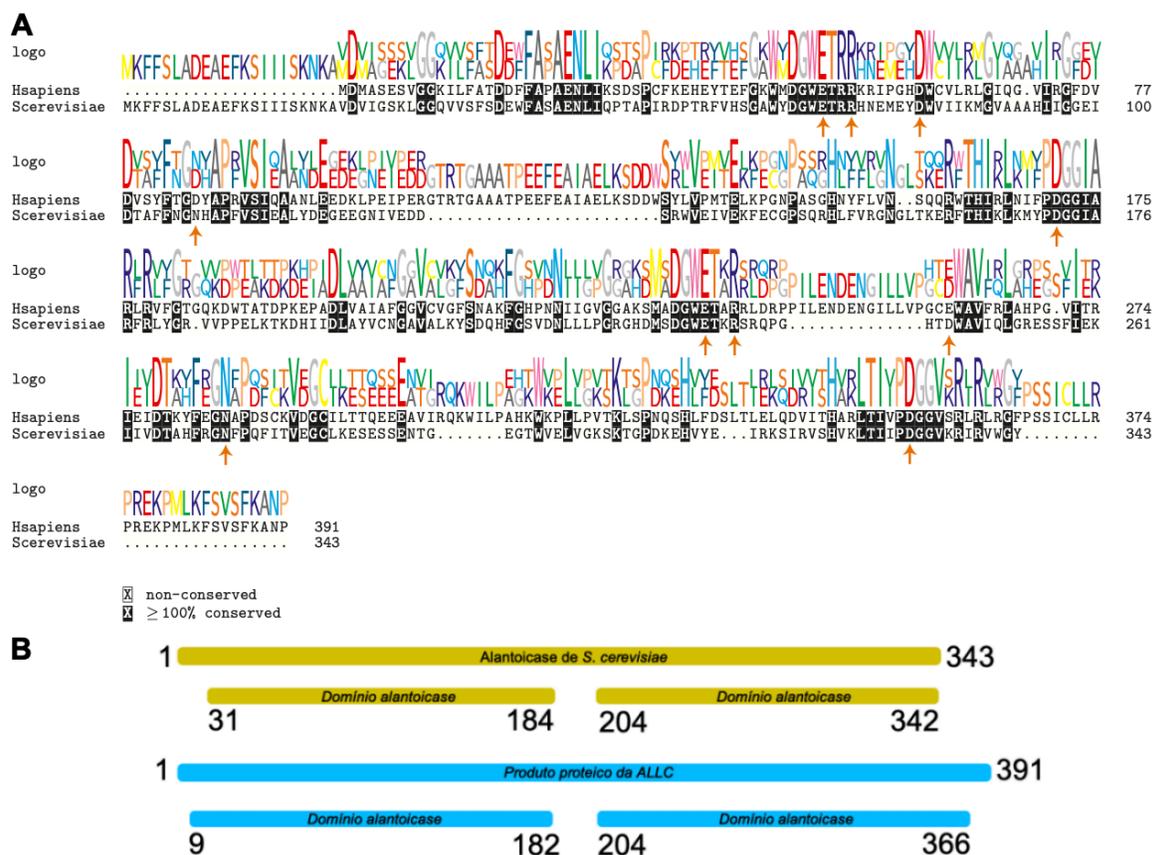


Figura 18. Alinhamento da estrutura primária do produto proteico da ALLC com sequência de alantoicase e domínios presentes na sequência. A) Alinhamento da sequência do produto proteico da URAD com sequências da alantoicase de *S. cerevisiae*, cuja estrutura está depositada no PDB (1O59). Os resíduos idênticos entre as sequências são indicados pelas caixas pretas e os resíduos do sítio catalítico da estrutura depositada são indicados pelas setas laranjas. B) Domínios alantoicase presentes nas sequências produto proteico da URAD e da enzima de *S. cerevisiae*

É possível notar no alinhamento (**Figura 18A**) que há diversas regiões da cadeia polipeptídica que são exclusivas da sequência de *H. sapiens*, havendo também uma região exclusiva da sequência de *S. cerevisiae*. Ao avaliar no BLASTp (ALTSCHUL *et al.*, 1990) a região N-terminal da sequência de *S. cerevisiae* (MKFFSLADEAEFKSIIISKNA) é identificado o alinhamento com sequências de outros fungos, além de bactérias e arqueas. Já para a porção C-terminal da sequência de *H. sapiens* (PSSICLLRPREKPMLKFSVSFKANP), assim como para alguns fragmentos no meio da

sequência (ERGTRTGAAATPEEFEAIAE e PPILENDENGILLV), os resultados indicam sequências presentes apenas em mamíferos placentários, incluindo diversos primatas.

Estas diferenças observadas nas sequências de proteínas entre *H. sapiens* e *S. cerevisiae* sugerem a ocorrência de um acúmulo de mutações na região gênica da alantoicase ao longo da evolução, que, eventualmente, levou a divergências na estrutura e, possivelmente, função do produto proteico. Enquanto a região N-terminal da sequência de *S. cerevisiae*, que parece estar conservada em outros fungos, bactérias e arqueas, poderia desempenhar uma função essencial nesses organismos, as regiões exclusivas na sequência de *H. sapiens*, que são encontradas apenas em mamíferos placentários, podem estar relacionadas a funções adaptativas ou especializadas que surgiram em mamíferos.

A análise utilizando o PFstats (FONSECA-JÚNIOR *et al.*, 2018) também foi realizada para as sequências de alantoicase. Tendo utilizado os filtros explicitados na metodologia, o número final de sequências foi 2848. Foram encontrados diversos resíduos do produto proteico da ALLC que apresentam uma conservação acima de 70%, estando eles indicados na **Tabela 6**. Foram observadas apenas 4 alterações, nos resíduos Val 79, Asp 85, Asn 212 e Glu 260, sendo os resíduos consenso Thr, Asn, Asp e Asp, respectivamente. Pode-se notar que todos estes resíduos conservados estão dentro dos domínios alantoicase indicados no InterPro (PAYSAN-LAFOSSE *et al.*, 2023), entre os resíduos 1 e 182 e 204 e 366, o que indica que estas regiões realmente correspondem aos domínios funcionais da enzima.

Tabela 6: Resíduos conservados presentes no produto proteico da ALLC

Resíduos conservados domínio 1											
G9	D17	G23	G48	W49	E50	T51	R53	D60	L66	D78	V79
Resíduos conservados domínio 2											
G204	N212	G226	G233	W234	E235	T236	R238	E260	L266	D278	T279
Resíduos conservados domínio 1											
F82	G84	D85	W132	P170	D171	G172	G173	R176	R178	G181	

Resíduos conservados domínio 2

F282	G284	N285	W317	P354	D355	G356	G357	R360	R362	G365
------	------	------	------	------	------	------	------	------	------	------

10.2 O produto proteico da URAD possui conservação no enovelamento com outros organismos modelos e conservação espacial dos resíduos críticos para a interação com a alantoína.

Considerando os resultados obtidos na análise da estrutura primária, prosseguiu-se com a avaliação da estrutura terciária da OHCU descarboxilase com intuito de melhor entender as características da conformação tridimensional desta enzima antes de realizar a modelagem do produto hipotético da expressão da URAD.

Ao analisar as estruturas de OHCU descarboxilase depositadas é possível notar uma clara conservação da estrutura tridimensional dessa enzima entre os diferentes organismos (CENDRON *et al.*, 2007; FRENCH; EALICK, 2010; KIM; PARK; RHEE, 2007; OH *et al.*, 2018), o que é condizente com os graus de identidade encontrados ao comparar a sequência da URAD com as estruturas primárias das descarboxilases cujas estruturas foram depositadas. A OHCU descarboxilase é composta, basicamente, por regiões ordenadas que assumem um padrão de α -hélices conectadas por regiões desordenadas, como mostrado na **Figura 19A**.

Contudo, há uma discordância na literatura em relação ao estabelecimento de uma estrutura quaternária. Enquanto para as estruturas de OHCU descarboxilase de *D. rerio*, *A. thaliana* e *P. tricornutum* é sugerida a formação de um dímero (**Figura 19B**), com cada um dos monômeros apresentando um sítio catalítico independente (CENDRON *et al.*, 2007; KIM; PARK; RHEE, 2007; OH *et al.*, 2018), para a OHCU descarboxilase de *K. pneumoniae* os autores afirmam que não é possível determinar, com certeza, o estado oligomérico da enzima (FRENCH; EALICK, 2010). Além disso, apesar de ter sido sugerido que as OHCU descarboxilases de *D. rerio*, *A. thaliana* e *P. tricornutum* formam um dímero, não há uma concordância em relação a interface que seria formada entre as subunidades monoméricas, conforme representado na **Figura 19B** (CENDRON *et al.*, 2007; KIM; PARK; RHEE, 2007; OH *et al.*, 2018).

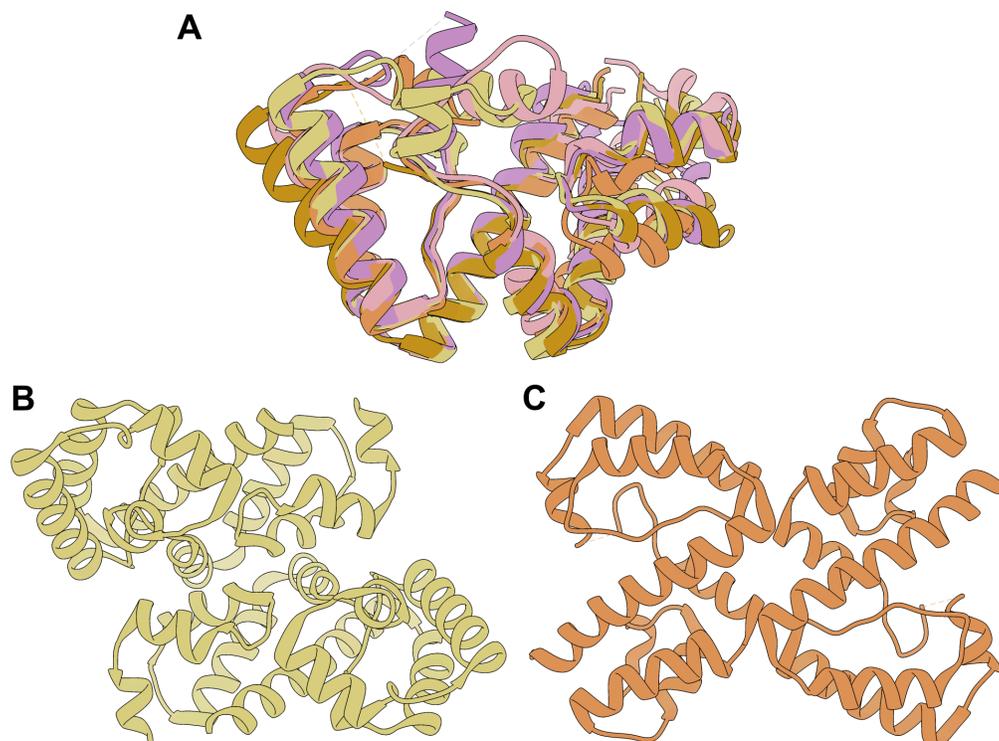


Figura 19. Estruturas de OHCU descarboxilase depositadas no PDB. A) Alinhamento das estruturas de OHCU descarboxilase de *D. rerio* (2O70) em caqui, *A. thaliana* (2Q37) em laranja, *A. tumefaciens* (2O8I) em roxo e *K. pneumoniae* (3O7H) em rosa, assim como o domínio OHCU descarboxilase da alantoína sintase de *P. tricornutum* (5Z5M) em ocre. Modos de interação propostos para as subunidades monoméricas da OHCU descarboxilase de B) *D. rerio* (2O70) e C) *A. thaliana* (2Q37).

Considerando essa discordância presente na literatura foi decidido que para avaliar a estrutura tridimensional da URAD seria realizada a modelagem da proteína hipotética em sua forma monomérica.

Foi, então, realizada a modelagem do produto de expressão da URAD por meio do ColabFold (MIRDITA *et al.*, 2022), sem estabelecer um molde específico. As estruturas representativas encontradas nos bancos de dado correspondem a OHCU descarboxilase de *D. rerio* (PDB 2O70), de *A. thaliana* (PDB 2Q37), de *K. pneumoniae* (PDB 3O7H, 3O7I, 3O7K e 5UOU), além da alantoína sintase de *P. tricornutum* (PDB 5Z5M) e a OHCU descarboxilase de *A. tumefaciens* (PDB 2O8I).

Além disso, foi realizada a modelagem por homologia utilizando o Modeller (ŠALI; BLUNDELL, 1993). Neste caso foram utilizadas a sequência primária do produto proteico da URAD e as coordenadas atômicas da OHCU descarboxilase de *D. rerio* (PDB 2O70), tendo sido geradas 3000 estruturas e feita a minimização e avaliação da estrutura de menor energia. Ademais, também foi utilizado o i-TASSER (YANG *et al.*, 2015) para realizar a

modelagem por *threading*. Os principais moldes encontrados nos bancos de dados utilizados para o *threading* foram as estruturas de OHCU descarboxilase de *Danio rerio* (PDB 2O70 e 2O73).

As cinco estruturas geradas no ColabFold apresentaram valores de pLDDT entre 94,9 e 95,1 ($95 \pm 0,0894$, sendo o coeficiente de variação 0,0941%) (**Anexo VI**). Os valores acima de 90 indicam uma alta confiança na precisão do modelo e, além disso, a variação mínima nos valores de pLDDT entre os cinco modelos sugere que as predições são consistentes entre si, ou seja, houve uma convergência para uma solução similar em cada tentativa de modelagem (JUMPER *et al.*, 2021). No caso da modelagem utilizando iTASSER foi obtido um resultado similar, tendo sido gerado, no final, apenas um modelo devido a alta convergência entre os cinco modelos construídos (YANG *et al.*, 2015), sendo que este único modelo apresentava um C-score de 1,10 e um TM-score de $0,86 \pm 0,07$. Assim, as predições do i-TASSER também apresentaram uma alta confiança e foram consistentes entre si. Utilizando ambas estratégias foi possível notar o mesmo padrão observado nas estruturas depositadas, sendo gerados modelos apresentando α -hélices conectadas por regiões desordenadas.

Ao contrário do ColabFold e do iTASSER, que tendem a convergir rapidamente para uma solução confiável devido aos métodos avançados de predição e refino, o que permite que, geralmente, um pequeno número de modelos gerados seja suficiente para obter uma predição de alta qualidade, o Modeller, que parte, basicamente, do alinhamento de sequências, gera milhares de modelos para capturar a variabilidade conformacional e garantir a obtenção de um modelo preciso. Ao analisar os resultados do Modeller foi possível notar que entre os 3000 modelos houve uma consistência no padrão das estruturas geradas, com os modelos de menor energia já apresentando o padrão de α -hélices conectadas por regiões desordenadas (**Anexo VII**), sugerindo que o algoritmo do Modeller está convergindo para uma solução e que esta conformação predominante foi identificada como a mais provável. Além disso, os cinco modelos de maior energia apresentaram valores de DOPE próximos ($-19.551,693 \pm 14,163$, sendo o coeficiente de variação 0,0724%), sendo esta baixa variação um indicativo que os modelos gerados são energeticamente favoráveis e similares entre si (**Anexo VII**).

Como cada ferramenta, utilizando métodos diferentes, chegou a uma solução similar e, além disso, houve convergência e consistência dos resultados entre diferentes ferramentas, tem-se um forte indicativo da precisão da estrutura modelada. Considerando estes resultados e sendo a proposição inicial do trabalho apenas a avaliação geral dos modelos, prosseguiu-se com a validação e análise apenas de cada um dos modelos que foi melhor classificado por cada uma das ferramentas.

No processo de validação os resultados do ERRAT (COLOVOS; YEATES, 1993) e do gráfico de Ramachandran gerado por meio do PROCHECK (LASKOWSKI *et al.*, 1993; RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963), indicam que os modelos gerados utilizando o ColabFold, o Modeller e o iTASSER apresentam uma boa qualidade.

No caso do resultado do ERRAT (COLOVOS; YEATES, 1993), o único modelo que apresentou um valor de qualidade geral, referente às interações não covalentes estabelecidas, inferior ao da estrutura de OHCU descarboxilase de *D. rerio* (PDB 2O70) foi o do Modeller, sendo os valores de qualidade 94,268 e 92,073 (**Anexos VIII e X**), respectivamente. Contudo, o segmento do modelo que apresentou uma menor qualidade, com um valor de erro superior a 99%, corresponde ao segmento da estrutura de OHCU descarboxilase de *D. rerio* que também apresentou uma menor qualidade. No caso dos modelos gerados pelo ColabFold e pelo iTASSER o valor de qualidade geral foi de 99,390 e 99,394 (**Anexos IX e XI**), respectivamente.

Já no gráfico de Ramachandran gerado por meio do PROCHECK (LASKOWSKI *et al.*, 1993; RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963) os modelos do ColabFold e do Modeller (**Anexos IX e X**) obtiveram resultados similares aos da estrutura de OHCU descarboxilase de *D. rerio* (PDB 2O70), com mais de 93% dos resíduos em regiões mais favoráveis e menos de 7% em regiões favoráveis adicionais. No caso do modelo gerado pelo iTASSER, um resíduo, a Ser 89, apresentou valores de ângulos diédricos ϕ (phi) e ψ (psi) em regiões menos favoráveis (**Anexo XI**). Contudo, isso não alterou a formação da alfa hélice na qual a Ser 89 está localizada e, apesar de sido observada uma rotação na cadeia lateral deste resíduo, quando comparado com o mesmo resíduo da OHCU descarboxilase de *D. rerio* (PDB 2O70), essa alteração não teria muito

impacto na estrutura e função da proteína, visto que a cadeia lateral da Ser 89 está voltada para o exterior da proteína.

Prosseguiu-se com o alinhamento da cadeia principal das estruturas tridimensionais dos modelos de maior pontuação gerados com a estrutura da OHCU descarboxilase *D. rerio* (PDB 2O70) utilizando o Chimera (PETTERSEN *et al.*, 2004). No alinhamento da enzima *D. rerio* com modelo do ColabFold (**Figura 20A**) foi obtido um RMSD (desvio da média quadrática, do inglês *Root Mean Square Deviation*) de 0,700 Å, com o do i-TASSER (**Figura 20B**) o valor de RMSD foi de 0,301 Å e com o do Modeller (**Figura 20C**) foi de 0,211 Å. Estes valores indicam que, no geral, houve uma boa superposição dos conjuntos de coordenadas atômicas das cadeias principais das estruturas comparadas. Ao avaliar os resíduos do sítio catalítico da descarboxilase *D. rerio* foi possível notar que, com exceção da His 67 do modelo do iTASSER, cuja cadeia lateral apresentou um ângulo de rotação diferente daquele observado para a estrutura da enzima *D. rerio* e para os modelos do ColabFold e do Modeller, houve uma boa superposição das coordenadas atômicas das cadeias laterais, conforme mostrado na **Figura 20D**.

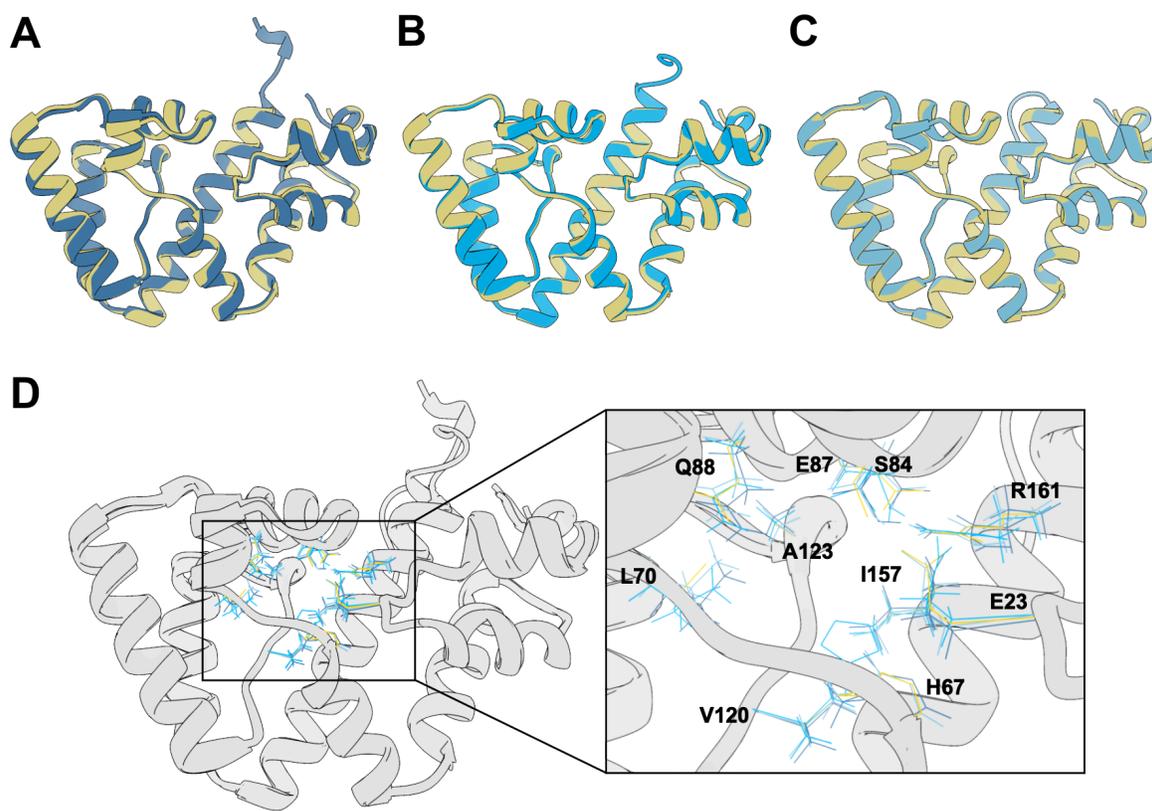
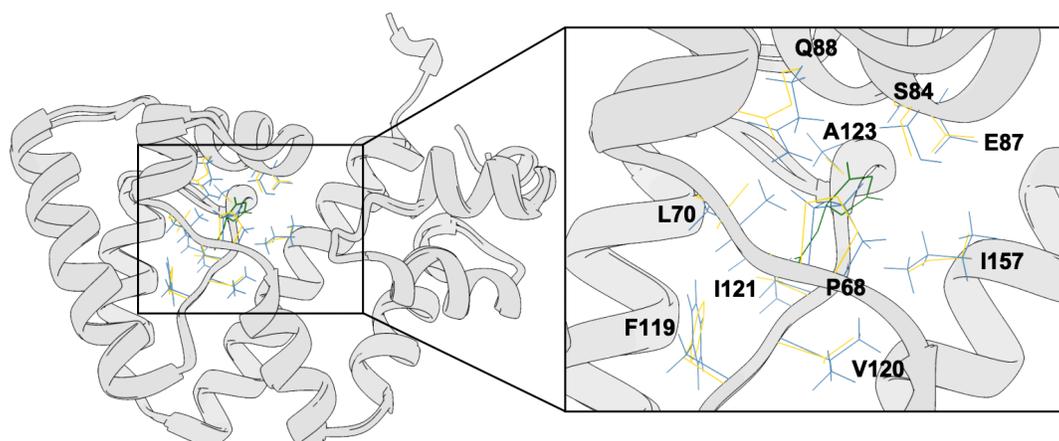


Figura 20. Modelos construídos para o produto proteico da URAD. A) Alinhamento da estrutura de OHCU descarboxilase de *D. rerio* (2O70) em caqui com o modelo de menor energia do

ColabFold, B) iTASSER e C) Modeller em tons de azul. D) Alinhamento da OHCU descarboxilase de *D. rerio* (2O70) com os modelos de menor energia do ColabFold (azul escuro), iTASSER (azul) e Modeller (azul claro), sendo indicados os resíduos do sítio catalítico da enzima de *D. rerio* em amarelo (Glu 23, His 67, Leu 70, Ser 84, Glu 87, Gln 88, Val 120, Ala 123, Ile 157 e Arg 161) e os resíduos correspondentes nos modelos.

Vale notar que no PDB também foi depositada a OHCU descarboxilase de *Danio rerio* em complexo com a alantoína, o produto da reação catalisada por esta enzima (PDB 2O73). A partir da estrutura da OHCU descarboxilase de *Danio rerio* em complexo com a alantoína é possível obter as interações estabelecidas entre a proteína e seu ligante utilizando o PDBSum (LASKOWSKI *et al.*, 2018). Neste caso, são formadas 6 interações do tipo ligação de hidrogênio envolvendo a cadeia principal ou lateral dos resíduos Ser 84, Glu 87, Gln 88, Ile 121 e Ala 123 com átomos de oxigênio e nitrogênio presentes na estrutura da alantoína. Além disso, são estabelecidas interações que envolvem os resíduos Pro 68, Leu 70, Phe 119, Val 120 e Ile 157.

Ao comparar o produto proteico da URAD com a OHCU descarboxilase de *D. rerio* é possível notar a conservação dos resíduos de aminoácidos que estabelecem as interações do tipo ligação de hidrogênio com a alantoína, tendo ocorrido apenas a substituição da Ile121 presente na proteína de *D. rerio* pela Leu121 presente no produto proteico da URAD, como pode ser visto na **Figura 21**. Ademais, os resíduos Pro68, Leu70, Phe119, Val120 e Ile157 também se fazem presentes no produto proteico da URAD (**Figura 21**) e, então, também poderiam contribuir com o estabelecimento de interações com a alantoína.



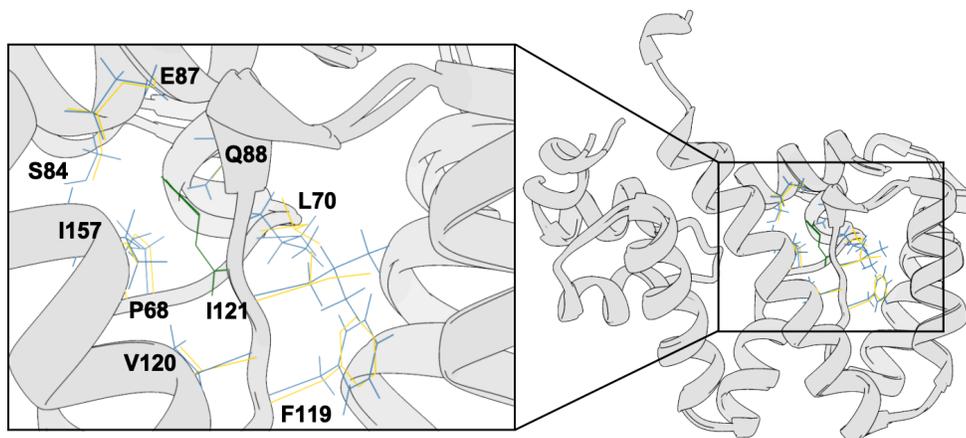


Figura 21. Sítio de interação da OHCU descarboxilase de *D. rerio* com a alantoína. A) Alinhamento da estrutura de OHCU descarboxilase de *D. rerio* em complexo com a alantoína (2O73) com o modelo de menor energia do ColabFold, sendo indicados os resíduos da enzima de *D. rerio* que interagem com a alantoína (laranja) em azul e os resíduos correspondentes no modelo do ColabFold em amarelo. Os resíduos Ser 84, Glu 87, Gln 88, Ile 121 e Ala 123 participam de interações do tipo ligação de hidrogênio e os resíduos Pro 68, Leu 70, Phe 119, Val 120 e Ile 157 participam de interações hidrofóbicas.

10.3 Avaliação da presença do gene URAD no vetor plasmidial

A partir da transformação de *E. coli* DH5 α e realização da extração do DNA plasmidial foi obtido um estoque de 80 μ L contendo $188,533 \pm 3,647$ ng/ μ L. Com este estoque foi realizado o processo de digestão do plasmídeo utilizando as enzimas de restrição *Xho*I, *Xba*I e *Nde*I (Promega). Os resultados obtidos, demonstrados nos geis de agarose da **Figura 22**, indicam que o inserto completo, que apresenta 534 pares de base, se fazia presente no vetor, que apresenta 5357 pares de base.

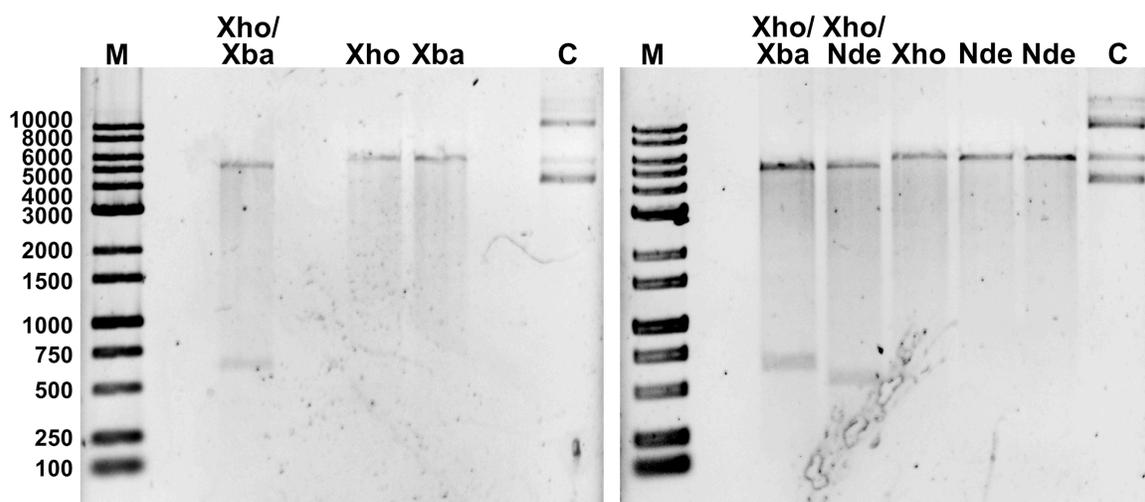


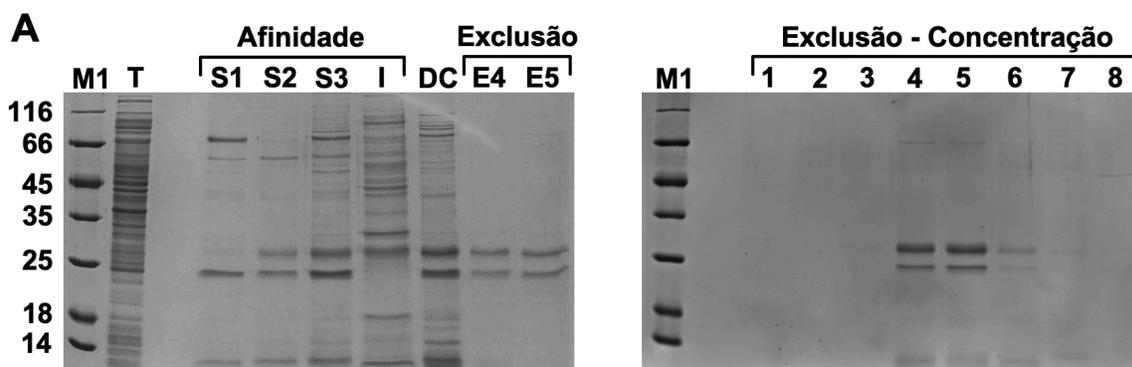
Figura 22. Digestão do plasmídeo. Geis de agarose nos quais foram aplicadas as amostras resultantes do protocolo de digestão com duas enzimas de restrição (*Xho*/*Xba* e *Xho*/*Nde*), com

apenas uma enzima (*Xho*, *Xba* e *Nde*) e com nenhuma enzima de restrição (*C* - controle). Na primeira linha o marcador de peso molecular (*1kb DNA Ladder RTU*).

10.4 OHCU descarboxilase de *H. sapiens* demonstrou atividade catalítica, realizando a descarboxilação do OHCU para gerar (S)-alantoína.

Tendo constatado a presença do inserto no vetor prosseguiu-se com a expressão da proteína recombinante. Primeiro, foi utilizada a *E. coli* BL21(DE3) pLys para expressar a OHCU descarboxilase de *H. sapiens*, tendo sido constatado, por meio do western blot, mostrado na **Figura 23B**, que a proteína estava sofrendo algum processo de clivagem, restando apenas uma fração de proteína íntegra, próxima aos 21 kDa, e sendo gerados fragmentos de massa molecular menor ou igual a 10 kDa.

Considerando que os fragmentos podem ser vistos em todas amostras aplicadas no gel e que ao tampão de lise é adicionado o inibidor de protease, pode-se supor que a proteína apresenta algum sítio suscetível a ação de proteases da *E. coli* BL21(DE3) pLys e, além disso, que uma proporção considerável das cadeias polipeptídicas geradas não foi enovelada da forma correta durante a expressão, o que favorece o processo de degradação da proteína. Esse processo de hidrólise da OHCU descarboxilase foi observado por Oh, J. *et al.* (2018) quando os autores tentaram expressar, além da alantoína sintase de *P. tricornutum*, a descarboxilase de diferentes organismos (OH *et al.*, 2018).



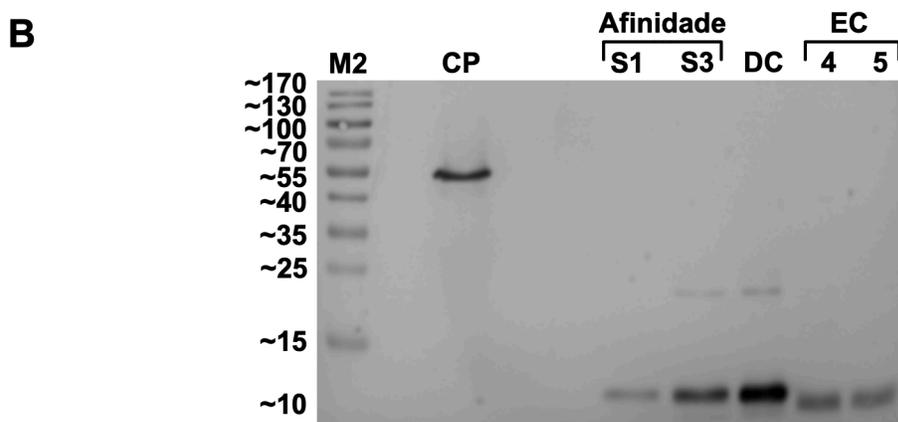
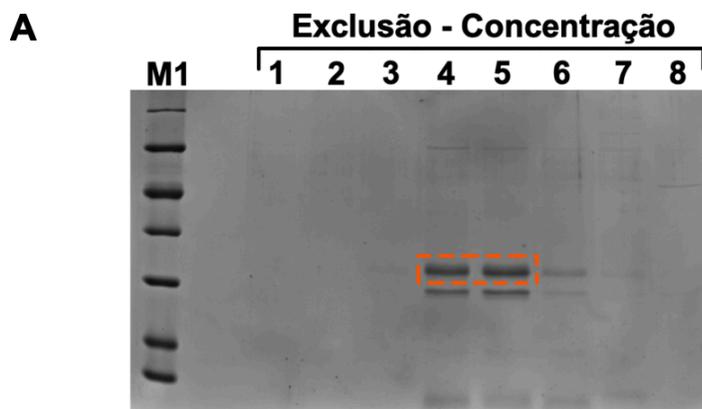


Figura 23. Expressão da OHCU descarboxilase de *H. sapiens* utilizando a *E. coli* BL21(DE3)pLys. A) Geis SDS-PAGE 12% nos quais foram aplicadas as amostras resultantes da expressão (T - total), da purificação por afinidade da fração solúvel (S1, S2 e S3) e insolúvel (I) obtidas após lise das células, da diálise e concentração (DC) das frações obtidas na afinidade, da purificação por exclusão (E4 e E5) e da concentração das amostras da exclusão (EC). B) Resultado do western blot, tendo sido utilizadas, além das amostras obtidas no processo de purificação, um controle positivo (CP - RBP ligada a HisTag).

Além disso, mesmo após a cromatografia por exclusão, não foi possível separar a proteína de interesse de uma proteína de *E. coli* de massa molecular próxima a 25 kDa (**Figura 24A**). Esta proteína de *E. coli* foi submetida ao protocolo de digestão e, então, sua sequência foi avaliada utilizando espectrometria de massa MALDI-TOF. A partir dos dados de espectrometria de massa, indicados nas **Figuras 24B** e **24C**, foi possível determinar que se tratava de uma proteína ligadora de DNA que atua na regulação da transcrição (DNA-binding transcriptional dual regulator CRP - ID UniProt P0ACJ8).



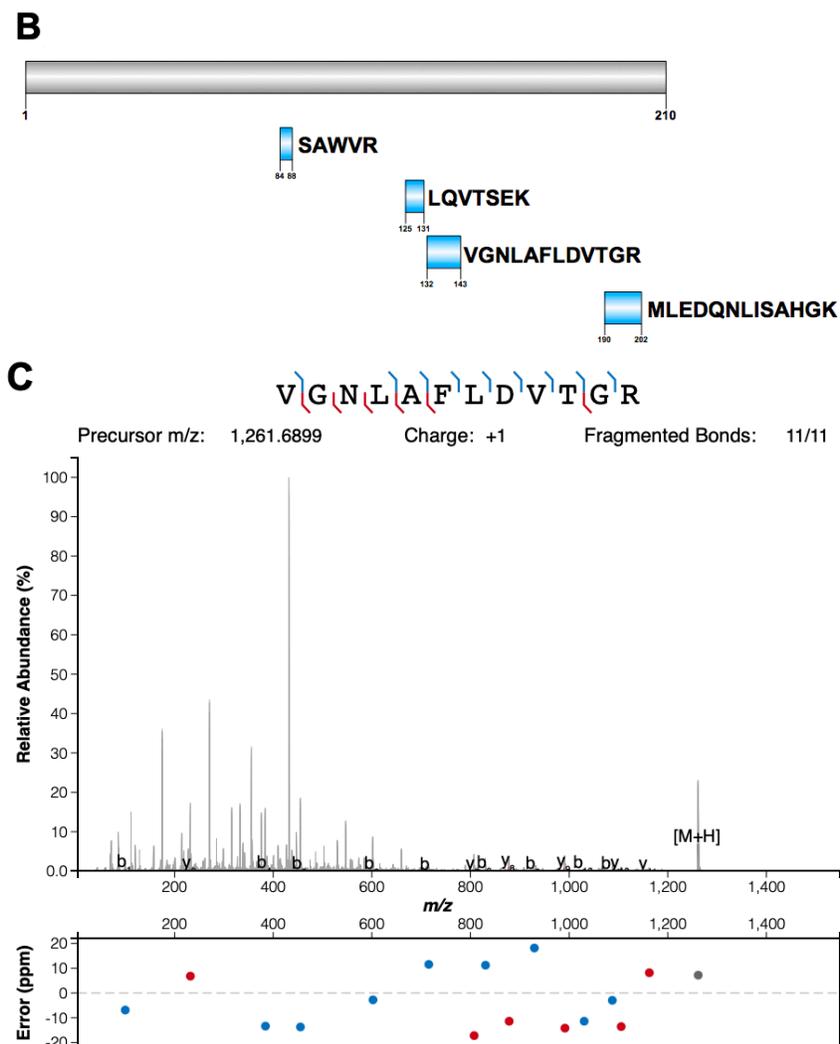


Figura 24. Avaliação da proteína contaminante de *E. coli* por espectrometria de massa MALDI-TOF. A) Gel SDS-PAGE 12% indicando a proteína que foi excisada e submetida ao protocolo de digestão por tripsina. B) Íons analisados por MALDI-TOF/MS e MALDI-TOF/MSMS após a digestão da proteína de *E. coli*. C) Resultado da fragmentação (MS/MS) do $[M + H]^+ = 1261.98$ m/z da proteína de *E. coli* digerida, sendo indicada a sequência do peptídeo, os íons de série b e y, assim como o erro associado.

Assim, considerando que esta proteína de *E. coli* dificilmente interferiria nas reações da via de degradação do ácido úrico, prosseguiu-se com os ensaios enzimáticos. Primeiro, foram coletados os espectros para soluções de ácido úrico e de alantoína, de forma a avaliar o perfis obtido para o substrato inicial da reação, o ácido úrico, e para o produto final das reações catalisadas pelo conjunto de enzimas uricase, HIUase e OHCU descarboxilase, ou seja, a alantoína.

Como pode ser visto pelos espectros na **Figura 25**, apesar de tanto o ácido úrico quanto a alantoína apresentarem um máximo de absorção próximo a 200 nm, os compostos

apresentam padrões de espectros bastante distintos. Enquanto a alantoína apresenta este único máximo de absorção próximo a 200 nm, o ácido úrico apresenta dois outros pontos com uma maior intensidade da absorção, um próximo a 235 nm e outro próximo a 295 nm.

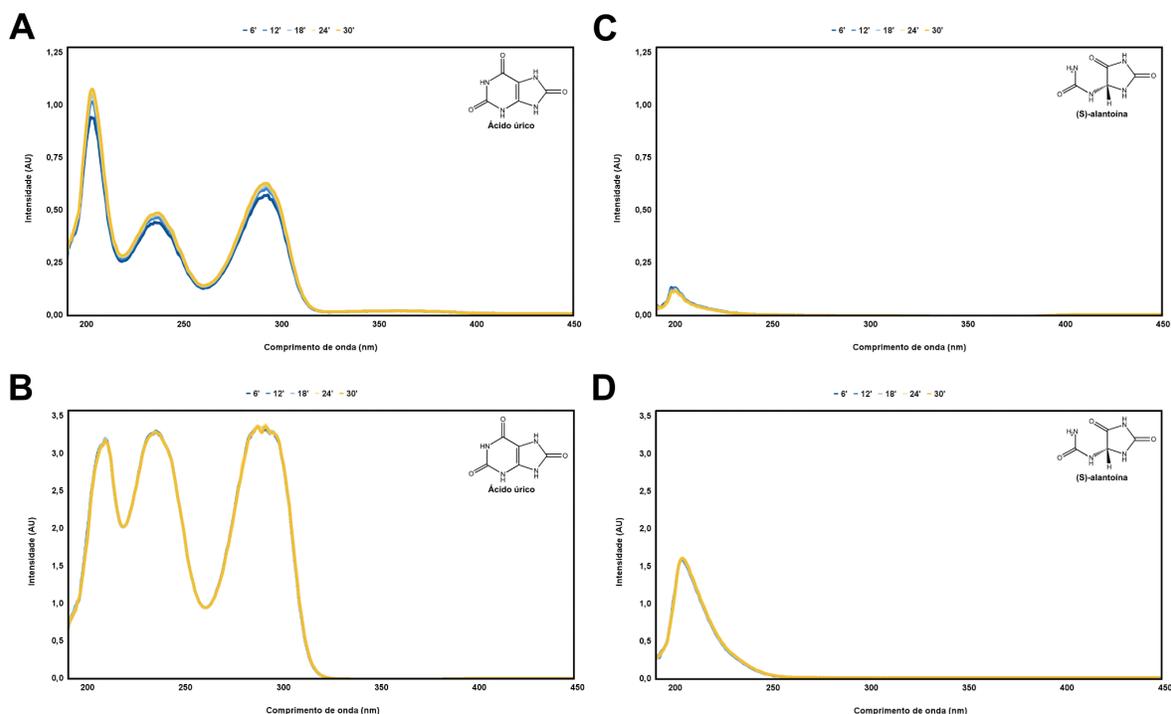


Figura 25. Perfis de absorção no UV do ácido úrico e da alantoína. Espectros obtidos para o ácido úrico a uma concentração de A) 30 μM e B) 300 μM e para a alantoína a uma concentração de C) 30 μM e D) 300 μM ao longo de 30 minutos (30') de experimento.

Tendo os espectros do do ácido úrico e da alantoína, foram realizados os ensaios enzimáticos utilizando a solução de ácido úrico, a uricase recombinante de *Candida sp.*, a HIUase recombinante de *H. seropedicae* e a fração purificada contendo a OHCU descarboxilase de *H. sapiens*. Considerando que na fração purificada contendo a descarboxilase havia outra proteína e que, portanto, a concentração calculada utilizando a absorbância a 280 nm correspondia ao somatório das concentrações das duas proteínas em solução, foram testadas duas condições, uma na qual a concentração de proteínas da fração purificada era 50 nM e outra na qual a concentração era 100 nM.

Foram coletados os espectros das quatro condições, 30 μM de ácido úrico e 50 nM de uricase; 30 μM de ácido úrico, 50 nM de uricase e 50 nM de HIUase; 30 μM de ácido úrico, 50 nM de uricase, 50 nM de HIUase e 50 nM da solução contendo descarboxilase; 30 μM de ácido úrico, 50 nM de uricase, 50 nM de HIUase e 100 nM da solução contendo

descarboxilase. Os espectros das diferentes condições foram coletados concomitantemente ao longo de uma hora de experimento (**Anexo XII**).

A partir dos ensaios enzimáticos, mostrados na **Figura 26**, podemos notar que, acompanhando a redução na intensidade nos comprimentos de onda 292, 302 e 256 nm, que são indicativos das atividades da uricase, da HIUase e da OHCU descarboxilase, respectivamente (FRENCH; EALICK, 2010; NYBORG *et al.*, 2016; SARMA *et al.*, 1999), a enzima de *H. sapiens* apresenta atividade catalítica, realizando a descarboxilação do OHCU para gerar a (S)-alantoína.

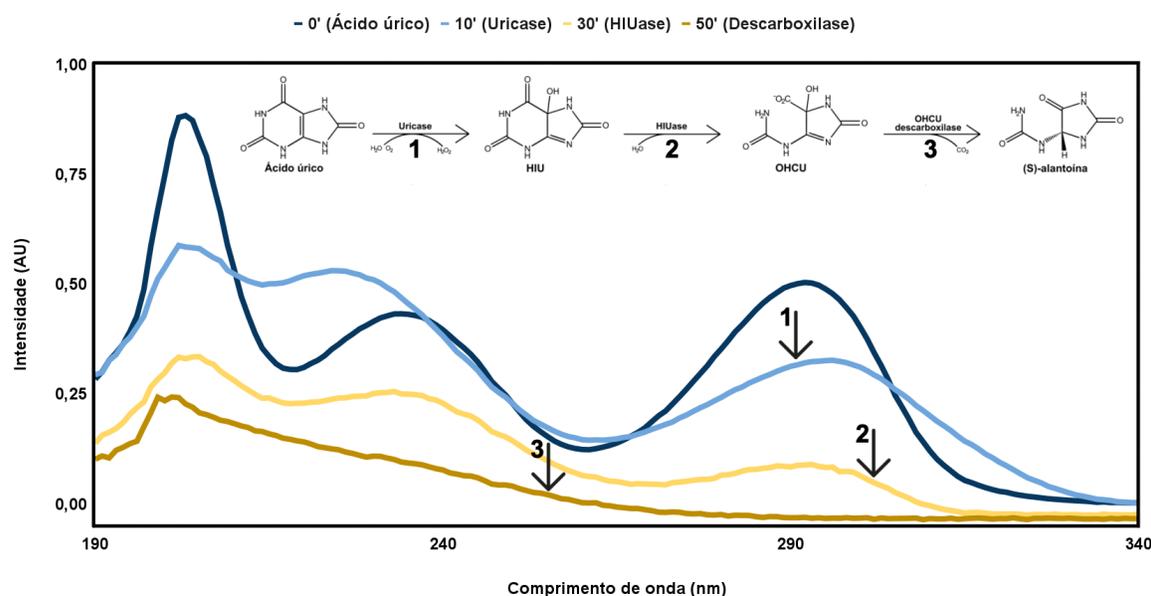


Figura 26. Ensaio enzimático. Resultado do ensaio enzimático conduzido por uma hora (60') utilizando 50 nM de uricase recombinante de *Candida sp.*, 50 nM de HIUase recombinante de *H. seropedicae* e 100 nM da mistura OHCU descarboxilase de *H. sapiens* e CRP de *E. coli*. As setas indicam os comprimentos de onda 292 (1), 302 (2) e 256 nm (3), que são indicativos das atividades da uricase, da HIUase e da OHCU descarboxilase, respectivamente.

10.5 Expressão da OHCU descarboxilase de *H. sapiens* utilizando a *E. coli* ArcticExpress (DE3) reduz a degradação da enzima

Considerando que o processo de degradação poderia ser devido ao mau enovelamento das cadeias polipeptídicas de OHCU descarboxilase de *H. sapiens* na bactéria, foi utilizada a *E. coli* ArcticExpress (DE3), que expressa chaperonas, que auxiliam no processo de enovelamento de cadeias polipeptídicas, para realizar a expressão da proteína recombinante.

A partir do SDS-PAGE (**Figura 27A**) e dos resultados do western blot (**Figura 27B**) é possível notar que o processo de degradação da proteína foi reduzido, mas não mitigado. Além disso, aparentemente estavam sendo gerados fragmentos diferentes dos observados na expressão utilizando a *E. coli* BL21(DE3)pLys, visto que são observadas proteínas com massa molecular próxima à da proteína íntegra, ou seja, próxima a 21 kDa. Além disso, vale ressaltar que os fragmentos parecem ter sido gerados após o processo de purificação por afinidade, durante a diálise, o que sugere a ocorrência da hidrólise da proteína em solução.

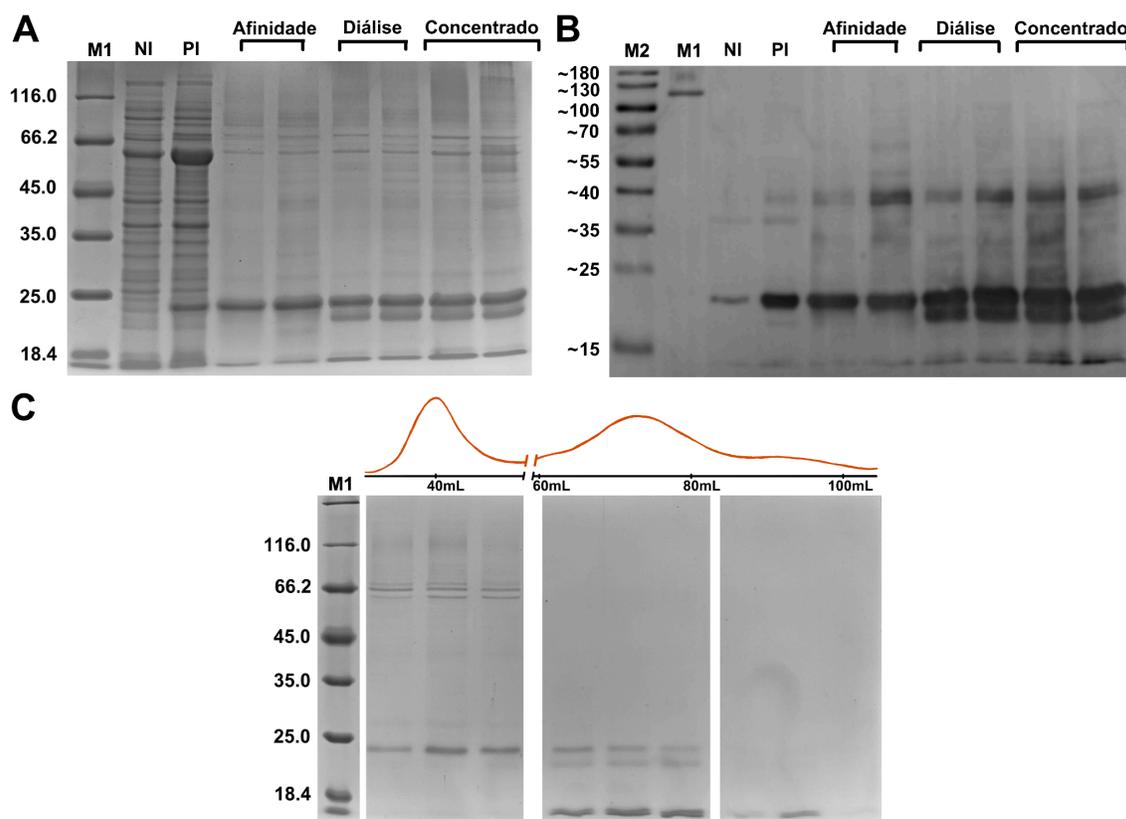


Figura 27. Expressão da OHCu descarboxilase de *H. sapiens* utilizando a *E. coli ArcticExpress (DE3)*. A) Gel SDS-PAGE 12% no qual foram aplicadas as amostras antes da indução da expressão (NI), após a indução à expressão (PI), após a purificação por afinidade, a diálise e a concentração da proteína recombinante. B) Resultado do western blot, no qual foram aplicadas as mesmas amostras. C) Frações obtidas após a cromatografia de exclusão por tamanho de acordo com o volume de eluição.

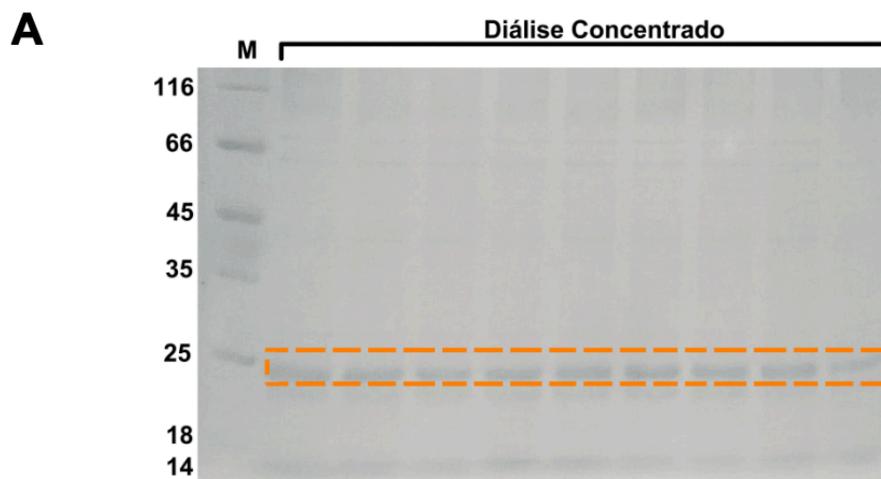
Para confirmar o resultado do western blot foi realizada a digestão da OHCu descarboxilase de *H. sapiens* e avaliação dos fragmentos gerados utilizando espectrometria de massa MALDI-TOF. Ao analisar os íons presentes no espectro de MALDI-TOF/MS foi observada uma correspondência destes com aqueles gerados na digestão teórica da OHCu

descarboxilase de *H. sapiens* utilizando a ferramenta PeptideMass do ExPASy, conforme indicado na **Tabela 7** (WILKINS *et al.*, 1997, 1998).

Tabela 7: Íons teóricos e experimentais obtidos na digestão da OHCU descarboxilase de *H. sapiens*

m/z dos peptídeos teóricos	m/z dos peptídeos experimentais	Erro experimental	Posição	Sequência
2120,0923	2119,7163	0,01774	63-81	NH ₂ -HFFAFIDALAQSGQEGILR-COOH
2099,9702	2099,5447	0,02027	22-40	NH ₂ -VNSMDLGEFVDVFGNATER-COOH
1325,6266	1325,2606	0,02762	82-93	NH ₂ -CHPDLAGSELQR-COOH
1124,6251	1124,2967	0,02921	131-140	NH ₂ -FGFPFVLAAR-COOH
1077,5687	1077,2474	0,02983	120-128	NH ₂ -LAELNAQYR-COOH
962,4901	962,1827	0,03195	94-102	NH ₂ -GTLTAESQR-COOH
817,4162	817,0858	0,04044	103-110	NH ₂ -EQSGAGLR-COOH
747,3631	746,9900	0,04995	111-117	NH ₂ -SLGADER-COOH
717,4141	716,9939	0,05861	165-171	NH ₂ -TALGEVK-COOH
700,4352	700,0147	0,06007	178-183	NH ₂ -LADLLR-COOH

Dentre os íons observados, 10 foram fragmentados (**Figura 28B**), sendo obtidos os espectros de MALDI-TOF/MSMS, utilizados para realizar o sequenciamento de cada um dos íons, conforme demonstrado para o íon $[M + H]^+ = 1124,62$ m/z na **Figura 28C**. Com estes resultados dos experimentos de espectrometria de massa foi possível constatar, por meio da elucidação parcial da estrutura primária, que a proteína recombinante se tratava realmente da OHCU descarboxilase de *H. sapiens*.



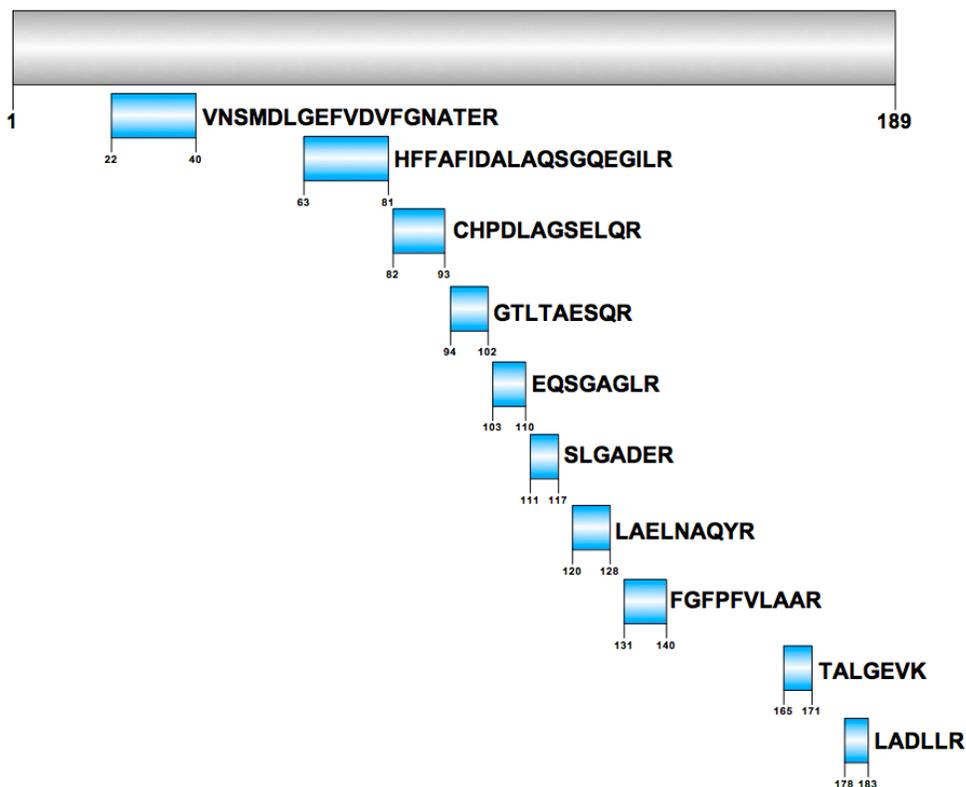
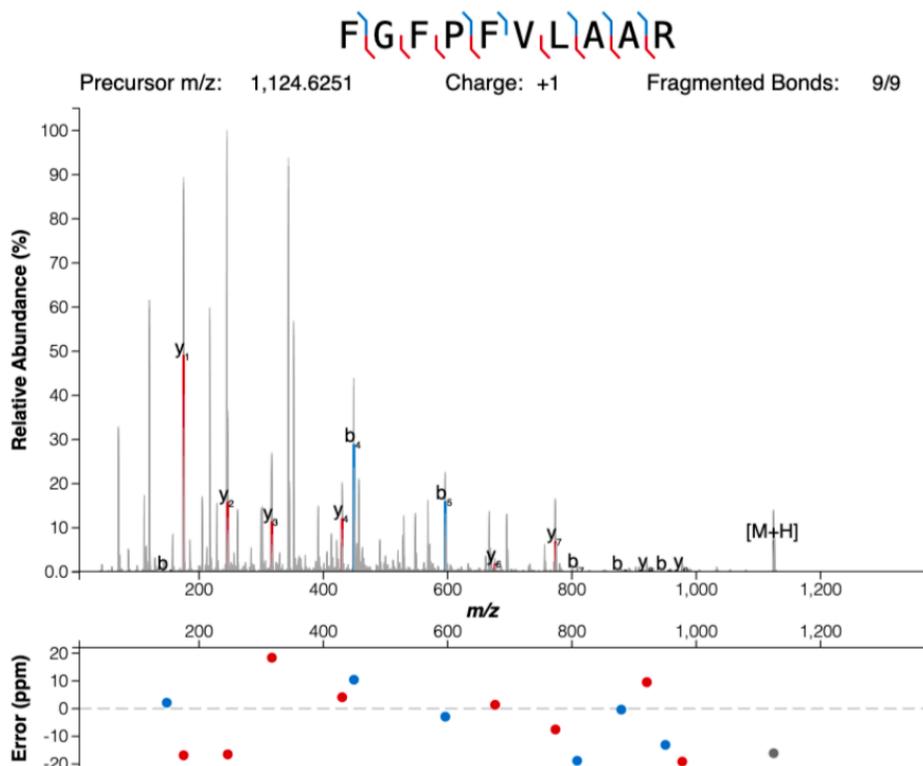
B**C**

Figura 28. Avaliação da OHCU descarboxilase de *H. sapiens* por espectrometria de massa MALDI-TOF. A) Gel SDS-PAGE indicando a proteína que foi excisada e submetida ao protocolo de digestão por tripsina B) Íons analisados por MALDI-TOF/MS e MALDI-TOF/MSMS após a digestão da OHCU descarboxilase de *H. sapiens*. O fragmento representado em amarelo foi o único

que foi gerado a partir de um erro de clivagem da tripsina. C) Resultado da fragmentação (MS/MS) do ion $[M + H]^+ = 1124,62 \text{ m/z}$ da OHCU descarboxilase de *H. sapiens* digerida, sendo indicada a sequência do peptídeo, os íons de série b e y, assim como o erro associado.

Por fim, os mesmos ensaios enzimáticos utilizando a uricase recombinante de *Candida sp.*, a HIUase recombinante de *H. seropedicae* e a fração purificada contendo a OHCU descarboxilase de *H. sapiens* foram repetidos e, além disso, considerando a presença de estruturas de OHCU descarboxilase em complexo com a alantoína que foram depositadas no PDB, foi realizado um ensaio no espectrofotômetro utilizando 100 nM de OHCU descarboxilase de *H. sapiens* juntamente com 300 μM de alantoína para verificar a ocorrência de interação entre a enzima e o produto da reação.

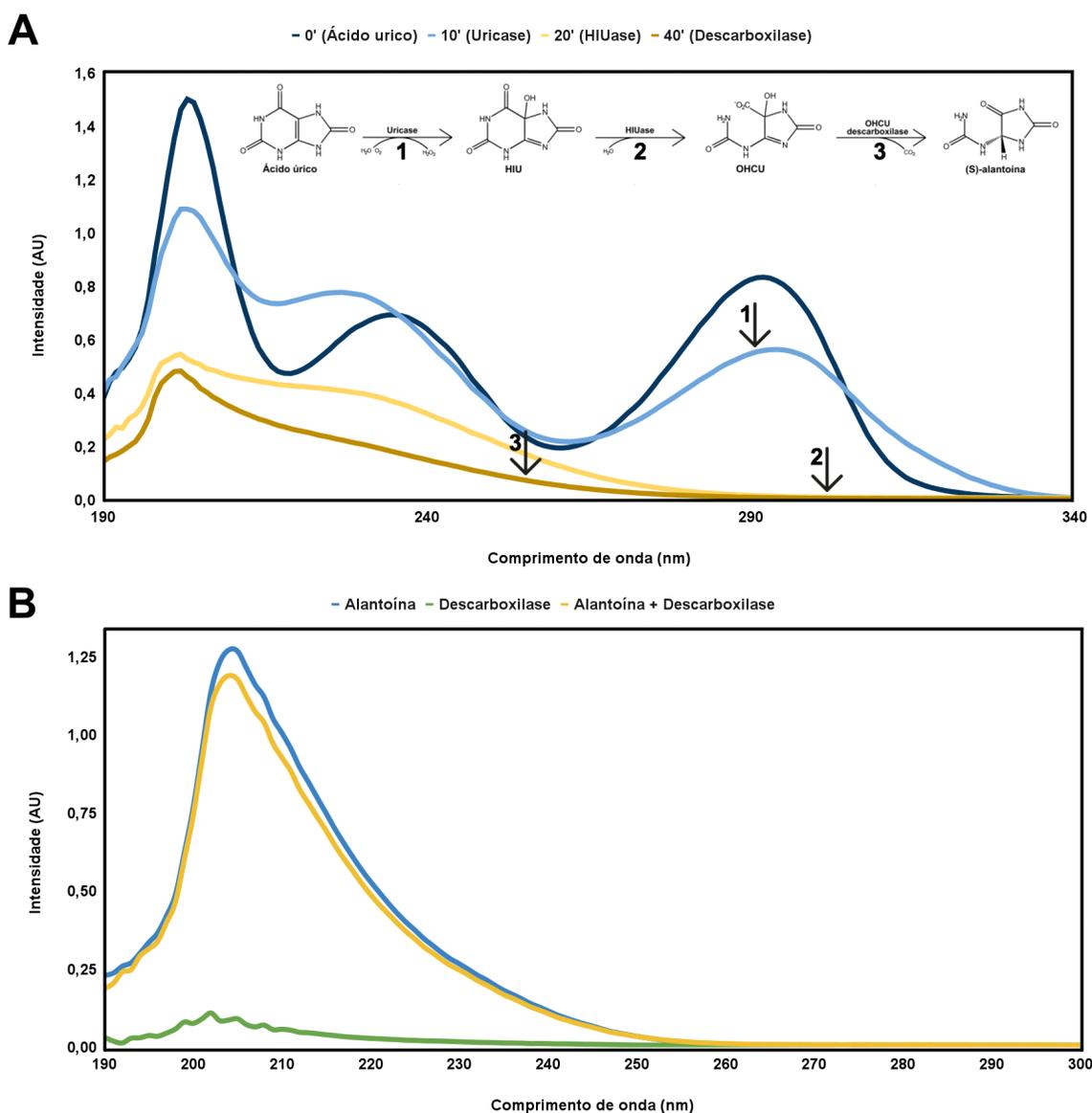


Figura 29. Ensaio enzimático e avaliação da interação com alantoína. A) Resultado do ensaio enzimático conduzido por uma hora (60') utilizando 50 nM de uricase recombinante de *Candida*

sp., 50 nM de HIUase recombinante de *Herbaspirillum seropedicae* e 100 nM de OHCU descarboxilase de *H. sapiens*. As setas indicam os comprimentos de onda 292 (1), 302 (2) e 256 nm (3), que são indicativos das atividades da uricase, da HIUase e da OHCU descarboxilase, respectivamente. B) Espectro obtido para alantoína (AL), OHCU descarboxilase de *H. sapiens* (URAD) e para a mistura alantoína e OHCU descarboxilase (AL + URAD).

Novamente podemos notar que, acompanhando a redução na intensidade nos comprimentos de onda 292, 302 e 256 nm (FRENCH; EALICK, 2010; NYBORG *et al.*, 2016; SARMA *et al.*, 1999), a enzima de *H. sapiens* apresenta atividade catalítica, realizando a descarboxilação do OHCU para gerar a (S)-alantoína (**Figura 29A**). Já no ensaio utilizando apenas a OHCU descarboxilase de *H. sapiens* e a alantoína, indicado na **Figura 29B**, foi possível observar uma redução da intensidade da absorção a 200 nm da alantoína, sugerindo uma interação entre a enzima e o produto da reação.

10.6 OHCU descarboxilase de *H. sapiens* assume um padrão de estrutura secundária predominantemente em α -hélice

Tendo sido realizada a avaliação da atividade e da estrutura primária da OHCU descarboxilase de *H. sapiens*, prosseguiu-se com a análise do seu padrão de enovelamento, de forma a verificar se o mesmo estaria condizente com os aqueles observados nas estruturas depositadas no PDB e com os modelos gerados utilizando o ColabFold, o Modeller e o iTASSER.

Primeiro, foram realizados experimentos de dicroísmo circular utilizando o espectropolarímetro, tendo sido coletados os espectros entre os comprimentos de onda 190 a 260 nm à 20°C, os espectros para os comprimentos de onda 220 e 222 nm submetendo a proteína ao aquecimento de 20 à 90°C e os espectros entre os comprimentos de onda 210 a 260 nm submetendo a proteína ao aquecimento de 20 à 100°C. Neste último caso não foram incluídos os comprimentos de onda de 190 a 209 nm para evitar que o valor de voltagem HT (*high-throughput*), que corresponde a voltagem sendo utilizada pelo detector, sofresse um aumento significativo, o que pode ocorrer com o aumento de temperatura.

Como mostrado nas **Figuras 30A e 30B**, os resultados do dicroísmo circular indicam que a OHCU descarboxilase de *H. sapiens* assume um padrão de α -hélice, apresentando pontos mínimos de elipticidade nos comprimentos de onda 208 e 222 nm (KELLY; JESS; PRICE, 2005). Já em relação a sua estabilidade térmica, ao avaliar o ponto de inflexão nos espectros coletados para os comprimentos de onda 220 e 222 nm (**Figura 30C**) e os perfis

obtidos nos espectros coletados entre 210 a 260 nm (**Figura 30D**), a enzima parece manter-se enovelada, quando aquecida, até atingir temperaturas entre 60 e 65°C, ocorrendo, a partir deste ponto, uma perda progressiva do padrão de α -hélice.

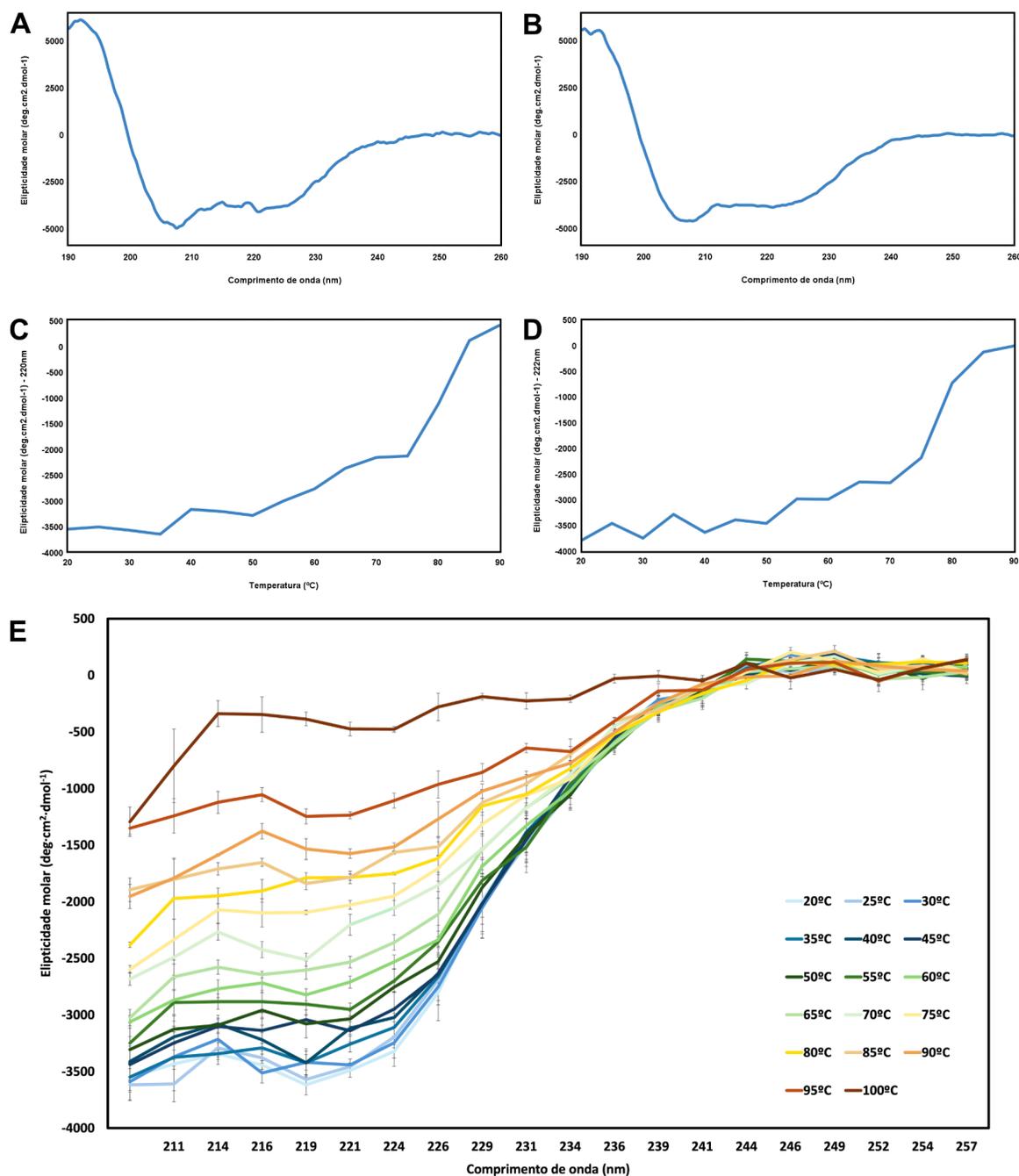


Figura 30. Experimentos de dicroísmo circular com OHCU descarboxilase de *H. sapiens*. Espectros entre os comprimentos de onda 190 a 260 nm à 20°C para solução A) 37,5 μM e B) 75 μM de OHCU descarboxilase de *H. sapiens*. Espectros para os comprimentos de onda C) 220 nm e D) 222 nm submetendo a proteína (37,5 μM) ao aquecimento de 20 à 90°C. E) Espectros entre os comprimentos de onda 210 a 260 nm submetendo a proteína (75 μM) ao aquecimento de 20 à 100°C. Os valores de elipticidade (eixo y) foram deconvoluídos levando em consideração a molaridade da amostra, o tamanho e a massa molecular da proteína e o caminho óptico da cubeta.

Para confirmar os resultados obtidos com os experimentos de dicroísmo circular, foi realizada a coleta do espectro ^1H 1D de ressonância magnética nuclear à 25°C . No espectro de RMN da OHCU descarboxilase de *H. sapiens*, indicado na **Figura 31**, podemos notar, próximo ao deslocamento químico de 10 ppm, os sinais característicos de hidrogênios mais desblindados presentes nos resíduos de aminoácidos aromáticos tirosina e triptofano, os quais se fazem presentes na enzima (Trp 33 e Tyr 111, além da Tyr presente na HisTag). Também se fazem presentes os sinais, abaixo de 0 ppm, dos hidrogênios mais blindados presentes em grupos metila de resíduos cuja cadeia lateral está voltada para o interior da estrutura da proteína. Além disso, há diversos sinais entre 0 a 2 ppm, que correspondem aos hidrogênios de resíduos de aminoácidos alifáticos (CAVANAGH, 2007).

Ademais, entre os deslocamentos químicos 4 e 3 ppm podemos notar os sinais dos hidrogênios α , ou seja, dos átomos de hidrogênio ligados ao carbono α dos resíduos de aminoácido, e entre 6 e 9 ppm estão os sinais dos hidrogênio amídicos, ou seja, dos átomos de hidrogênio ligados aos nitrogênios de grupamentos amina (CAVANAGH, 2007).

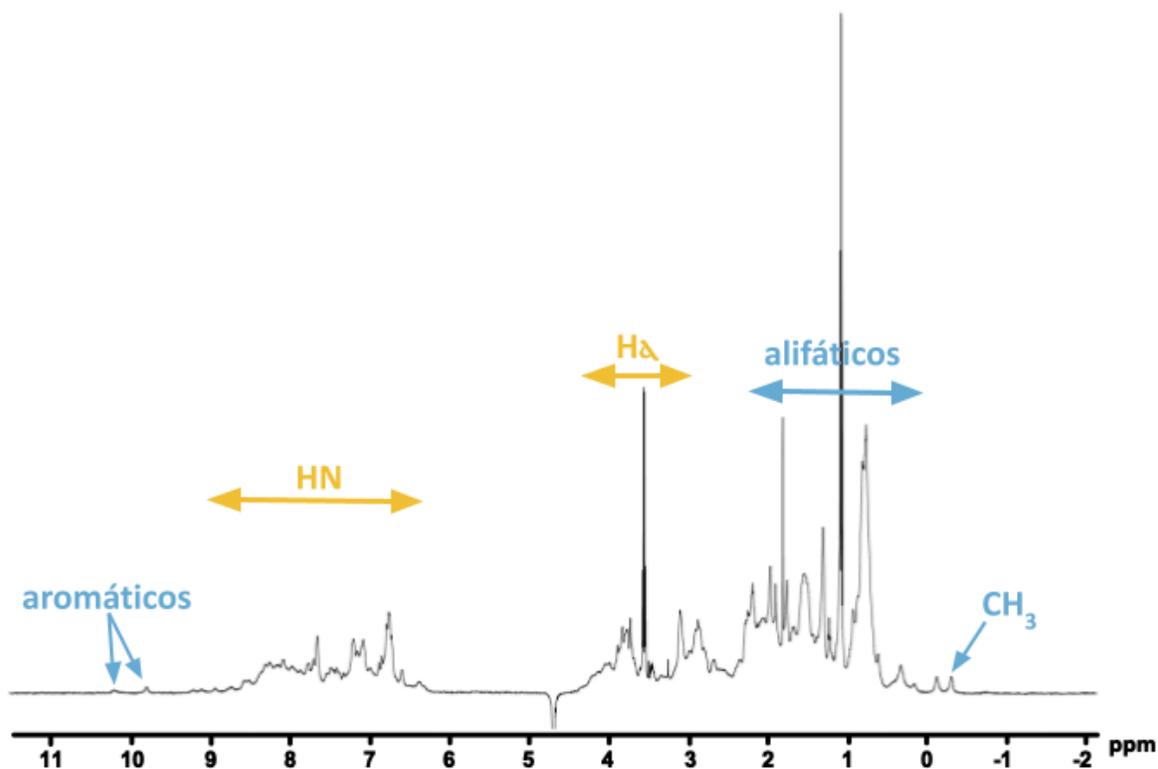


Figura 31. Avaliação da OHCU descarboxilase de *H. sapiens* por ressonância magnética nuclear. Espectro de RMN obtido para solução $100\ \mu\text{M}$ da OHCU descarboxilase de *H. sapiens*, sendo indicados os sinais característicos de hidrogênios de resíduos de aminoácidos aromáticos e alifáticos, de hidrogênios α e amídicos, assim como hidrogênios de grupos metila voltados para o interior da estrutura.

Assim, considerando que os sinais encontram-se bem distribuídos no espectro de RMN, os resultados da ressonância magnética nuclear corroboram os resultados obtidos no dicroísmo circular, indicando que a OHCU descarboxilase de *H. sapiens* está enovelada. Além disso, vale ressaltar que não foram observados sinais entre 5 e 6 ppm, o que é um indicativo da ausência de padrões de estrutura secundária em fitas- β , assim, apesar de não podermos afirmar, por RMN, que a proteína está enovelada segundo um padrão de α -hélice, podemos inferir que a proteína não apresenta fitas- β em sua estrutura.

11. CONCLUSÃO E PERSPECTIVAS

Os resultados do GTE_x, que mostram uma expressão tecido-específica da UOX, URAHP, URAD e ALLC, majoritariamente em órgãos que não estão envolvidos na degradação de purinas e na excreção do ácido úrico, juntamente com os resultados das análises de transcriptoma, que não indicaram expressão diferencial da UOX, URAHP, URAD e ALLC em afecções associadas ao acúmulo de ácido úrico, motiva uma reavaliação das funções destas regiões gênicas além das vias metabólicas relacionadas à hiperuricemia.

Considerando a possibilidade dos transcritos da UOX, URAHP, URAD e ALLC estarem envolvidos na regulação de vias não associadas à degradação de purinas e à excreção do ácido úrico, bem como na patogênese de doenças não relacionadas à hiperuricemia, faz-se necessária a análise de dados de transcriptoma utilizando dados de células de tecidos nos quais cada uma destas regiões gênicas é mais expressa, de forma a buscar entender seus papéis no organismo.

Os resultados obtidos nas análises dos produtos proteicos da URAD e da ALLC indicam a existência de uma OHC_U descarboxilase humana que mantém o padrão estrutural das enzimas desta família, sendo composta por α -hélices, e, além disso, cuja atividade também foi preservada. Contudo, as alterações que ocorreram na sequência da OHC_U descarboxilase de *H. sapiens* ao longo da evolução parecem ter levado ao surgimento de sítios mais suscetíveis à hidrólise no interior da cadeia polipeptídica. Assim, faz-se necessária a análise dos fragmentos gerados a partir da hidrólise da enzima, de forma a elucidar o ponto de clivagem na sequência da OHC_U descarboxilase de *H. sapiens*.

Somados à determinação do sítio de hidrólise, resultados de análises futuras utilizando o PFstats serão essenciais para o desenho de uma cadeia polipeptídica análoga da OHC_U descarboxilase de *H. sapiens* que apresente uma maior estabilidade e, além disso, uma maior eficiência catalítica. Assim, os *insights* sobre a estabilidade e função da OHC_U descarboxilase podem ser aplicados na engenharia de proteínas para aplicações biotecnológicas. Contudo, para isso também são necessárias investigações utilizando técnicas complementares, como cristalografia de proteína de difração de raios-X ou RMN 2D e 3D, de forma a obter uma visão mais detalhada da estrutura da OHC_U descarboxilase.

12. REFERÊNCIAS

- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. *Journal of Molecular Biology*, v. 215, n. 3, p. 403–410, out. 1990.
- ALVAREZ-LARIO, B.; MACARRON-VICENTE, J. Uric acid and evolution. *Rheumatology*, v. 49, n. 11, p. 2010–2015, 1 nov. 2010.
- BEN SALEM, C. *et al.* Drug-induced hyperuricaemia and gout. *Rheumatology*, p. kew293, 7 ago. 2016.
- BERMAN, H. M. The Protein Data Bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242, 1 jan. 2000.
- BRADEMAN, D. R. *et al.* Interactive Peptide Spectral Annotator: A Versatile Web-based Tool for Proteomic Applications. ***Molecular & Cellular Proteomics***, v. 18, n. 8, p. S193–S201, ago. 2019.
- BRIGHTWELL, C. R. *et al.* Muscle fibrosis and maladaptation occur progressively in CKD and are rescued by dialysis. *JCI Insight*, v. 6, n. 24, p. e150112, 22 dez. 2021.
- CARRIJO DE OLIVEIRA, L. *et al.* Reenacting the Birth of a Function: Functional Divergence of HIUases and Transthyretins as Inferred by Evolutionary and Biophysical Studies. *Journal of Molecular Evolution*, v. 89, n. 6, p. 370–383, jul. 2021.
- CAVANAGH, J. (ED.). *Protein NMR spectroscopy: principles and practice*. 2nd ed. Amsterdam ; Boston: Academic Press, 2007.
- CENDRON, L. *et al.* The Structure of 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase Provides Insights into the Mechanism of Uric Acid Degradation. *Journal of Biological Chemistry*, v. 282, n. 25, p. 18182–18189, jun. 2007.
- CHEETHAM, S. W.; FAULKNER, G. J.; DINGER, M. E. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics*, v. 21, n. 3, p. 191–201, mar. 2020.
- CHI, K. *et al.* LncRNA-HOTAIR promotes endothelial cell pyroptosis by regulating the miR-22/NLRP3 axis in hyperuricaemia. *Journal of Cellular and Molecular Medicine*, v. 25, n. 17, p. 8504–8521, set. 2021.
- COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, v. 2, n. 9, p. 1511–1519, set. 1993.
- CULLEN, M. R. *et al.* Disease-Specific Health Disparities: A Targeted Review Focusing on Race and Ethnicity. *Healthcare*, v. 10, n. 4, p. 603, 23 mar. 2022.
- CUNNINGHAM, F. *et al.* Ensembl 2022. *Nucleic Acids Research*, v. 50, n. D1, p. D988–D995, 7 jan. 2022.
- DE SOUZA RUGOLO, L. M. S.; BENTLIN, M. R.; TRINDADE, C. E. P. Preeclampsia: Effect on the Fetus and Newborn. *NeoReviews*, v. 12, n. 4, p. e198–e206, 1 abr. 2011.

DEVESON, I. W. *et al.* Universal Alternative Splicing of Noncoding Exons. *Cell Systems*, v. 6, n. 2, p. 245–255.e5, fev. 2018.

FONSECA-JÚNIOR, N. J. *et al.* PFstats: A Network-Based Open Tool for Protein Family Analysis. *Journal of Computational Biology*, v. 25, n. 5, p. 480–486, maio 2018.

FRANKISH, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, v. 47, n. D1, p. D766–D773, 8 jan. 2019.

FRANKISH, A. *et al.* GENCODE 2021. *Nucleic Acids Research*, v. 49, n. D1, p. D916–D923, 8 jan. 2021.

FREEDMAN, A. H.; GASPAR, J. M.; SACKTON, T. B. Short paired-end reads trump long single-end reads for expression analysis. *BMC Bioinformatics*, v. 21, n. 1, p. 149, dez. 2020.

FRENCH, J. B.; EALICK, S. E. Structural and Mechanistic Studies on *Klebsiella pneumoniae* 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase. *Journal of Biological Chemistry*, v. 285, n. 46, p. 35446–35454, nov. 2010.

GIACCHERINI, M. *et al.* A scan of all coding region variants of the human genome, identifies 13q12.2-rs9579139 and 15q24.1-rs2277598 as novel risk loci for pancreatic ductal adenocarcinoma. *Carcinogenesis*, v. 44, n. 8–9, p. 642–649, 2 dez. 2023.

GRIFFEN, T. L. *et al.* Multivariate transcriptome analysis identifies networks and key drivers of chronic lymphocytic leukemia relapse risk and patient survival. *BMC Medical Genomics*, v. 14, n. 1, p. 171, dez. 2021.

HU, J. *et al.* Long non-coding RNA ANRIL-mediated inflammation response is involved in protective effect of rhein in uric acid nephropathy rats. *Cell & Bioscience*, v. 9, n. 1, p. 11, dez. 2019.

HU, Y. *et al.* NMR-Based Methods for Protein Analysis. *Analytical Chemistry*, v. 93, n. 4, p. 1866–1879, 2 fev. 2021.

HUANG, Q. *et al.* PaxDb 5.0: Curated Protein Quantification Data Suggests Adaptive Proteome Changes in Yeasts. *Molecular & Cellular Proteomics*, v. 22, n. 10, p. 100640, out. 2023.

JIANG, G. *et al.* A comprehensive workflow for optimizing RNA-seq data analysis. *BMC Genomics*, v. 25, n. 1, p. 631, 24 jun. 2024.

JIANG, X.; HOLMES, C.; MCVEAN, G. The impact of age on genetic risk for common diseases. *PLOS Genetics*, v. 17, n. 8, p. e1009723, 26 ago. 2021.

JUMPER, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*, v. 596, n. 7873, p. 583–589, 26 ago. 2021.

KEEBAUGH, A. C.; THOMAS, J. W. The Evolutionary Fate of the Genes Encoding the Purine Catabolic Enzymes in Hominoids, Birds, and Reptiles. *Molecular Biology and Evolution*, v. 27, n. 6, p. 1359–1369, 1 jun. 2010.

KELLY, S. M.; JESS, T. J.; PRICE, N. C. How to study proteins by circular dichroism.

Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, v. 1751, n. 2, p. 119–139, ago. 2005.

KESNER, J. S. *et al.* Noncoding translation mitigation. *Nature*, v. 617, n. 7960, p. 395–402, 11 maio 2023.

KIM, K.; PARK, J.; RHEE, S. Structural and Functional Basis for (S)-Allantoin Formation in the Ureide Pathway. *Journal of Biological Chemistry*, v. 282, n. 32, p. 23457–23464, ago. 2007.

KIM, S. *et al.* PubChem 2023 update. *Nucleic Acids Research*, v. 51, n. D1, p. D1373–D1380, 6 jan. 2023.

KWAN, T.; THOMPSON, S. R. Noncanonical Translation Initiation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, v. 11, n. 4, p. a032672, abr. 2019.

LASKOWSKI, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, v. 26, n. 2, p. 283–291, 1 abr. 1993.

LASKOWSKI, R. A. *et al.* PDBsum: Structural summaries of PDB entries. *Protein Science*, v. 27, n. 1, p. 129–134, jan. 2018.

LEINONEN, R. *et al.* The Sequence Read Archive. *Nucleic Acids Research, Database*, p. D19–D21, 1 jan. 2011.

LEMONS, R. P., *et al.* Evolution-based protein engineering: Functional switching between transthyretins and 5-hydroxyisourate hydrolases. *Journal of Biomolecular Structure and Dynamics*, 2024. [article in press]

LI, D. *et al.* MALDI-TOF Mass Spectrometry in Clinical Analysis and Research. *ACS Measurement Science Au*, v. 2, n. 5, p. 385–404, 19 out. 2022.

LI, S.-X. *et al.* Octameric Structure of the Human Bifunctional Enzyme PAICS in Purine Biosynthesis. *Journal of Molecular Biology*, v. 366, n. 5, p. 1603–1614, mar. 2007.

LI, W.; LI, Y. The pseudogene URAHP promotes proliferation and regulates the pathogenesis of preeclampsia. *American Journal of Translational Research*, v. 12(8), p. 4715–4727, 15 ago. 2020.

LIU, X. *et al.* Extracellular vesicles isolated from hyperuricemia patients might aggravate airway inflammation of COPD via senescence-associated pathway. *Journal of Inflammation*, v. 19, n. 1, p. 18, 2 nov. 2022.

LOVE, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLOS Computational Biology*, v. 16, n. 2, p. e1007664, 25 fev. 2020.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, n. 12, p. 550, dez. 2014.

LÜSCHER, B. P. *et al.* Hyperuricemia during Pregnancy Leads to a Preeclampsia-Like Phenotype in Mice. *Cells*, v. 11, n. 22, p. 3703, 21 nov. 2022.

MACIEJEWSKI, M. W. *et al.* NMRbox: A Resource for Biomolecular NMR Computation. *Biophysical Journal*, v. 112, n. 8, p. 1529–1534, abr. 2017.

MALLAT, S. G. *et al.* Hyperuricemia, Hypertension, and Chronic Kidney Disease: an Emerging Association. *Current Hypertension Reports*, v. 18, n. 10, p. 74, out. 2016.

MATTICK, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*, 3 jan. 2023.

MELÉ, M. *et al.* Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Research*, v. 27, n. 1, p. 27–37, jan. 2017.

MIGENDA, N.; MÖLLER, R.; SCHENCK, W. Adaptive dimensionality reduction for neural network-based online principal component analysis. *PLOS ONE*, v. 16, n. 3, p. e0248896, 30 mar. 2021.

MIRDITA, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods*, v. 19, n. 6, p. 679–682, jun. 2022.

MIRDITA, M.; STEINEGGER, M.; SÖDING, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*, v. 35, n. 16, p. 2856–2858, 15 ago. 2019.

MISTRY, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research*, v. 49, n. D1, p. D412–D419, 8 jan. 2021.

MONTAGNANA, M. *et al.* The role of ethnicity, age and gender in venous thromboembolism. *Journal of Thrombosis and Thrombolysis*, v. 29, n. 4, p. 489–496, maio 2010.

NELSON, D. L.; COX, M. M.; LEHNINGER, A. L. *Lehninger principles of biochemistry*. Seventh edition ed. New York NY: W.H. Freeman, 2017.

NOTREDAME, C.; HIGGINS, D. G.; HERINGA, J. T-coffee: a novel method for fast and accurate multiple sequence alignment 1 Edited by J. Thornton. *Journal of Molecular Biology*, v. 302, n. 1, p. 205–217, set. 2000.

NYBORG, A. C. *et al.* A Therapeutic Uricase with Reduced Immunogenicity Risk and Improved Development Properties. *PLOS ONE*, v. 11, n. 12, p. e0167935, 21 dez. 2016.

OH, J. *et al.* Diatom Allantoin Synthase Provides Structural Insights into Natural Fusion Protein Therapeutics. *ACS Chemical Biology*, v. 13, n. 8, p. 2237–2246, 17 ago. 2018.

PATRO, R. *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, v. 14, n. 4, p. 417–419, abr. 2017.

PAYSAN-LAFOSSE, T. *et al.* InterPro in 2022. *Nucleic Acids Research*, v. 51, n. D1, p. D418–D427, 6 jan. 2023.

PESSOA, J. *et al.* Functional characterization of *Arabidopsis thaliana* transthyretin-like protein. *BMC Plant Biology*, v. 10, n. 1, p. 30, dez. 2010.

PETERSEN, M. C.; SHULMAN, G. I. *Mechanisms of Insulin Action and Insulin*

Resistance. *Physiological Reviews*, v. 98, n. 4, p. 2133–2223, 1 out. 2018.

PETTERSEN, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, v. 25, n. 13, p. 1605–1612, out. 2004.

PIÑERO, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*, p. gkz1021, 4 nov. 2019.

PRAPUNPOJ, P. *et al.* Evolution of structure, ontogeny of gene expression, and function of *Xenopus laevis* transthyretin. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, v. 279, n. 6, p. R2026–R2041, 1 dez. 2000.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, v. 7, n. 1, p. 95–99, jul. 1963.

RAMAZZINA, I. *et al.* Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. *Nature Chemical Biology*, v. 2, n. 3, p. 144–148, mar. 2006.

REYES-PALOMARES, A. *et al.* Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nature Communications*, v. 11, n. 1, p. 1673, 3 abr. 2020.

RICHARDSON, S. J. Tweaking the Structure to Radically Change the Function: The Evolution of Transthyretin from 5-Hydroxyisourate Hydrolase to Triiodothyronine Distributor to Thyroxine Distributor. *Frontiers in Endocrinology*, v. 5, 11 fev. 2015.

RION, N.; RÜEGG, M. A. LncRNA-encoded peptides: More than translational noise? *Cell Research*, v. 27, n. 5, p. 604–605, maio 2017.

ROCK, K. L.; KATAOKA, H.; LAI, J.-J. Uric acid as a danger signal in gout and its comorbidities. *Nature Reviews Rheumatology*, v. 9, n. 1, p. 13–23, jan. 2013.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>. , [s.d.].

ŠALI, A.; BLUNDELL, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, v. 234, n. 3, p. 779–815, dez. 1993.

SAMUELS, M. L.; WITMER, J. A.; SCHAFFNER, A. A. *Statistics for the life sciences*. Fifth edition. Global edition ed. Boston Columbus Indianapolis: Pearson, 2016.

SARMA, A. D. *et al.* Identification and Purification of Hydroxyisourate Hydrolase, a Novel Ureide-metabolizing Enzyme. *Journal of Biological Chemistry*, v. 274, n. 48, p. 33863–33865, nov. 1999.

SCHULTZ, A. C.; NYGAARD, P.; SAXILD, H. H. Functional Analysis of 14 Genes That Constitute the Purine Catabolic Pathway in *Bacillus subtilis* and Evidence for a Novel Regulon Controlled by the PucR Transcription Activator. *Journal of Bacteriology*, v. 183, n. 11, p. 3293–3302, jun. 2001.

SHEN, M.; SALI, A. Statistical potential for assessment and prediction of protein

structures. *Protein Science*, v. 15, n. 11, p. 2507–2524, nov. 2006.

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, jan. 2011.

SONESON, C.; LOVE, M. I.; ROBINSON, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, v. 4, p. 1521, 30 dez. 2015.

STAMP, L. K. *et al.* Management of gout in chronic kidney disease: a G-CAN Consensus Statement on the research priorities. *Nature Reviews Rheumatology*, v. 17, n. 10, p. 633–641, out. 2021.

STATELLO, L. *et al.* Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, v. 22, n. 2, p. 96–118, fev. 2021.

STEVENSON, W. S. *et al.* Deficiency of 5-hydroxyisourate hydrolase causes hepatomegaly and hepatocellular carcinoma in mice. *Proceedings of the National Academy of Sciences*, v. 107, n. 38, p. 16625–16630, 21 set. 2010.

SZKLARCZYK, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, v. 51, n. D1, p. D638–D646, 6 jan. 2023.

TAKAHASHI, M. *et al.* Fetal growth restriction as the initial finding of preeclampsia is a clinical predictor of maternal and neonatal prognoses: a single-center retrospective study. *BMC Pregnancy and Childbirth*, v. 21, n. 1, p. 678, dez. 2021.

The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on dbGaP accession number phs000424.v8.p2 on 09/03/2023. , [s.d.].

THE UNIPROT CONSORTIUM *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, v. 51, n. D1, p. D523–D531, 6 jan. 2023.

TORRES, R. J.; PUIG, J. G. Hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency: Lesch-Nyhan syndrome. *Orphanet Journal of Rare Diseases*, v. 2, n. 1, p. 48, dez. 2007.

ULE, J.; BLENCOWE, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, v. 76, n. 2, p. 329–345, out. 2019.

VASILIOU, V. *et al.* ALDH16A1 is a novel non-catalytic enzyme that may be involved in the etiology of gout via protein–protein interactions with HPRT1. *Chemico-Biological Interactions*, v. 202, n. 1–3, p. 22–31, fev. 2013.

VOET, D.; VOET, J. G. *Biochemistry*. 4. ed ed. Hoboken, NJ: Wiley, 2011.

WICKHAM, H. *ggplot2: elegant graphics for data analysis*. Second edition ed. Switzerland: Springer, 2016.

WILKINS, M. R. *et al.* Detailed peptide characterization using PEPTIDEMASS – a World-Wide-Web-accessible tool. *ELECTROPHORESIS*, v. 18, n. 3–4, p. 403–408, jan. 1997.

WILKINS, M. R. *et al.* Protein Identification and Analysis Tools in the ExPASy Server. Em: LINK, A. J. (Ed.). 2-D Proteome Analysis Protocols. New Jersey: Humana Press, 1998. v. 112p. 531–552.

WU, Y. *et al.* Association of uric acid with progression to preeclampsia and development of adverse conditions in gestational hypertensive pregnancies. *American Journal of Hypertension*, v. 25, n. 6, p. 711–717, 1 jun. 2012.

XING, J. *et al.* LncRNA-Encoded Peptide: Functions and Predicting Methods. *Frontiers in Oncology*, v. 10, p. 622294, 14 jan. 2021.

XIONG, Q.; LIU, J.; XU, Y. Effects of Uric Acid on Diabetes Mellitus and Its Chronic Complications. *International Journal of Endocrinology*, v. 2019, p. 1–8, 13 out. 2019.

XU, Q. *et al.* Crystal structure of an allantoinase (YIR029W) from *Saccharomyces cerevisiae* at 2.4 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, v. 56, n. 3, p. 619–624, 15 ago. 2004.

YANG, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, v. 12, n. 1, p. 7–8, jan. 2015.

YANG, P.-C.; MAHMOOD, T. Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, v. 4, n. 9, p. 429, 2012.

YANG, X. *et al.* Landscape of Dysregulated Placental RNA Editing Associated With Preeclampsia. *Hypertension*, v. 75, n. 6, p. 1532–1541, jun. 2020.

ZANOTTI, G. *et al.* Structure of Zebra fish HIUase: Insights into Evolution of an Enzyme to a Hormone Transporter. *Journal of Molecular Biology*, v. 363, n. 1, p. 1–9, out. 2006.

ZHU, A. *et al.* Nonparametric expression analysis using inferential replicate counts. *Nucleic Acids Research*, v. 47, n. 18, p. e105–e105, 10 out. 2019.

ZHU, A.; IBRAHIM, J. G.; LOVE, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, v. 35, n. 12, p. 2084–2092, 1 jun. 2019.

13. ANEXOS

Anexo I: Organização dos éxons e íntrons dos transcritos da UOX, URAHP, URAD e ALLC

Gene	Transcrito	Exon / Início / Fim / Número de bases (bp)
Urato Oxidase (UOX) Cromossomo 1: 84, 363, 706-84, 397, 831 fita reversa	UOX-201	Exon 1 84397804 84397831 28 bp Exon 2 84384598 84384811 214 bp Exon 3 84381994 84382108 115 bp Exon 4 84381814 84381892 79 bp Exon 5 84372043 84372229 187 bp Exon 6 84369910 84370027 118 bp Exon 7 84367657 84367744 88 bp Exon 8 84365428 84365497 70 bp
	UOX-202	Exon 1 84384593 84384801 209 bp Exon 2 84381814 84381892 79 bp Exon 3 84372042 84372228 187 bp Exon 4 84369904 84370027 124 bp Exon 5 84367661 84367740 80 bp
	UOX-203	Exon 1 84365809 84365841 33 bp Exon 2 84365589 84365681 93 bp Exon 3 84364958 84365495 438 bp
	UOX-204	Exon 1 84381814 84381891 78 bp Exon 2 84372042 84372228 187 bp Exon 3 84369904 84370027 124 bp Exon 4 84367650 84367740 91 bp Exon 5 84365589 84365681 93 bp Exon 6 84364389 84365495 1107 bp
	UOX-205	Exon 1 84370383 84370400 18 bp Exon 2 84369904 84370027 124 bp Exon 3 84367650 84367740 91 bp Exon 4 84363706 84364016 311 bp

Gene	Transcrito	Exon / Início / Fim / Número de bases (bp)
5-hidroxiisourato hidrolase (URAHP) Cromossomo 16: 90, 043, 914-90, 047, 597 fita reversa	URAHP-201	Exon 1 90047538 90047597 60 bp Exon 2 90046957 90047096 140 bp Exon 3 90044502 90044623 122 bp Exon 4 90043914 90043989 76 bp
	URAHP-202	Exon 1 90044502 90044610 109 bp Exon 2 90044114 90044233 120 bp Exon 3 90043429 90043989 561 bp
	URAHP-203	Exon 1 90047538 90047625 88 bp Exon 2 90046957 90047096 140 bp Exon 3 90044502 90044623 122 bp Exon 4 90043648 90043989 342 bp Exon 5 90039761 90040417 657 bp

	URAHP-204	Exon 1 90047743 90047773 31 bp Exon 2 90046957 90047096 140 bp Exon 3 90044502 90044623 122 bp Exon 4 90042909 90043989 1088 bp
	URAHP-205	Exon 1 90046957 90047190 234 bp Exon 2 90044502 90044623 122 bp Exon 3 90043802 90043989 188 bp

Gene	Transcrito	Exon / Início / Fim / Número de bases (bp)
2-oxo-4-hidroxi-4-carboxi-5-ureidoimidazolina descarboxilase (URAD) Cromossomo 13: 27, 977, 717-27, 988, 693 fita reversa	URAD-201	Exon 1 27988463 27988654 192 bp Exon 2 27977714 27978452 739 bp

Gene	Transcrito	Exon / Início / Fim / Número de bases (bp)
Alantoicase (ALLC) Cromossomo 2: 3, 658, 200-3, 702, 671 fita principal	ALLC-201	Exon 1 3658195 3658294 100 bp Exon 2 3671096 3671190 95 bp Exon 3 3674075 3674125 51 bp Exon 4 3678468 3678555 88 bp Exon 5 3679869 3679994 126 bp Exon 6 3681634 3681713 80 bp Exon 7 3682942 3683074 133 bp Exon 8 3695717 3695872 156 bp Exon 9 3696275 3696348 74 bp Exon 10 3697348 3697456 109 bp Exon 11 3701512 3701636 125 bp Exon 12 3702363 3702671 309 bp
	ALLC-202	Exon 1 3695091 3695265 175 bp Exon 2 3695717 3695872 156 bp Exon 3 3696275 3696348 74 bp Exon 4 3697348 3697456 109 bp Exon 5 3701512 3701636 125 bp Exon 6 3702363 3702448 86 bp
	ALLC-203	Exon 1 3694621 3695872 1252 bp Exon 2 3696275 3696348 74 bp Exon 3 3697348 3697456 109 bp Exon 4 3701512 3701636 125 bp Exon 5 3702363 3702671 309 bp

Anexo II: Dados completos do Ensembl versão 109 - Fevereiro de 2023

Enzima	Gene	Localização	ID do gene	Transcritos	ID do transcrito	Tipo de transcrito
Urato oxidase	UOX (pseudogene)	Cromossomo 1: 84, 363, 706-84, 397, 831 fita reversa	ENSG00000240520	UOX-201	ENST00000471089	pseudogene unitário transcrito transcrito processado transcrito processado transcrito processado transcrito processado
				UOX-202	ENST00000483236	
				UOX-203	ENST00000638373	
				UOX-204	ENST00000638856	
				UOX-205	ENST00000640983	
5-hidroxiisourato hidrolase	URAHP (pseudogene)	Cromossomo 16: 90, 043, 914-90, 047, 597 fita reversa	ENSG00000222019	URAHP-201	ENST00000610227	transcrito não processado lncRNA lncRNA lncRNA lncRNA
				URAHP-202	ENST00000409768	
				URAHP-203	ENST00000409873	
				URAHP-204	ENST00000517889	
				URAHP-205	ENST00000521551	
2-oxo-4-hidroxi-4-car boxi-5-ureidoimidazo lina descarboxilase	URAD	Cromossomo 13: 27, 977, 717-27, 988, 693 fita reversa	ENSG00000183463	URAD-201	ENST00000332715	codificante de proteína
Alantoinase	NA	NA	NA	NA	NA	NA
Alantoicase	ALLC	Cromossomo 2: 3, 658, 200-3, 702, 671 fita principal	ENSG00000151360	ALLC-201 ALLC-202 ALLC-203	ENST00000252505 ENST00000471711 ENST00000476389	codificante de proteína codificante de proteína, CDS não definida retenção de íntron

Anexo III: Dados completos do Ensembl versão 111 - Janeiro de 2024

Enzima	Gene	Localização	ID do gene	Transcritos	ID do transcrito	Tipo de transcrito
Urato oxidase	UOX (pseudogene)	Cromossomo 1: 84, 365, 428-84, 397, 831 fita reversa	ENSG00000240520	UOX-201	ENST00000471089	pseudogene unitário transcrito
5-hidroxiisourato hidrolase	URAHP (pseudogene)	Cromossomo 16: 90, 043, 914-90, 047, 597 fita reversa	ENSG00000222019	URAHP-201	ENST00000610227	transcrito não processado
5-hidroxiisourato hidrolase	URAHP (pseudogene)	Cromossomo 16: 90, 039, 761-90, 047, 773 fita reversa	ENSG00000290738	URAHP-202 URAHP-203 URAHP-204 URAHP-205	ENST00000409768 ENST00000409873 ENST00000517889 ENST00000521551	lncRNA lncRNA lncRNA lncRNA
2-oxo-4-hidroxi-4-carb oxi-5-ureidoimidazolin a descarboxilase	URAD	Cromossomo 13: 27, 977, 717-27, 988, 693 fita reversa	ENSG00000183463	URAD-201 URAD-202	ENST00000332715 ENST00000713555	codificante de proteína codificante de proteína, CDS não definida
Alantoinase	NA	NA	NA	NA	NA	NA
Alantoicase	ALLC	Cromossomo 2: 3, 658, 200-3, 702, 671 fita principal	ENSG00000151360	ALLC-201 ALLC-202 ALLC-203	ENST00000252505 ENST00000471711 ENST00000476389	codificante de proteína codificante de proteína, CDS não definida retenção de íntron

Anexo IV: Sequências das proteínas avaliadas

> A6NGE7|URAD_HUMAN|Homo sapiens (9606)

MDIEKVN~~SMDLGEFVDV~~FGNATERCPLIAAAVWSQRPFSDLEDLEKHF~~FAFID~~ALAQSGQEGILRCHPDLAGSEL
QRGTLTAESQREQSGAGLRS~~LGADERLRLAELNAQYRARFGFPFVLAARFSDRTAVPRELARLLCPSAQELRTA~~
LGEVKKIGSLRLADLLRADPAKL

> 2070_1|OHCU decarboxylase|Danio rerio (7955)

MDINVVNALAYEDFVKLFGNVVEKCP~~LISAAI~~WSYRPFKDLADIEARISEFIHSLPDSGKEGILRCHPDLAGRDL
QSGTLTPESQEEQSQAGMTTLD~~SAEIVHMYRLNSEYKERFGFPFVICARLNNKADIVRQLSERLKNRRTAELECA~~
IEEVKKICSLRLHSIVLSDIQTKL

>2Q37_1|OHCU decarboxylase|Arabidopsis thaliana (3702)

~~MGSSHHHHSSGENLYFOGH~~MAMEIGEDEWKVCCGSSEFAKOMSTSGPLTSQEAIYTARDIWFN~~QVNVTDWLEAF~~
SAHPQIGNTPSPSINSDFARRSVSEQSTAFATTSASALQELAEWNVLYKKKFGFIFIICASGRTHAEMLHALKER
YENRPIVELEIAAMEQMKITELRMAKLFSDK

>Q9LVM5|TTHL_ARATH|Arabidopsis thaliana (3702)

~~MAMEIGEDEWKVCCGSSEFAKOMSTSGPLTSQEAIYTARDIWFNOVNVTDWLEAFSAHPQIGNTPSPSINSDFAR~~
~~RSVSEQSTAFATTSASALQELAEWNVLYKKKFGFIFIICASGRTHAEMLHALKERYENRPIVELEIAAMEQMKIT~~
~~ELRMAKLFSDKAKVISETDSSSSPVSTKPDRLRLIIGGHLNVAAEAKAPKRSRPPIITTHVLDVSRGAPAAAGVEVH~~
~~LEVWSGTTGPSFVHGGGGVWSSVGTSA~~TD~~DRDGRSGPLMDLVDALNPGTYRISFDTAKYSPGCFFPYVSIVFOVTE~~
~~SOKWEHFHVPLLLAPFSFSTYRGS~~

>5Z5M_1|Predicted protein|Phaeodactylum tricorutum(556484)

~~MSAPVTVKDLLSKPSAEIASFLGGIYEHSAWVAEALVKDAESLASIETISQLAAAMKAI~~VNKSSKDQKLELLCAH~~~~
~~PDLCEKVGKLEMLTVESOEEO~~SRSGLOS~~LTDAELERFNSLNGAYRDOC~~GFPFILAVRNATKHTVLAALGGRVOHT~~~~
~~PEOEFMVALEOVHKIAWMRL~~LSKIDTSDAQ~~GFLTCHVLD~~TGNGCPA~~EKMRIHLHRLSPPEMAGLVGEFVTND~~DGR
~~LEGGPALKGGKEFTVGOYEWTF~~FCGEYFASKGTF~~TSGOPFLDTIPLRFGIDNPDDHYHVPLLVSPWSFSTYRG~~

>307J_1|OHCU decarboxylase|Klebsiella pneumoniae (272620)

~~MGSDKIHHHHHHSSGENLYFOGH~~MIALSQFNLSKDEAAGLLAPCVALPAWGETLVSLRPFASRHALLQTAREAM
ANWGEDELNAALSAHPRIGE~~KPTGSQ~~AHAALSRQE~~QSSV~~DSENERLAQALREGNARYEARFGRVFLIRAKGRSGE
EILQALTRRLQHTADEEVAEALAQLREITMLRLEGAIGE

>208I_1|Atu2327|Agrobacterium tumefaciens str. (176299)

MVTREEFVARFGGVFEHSPFIAERAYDAGGAGLELTAKAVHGALCAQFRVASEAERLGVLR~~AHPDLAGKLA~~IAGE
LTADSRNEQAGAGLDR~~LSPQEHARFTQLNSAYTEKFGFPFI~~I~~AVKGLNRHDILSAFDTRIDNNAQEFATATGQV~~
EKIAWLRLASMLPEG

>1059_1|Allantoicase|Saccharomyces cerevisiae (4932)

~~MGSDKIHHHHHH~~MKFFSLADEAEFKSIIISK~~NKAVD~~VIGSKLGGQVVSFSDEWFASAENLIQPTAPIRD~~PTRFVH~~
SGAWYDGWETRRHNEMEYD~~WVI~~IKMGVAAAHIGGEIDTAFFNGNHAPFVSIEALYDEGE~~EGNIVEDDSRWVEIV~~
EKFE~~CGPSQRHLFVRGNLTKERFTHIKL~~KMPDGGIARFRLYGRVVPPELKT~~KDHI~~IDLAYCNGAVALKYS~~DQ~~
HFGSV~~DNLLLPGRGH~~DMSDGWETKRSRQPGHTD~~WAVIQ~~LGRESSFIEKIVDTA~~HFRGNFPQFITVEGCL~~KESES
SENTGEGT~~WVELVGSKTGP~~DKEHVYEIRKSIRVSHV~~KLTII~~PDGGVKRIRVWGY

> Q8N6M5|ALLC_HUMAN|Homo sapiens (9606)

MDMASESVGGKILFATDDFFAPAENLIKSDSPCFKEHEYTEFGK~~WMDGWETRRKRI~~PGHDWC~~VLR~~LG~~IQ~~GVIRG~~F~~
DVDVSYFTGDYAPR~~VS~~IQ~~AA~~NLEEDK~~LPEI~~PERGTRTGA~~AA~~TPEEFEAIAELKSD~~DWSYL~~VPMT~~ELK~~PGNPASGH
NYFLVNSQQRWTHIRLNI~~F~~PDGGIARL~~RV~~FGTGQK~~DWTATDP~~KEPADLVAIAFGGVCVGF~~SNAKFGHPNNI~~IGVG
GAKSMADGWETARR~~LDRPPI~~LENDENGILLVPGCEWAV~~FRLAHPG~~VITRIEIDTKYFEGNAP~~DSCKVDGCIL~~T~~TQ~~
EEEAVIRQKWILPAHKWK~~PLLPVTKL~~SPNQSHLFD~~SLTLELQ~~DVITHARLTIVPDGGV~~SRLRLR~~GRFPSSIC~~LLRP~~
REK~~PMLKFSV~~SFKANP

Anexo V: Scripts para modelagem por homologia utilizando o Modeller

Sequência de interesse

```
>P1;A6NGE7
sequence:A6NGE7:::0.00: 0.00
MDIEKVN SMDLGEFVDVFGNATERCPLIAAAVWSQRPFS DLEDLEKHFFAFIDALAQSG
QEGILRCHPDLAGSELQRGTLTAESQREQSGAGLRSLGADERLRLAELNAQYRARFGFPF
VLAARFSDRTAVPRELARRLCPSAQELRTALGEVKKIGSLRLADLLRADPAKL*
```

Alinhamento

```
from modeller import *

env = environ()
aln = alignment(env)

mdl = model(env, file='2o70', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='2o70', atom_files='2o70.pdb')

aln.append(file='A6NGE7.ali', align_codes='A6NGE7')

aln.align2d()
aln.write(file='A6NGE7-2o70.ali', alignment_format='PIR')
aln.write(file='A6NGE7-2o70.pap', alignment_format='PAP')
```

Construção dos modelos

```
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()
env.io.hetatm = False

a = automodel(env,
              alnfile = 'A6NGE7-2o70.ali',
              knowns = ('2o70'),
              sequence = 'A6NGE7',
              assess_methods = (assess.DOPE, assess.GA341))

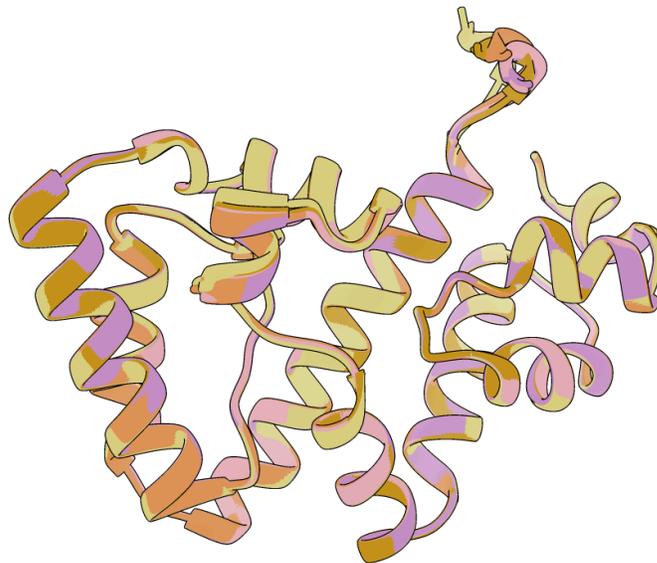
a.starting_model = 1
a.ending_model = 3000

a.make()
```

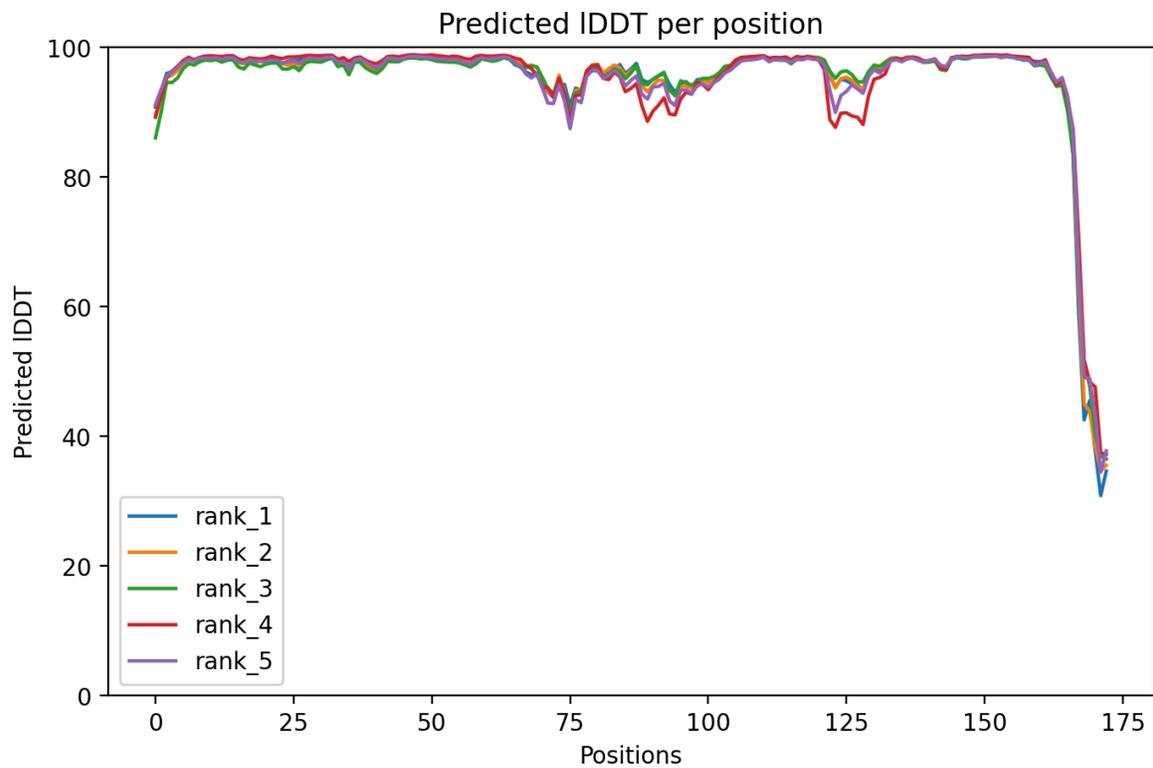
Anexo VI: Alinhamento das cinco estruturas geradas pelo ColabFold e seus valores de pLDDT

Modelos dos ColabFold rank 1 em caqui, 2 em laranja, 3 em roxo, 4 em rosa e 5 em ocre

(RMSD 0,239Å)



Valores de pLDDT para rank 001 (95,1), 002(95,1), 003(95), 004 (94,9) e 005 (94,9)

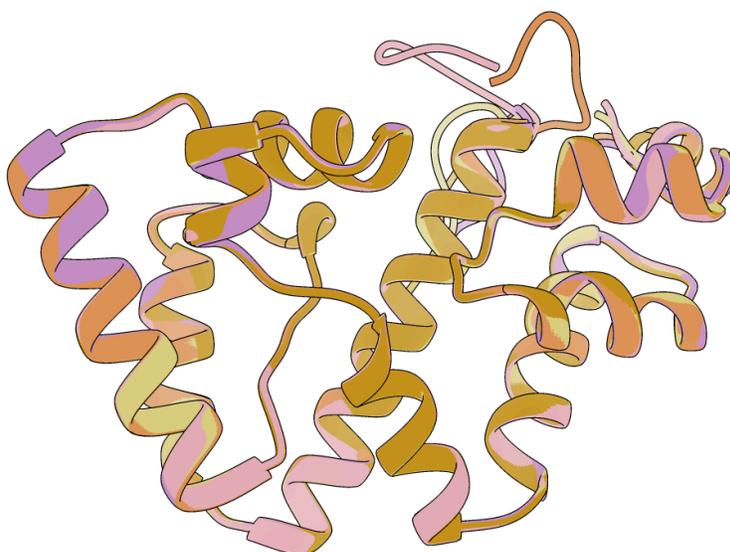


Anexo VII: Alinhamento das cinco estruturas de maior e menor energia geradas pelo Modeller e seus valores de DOPE

Modelos de maior energia do Modeller rank 1 (DOPE -18.427,631) em caqui, 2 (DOPE -18.502,602) em laranja, 3 (DOPE -18.554,492) em roxo, 4 (DOPE -18.583,896) em rosa e 005 (DOPE -18.585,973) em ocre (RMSD 0,211Å)

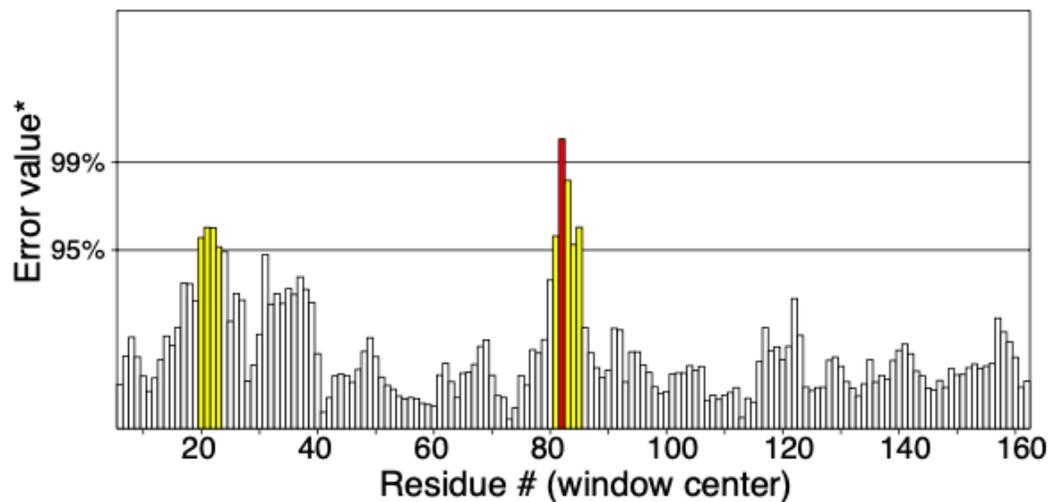


Modelos de menor energia do Modeller rank 1 (DOPE -19.572,543) em caqui, 2 (DOPE -19.562,246) em laranja, 3 (DOPE -19.547,309) em roxo, 4 (DOPE -19.544,213) em rosa e 005 (DOPE -19.532,156) em ocre (RMSD 0,163Å)



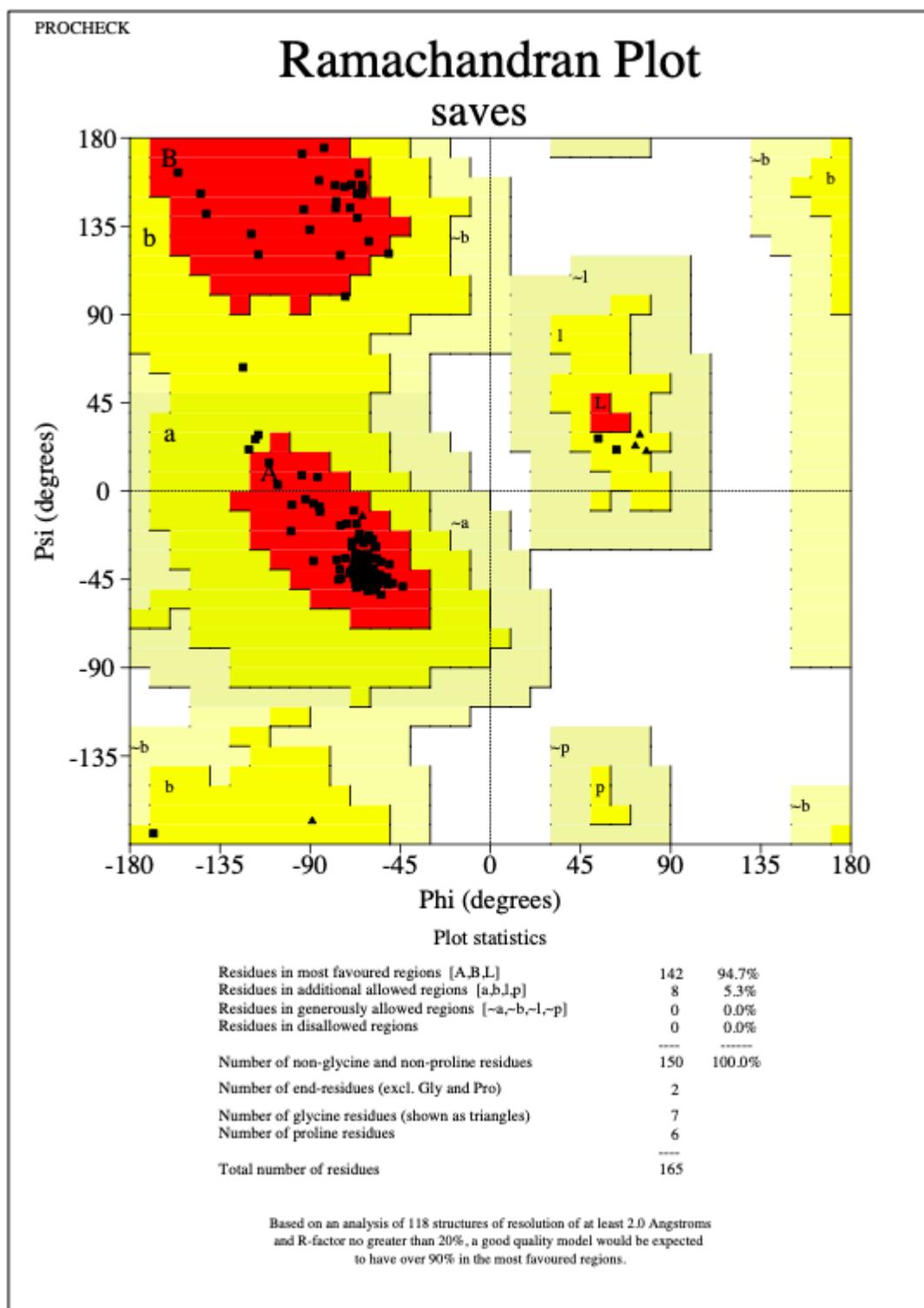
Anexo VIII: Validação da OHCU descarboxilase de *D. rerio* (PDB 2O70)

Program: ERRAT2
File: 2o70.pdb
Chain#:A
Overall quality factor**: 94.268



*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

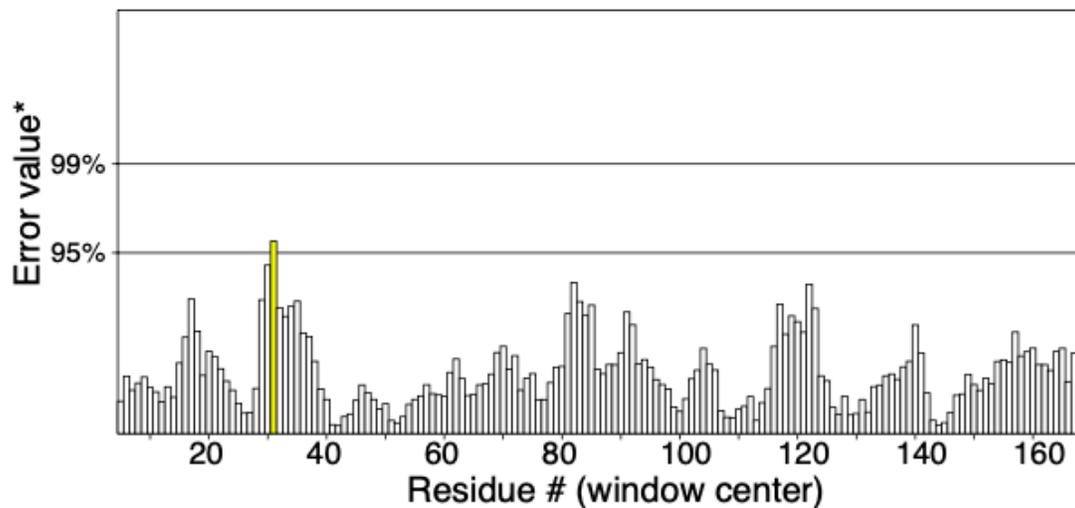
**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.



saves_01.ps

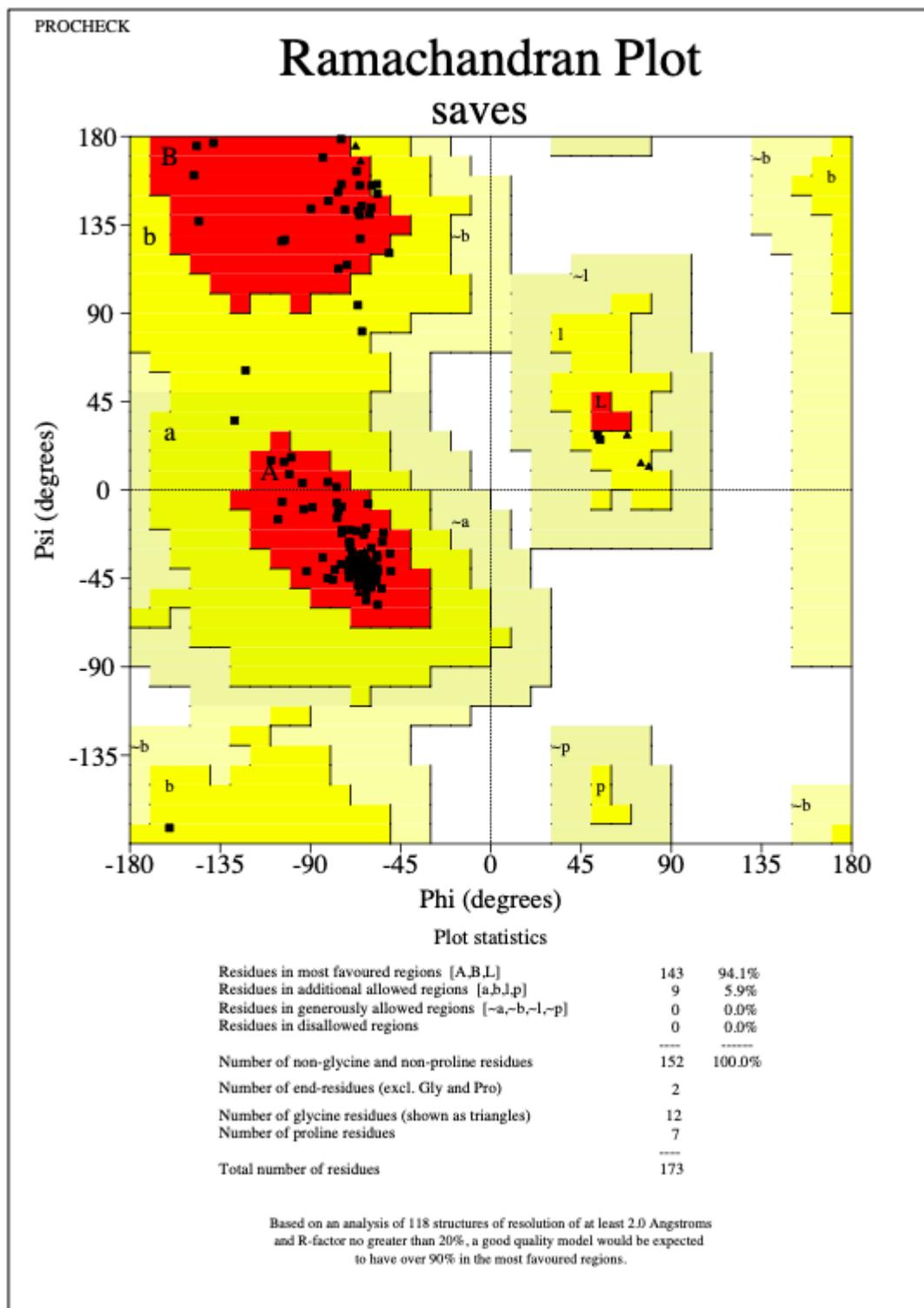
Anexo IX: Validação dos modelo construído utilizando o ColabFold

Program: ERRAT2
File: Colab.pdb
Chain#:A
Overall quality factor**: 99.390



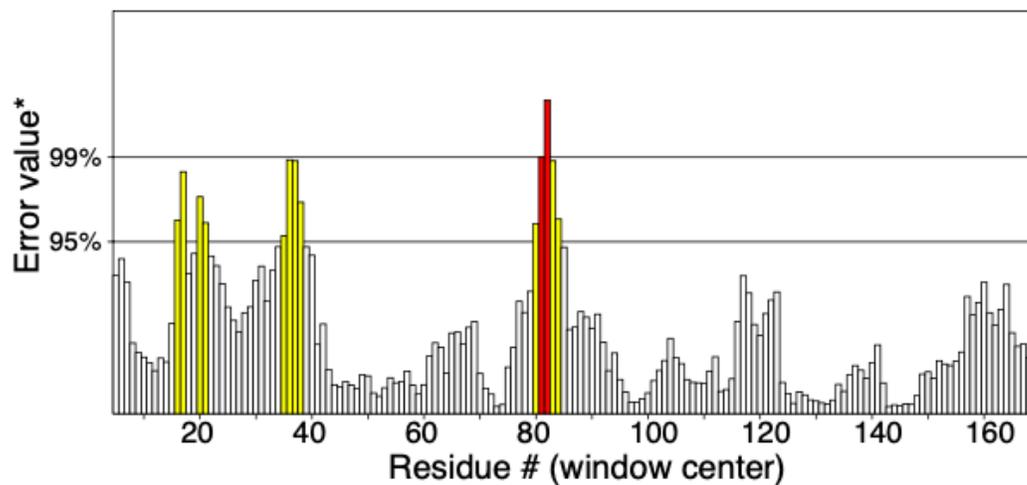
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.



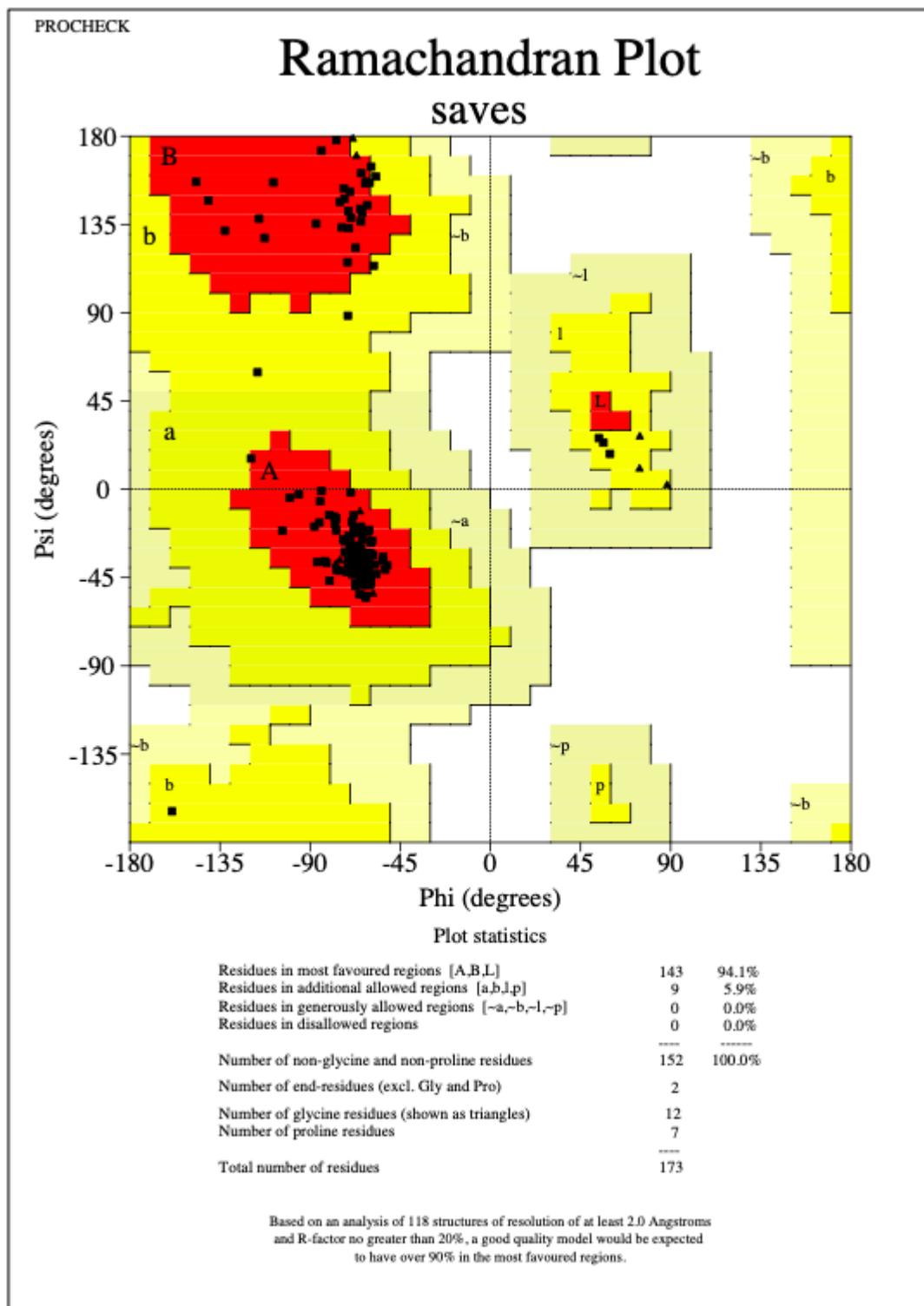
Anexo X: Validação do modelo construído utilizando o Modeller

Program: ERRAT2
File: Modeller.pdb
Chain#:A
Overall quality factor** : 92.073



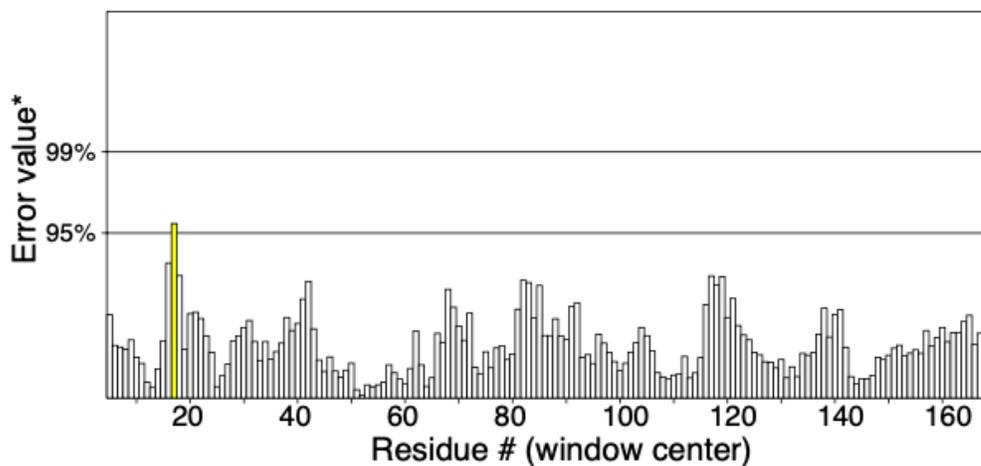
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.



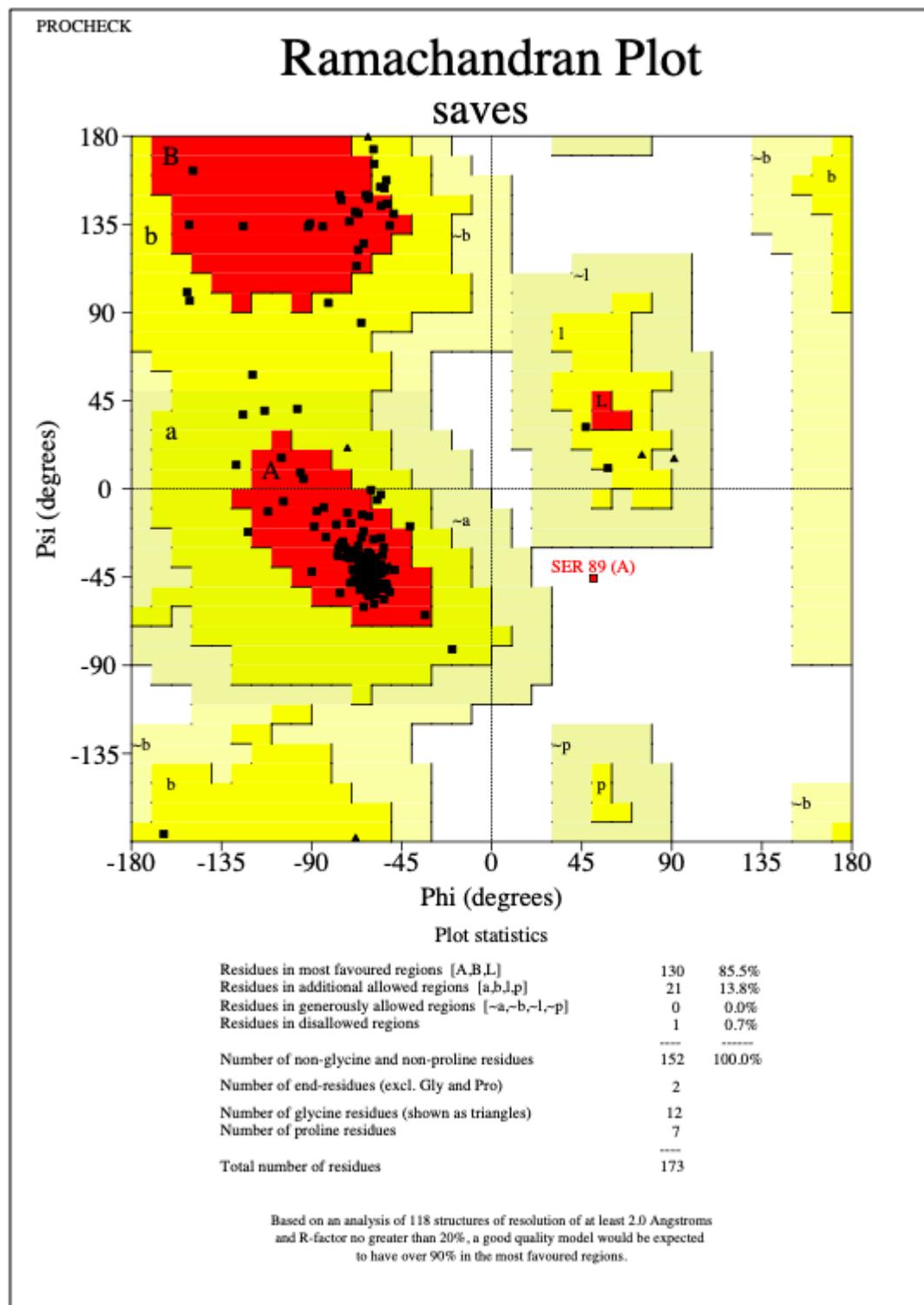
Anexo XI: Validação do modelo construído utilizando o iTASSER

Program: ERRAT2
File: iTasser.pdb
Chain#:A
Overall quality factor**: 99.394



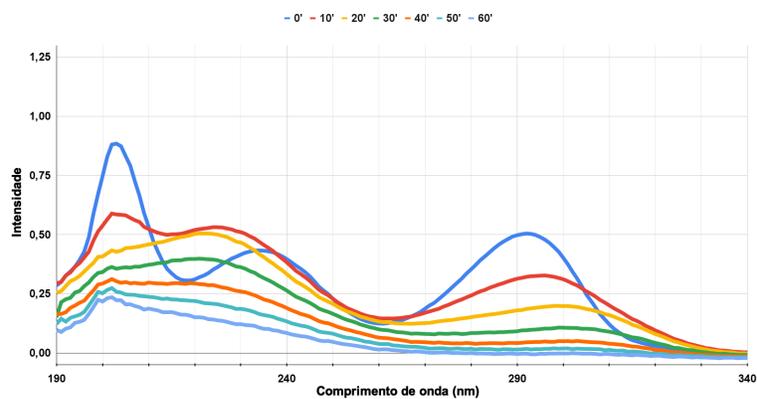
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.

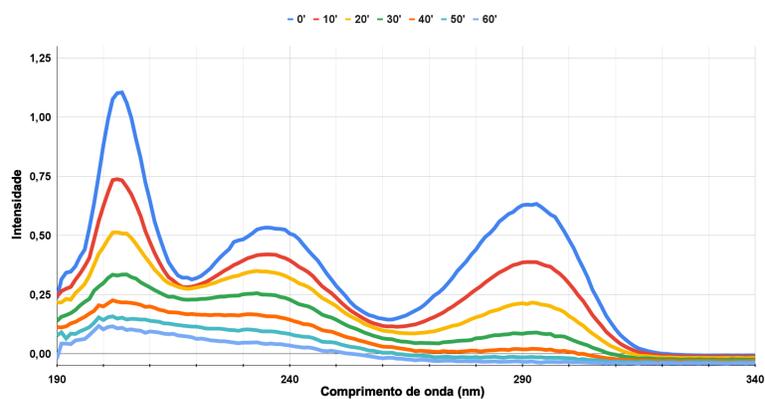


Anexo XII: Espectros dos ensaios enzimáticos individuais

30 μM de ácido úrico + 50 nM de uricase recombinante de *Candida sp.* e a HIUase recombinante de *Herbaspirillum seropedicae*.



30 μM de ácido úrico + 50 nM uricase recombinante de *Candida sp.* + 50 nM de HIUase recombinante de *Herbaspirillum seropedicae*.



30 μM de ácido úrico + 50 nM uricase recombinante de *Candida sp.* + 50 nM de HIUase recombinante de *Herbaspirillum seropedicae* + 50 nM OHCU descarboxilase de *H. sapiens*

