

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA, ECOLOGIA E EVOLUÇÃO

Alexandre Eustáquio Murta Lemos

CARACTERIZAÇÃO *IN SILICO* DA DARPP-32 EM SERES HUMANOS

Belo Horizonte

2023

Alexandre Eustáquio Murta Lemos

CARACTERIZAÇÃO *IN SILICO* DA DARPP-32 EM SERES HUMANOS

Dissertação apresentada ao Programa de Pós-Graduação em Genética, Ecologia e Evolução da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Mestre em Genética.

Orientadora: Laila Alves Nahum

Belo Horizonte

2023

043

Lemos, Alexandre Eustáquio Murta.

Caracterização in silico da DARPP-32 em seres humanos [manuscrito] / Alexandre Eustáquio Murta Lemos. – 2023.

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

Orientadora: Laila Alves Nahum.

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

1. Genética. 2. Fosfoproteína 32 Regulada por cAMP e Dopamina . 3. Homologia de Genes. 4. Bioquímica. 5. Domínios Proteicos. 6. Proteínas Intrinsicamente Desordenadas. 7. Alinhamento de Sequência. I. Nahum, Laila Alves. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 575



UNIVERSIDADE FEDERAL DE MINAS GERAIS
 Instituto de Ciências Biológicas
 Programa de Pós-Graduação em Genética

ATA DE DEFESA DE DISSERTAÇÃO

ATA DA DEFESA DE DISSERTAÇÃO	342/2023 ENTRADA
ALEXANDRE EUSTÁQUIO MURTA LEMOS	2º/2020 CPF: 066.335.056-55

Às treze horas e trinta minutos do dia **29 de junho de 2023**, reuniu-se a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Caracterização in silico da DARPP-32 em seres humanos**", requisito para obtenção do grau de Mestre em **Genética**. Abrindo a sessão, a Presidente da Comissão, **Laila Alves Nahum**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Laila Alves Nahum	Fiocruz Minas	666.211.786-20	APROVADO
Luciana Maria Silva	FUNED	977.664.456-20	APROVADO
Lucas Bleicher	UFMG	813.528.593-00	APROVADO

Pelas indicações, o candidato foi considerado: **APROVADO**

O resultado final foi comunicado publicamente ao candidato 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.

Belo Horizonte, 29 de junho de 2023.

Laila Alves Nahum

Luciana Maria Silva

Lucas Bleicher



Documento assinado eletronicamente por **Lucas Bleicher, Professor do Magistério Superior**, em 30/06/2023, às 16:37, 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 **Luciana Maria Silva Lopes, Usuário Externo**, em 03/07/2023, às 14:28, 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 **Laila Alves Nahum, Usuário Externo**, em 04/07/2023, às 13:59, 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 **2423064** e o código CRC **B1ED3106**.

AGRADECIMENTOS

A DEUS, pela dádiva da vida, por ser meu porto seguro quando tudo parece desabar, por abrir caminhos, por fortalecer a minha mente, corpo e espírito para que eu possa cumprir minha jornada;

À minha mãe, exemplo maior de vida, caráter, dedicação, empatia e superação. Não há como exprimir a importância dela em minha vida;

Ao meu pai, José Eustáquio de Abreu Lemos. Ao Senhor, que aí de cima, motiva-me a prosseguir com fé, entusiasmo e responsabilidade;

Às minhas irmãs, Luciana, Letícia e Patrícia e ao meu irmão, Carlos Henrique, vocês são guerreiros vitoriosos, amigos maiores, que nossos laços nunca se desfaçam;

Ao meu amado sobrinho, João Pedro, benção e anjo em minha vida, você proporciona paz, alegria e força;

Aos meus avós, moradores eternos em meu coração, os seus causos ficarão, eternamente, na memória;

Aos meus tios, primos e primas, inúmeros em quantidade, mas insuperáveis em qualidade;

Aos meus cunhados, Nilo e Ademar, obrigado pela amizade, que venham muitos churrascos, passeios e batuques;

À Elvira Cynthia Alves Horácio, sua contribuição é imensurável, inestimável, grato pela parceria, ensinamentos e inspiração, este trabalho encontra-se em dívida contigo;

À minha orientadora, Dra. Laila Alves Nahum, obrigado pela atenção, pela acolhida, pelo aprendizado, pela paciência e a oportunidade ímpar em minha vida;

Ao Programa de Pós-Graduação em Genética do ICB-UFMG pela acolhida e oportunidade em realizar esta realização em minha vida, gratidão especial ao corpo docente, a Raíssa e a Izabela, aos coordenadores, ao CAE, todos os discentes e as demais pessoas, que contribuem para o enriquecimento deste curso, que o sucesso e a evolução profissional sejam uma constante para todos vocês;

A UFMG e ao Instituto René Rachou - FIOCUZ Minas, a grandeza do trabalho de vocês é transcendental, marcará para sempre minha vida e, também, de incontáveis pessoas de diferentes lugares, obrigado pela acolhida e oportunidade;

Ao Dr. Robson Kazunori, muito obrigado pelo cuidado, pelas dicas e a confiança depositada em mim;

A equipe EvoBioTeam, uma parceria e conexão fortes;

Aos meus amigos Victor, Thiago, Jorge, Edson e Gabriel, vocês estão longe, mas o endereço em meu coração é permanente;

À minha velha turma da pracinha na Rua Brasil no bairro Jardim América, muitas bênçãos, alegria e sucesso para todos;

Ao meu amigo, Arão, obrigado pela amizade e as inúmeras sessões terapêuticas;

À Cida e ao Jorge do Mult Informática, vocês foram suporte em momentos críticos para a realização deste trabalho, gratidão eterna;

Aos amigos, Marcelo, Gustavo, Rodrigo e Túlio, que venham muitas cachoeiras e sushis e sashimis;

Aos meus amigos Daniel e João Paulo, parceiros leais de inúmeras risadas e mútua colaboração sempre, “a gente se aguenta, né?!”;

Ao meu eterno professor, Tarcísio, os ensinamentos adquiridos e a inspiração ultrapassam fronteiras;

Ao meu cachorro, Duke;

A todos aqueles não citados, mas que marcaram para sempre minha existência;

E, também, aos livros, aos gibis, que li, as músicas, que ouvi, as praias, as serras, as cachoeiras onde fui e as milhares de maravilhas, que mostram a grandeza e a vastidão deste mundo a se desbravar.

“Vai, segue o seu caminho”
(Ataulfo Alves)

RESUMO

Introdução: PPP1R1A, PPP1R1B e PPP1R1C são membros de uma família proteica conhecida pelo efeito inibitório da proteína fosfatase 1 (PP1). Dentre elas, a PPP1R1B (DARPP-32) é a mais estudada, pois as investigações relacionam estas proteínas, especialmente, a DARPP-32, a várias doenças humanas como o câncer, a esquizofrenia e a dependência química. Trabalhos *in silico* a respeito dessas proteínas são escassos e a estrutura tridimensional ainda é desconhecida. **Justificativa:** Compreender as propriedades e a função de um grupo particular de proteínas, que podem proporcionar o entendimento sobre a bioquímica, a fisiologia e, também, os efeitos biológicos que podem auxiliar em trabalhos futuros acerca de alternativas terapêuticas contra as diversas enfermidades em que estas proteínas estão envolvidas. **Objetivos:** Avaliar a biodiversidade molecular da DARPP-32 e os potenciais homólogos que ela possui. a) Identificar os potenciais homólogos da DARPP-32 em humanos; b) Caracterizar as propriedades bioquímicas; c) Avaliar a arquitetura proteica dos alvos moleculares selecionados; d) Construir alinhamentos de múltiplas sequências; e) Identificar padrões de conservação e divergência nessas sequências. **Metodologia:** a) Levantamento bibliográfico; b) Buscas em bancos de dados e ferramentas de bioinformática; c) Busca por sequências e identificação de potenciais homólogos; d) Avaliação da arquitetura proteica; e) Classificação das sequências proteicas; f) Caracterização bioquímica das proteínas de interesse; g) Construção dos alinhamentos de múltiplas sequências. **Resultados e Discussão:** Foram identificados potenciais homólogos de DARPP-32. As PPP1R1A, PPP1R1B e PPP1R1C compartilham um domínio (Pfam: PF05395), que determinam as propriedades bioquímicas delas, como a desordem estrutural e as interações moleculares. Apenas a PPP1R1B possui uma região rica em glutamato. As propriedades físico-químicas foram analisadas, identificando as sub-regiões de desordem nessas proteínas. Sendo que, os resultados mostram que a DARPP-32 e outros membros dessa família de proteínas são Proteínas Intrinsecamente Desordenadas (PIDs) possuindo resíduos potencialmente fosforilados, especialmente, serina e treonina, e outras modificações pós-traducionais. **Conclusão e Perspectivas:** Este trabalho ajudará na compreensão da estrutura e função dessas proteínas, (especialmente a DARPP-32), auxiliando pesquisas a respeito da evolução molecular e terapias contra enfermidades em que estão envolvidas.

Palavras-chave: darpp-32; homologia; caracterização bioquímica; domínio proteico; proteínas intrinsecamente desordenadas; alinhamento de sequência.

ABSTRACT

Introduction: PPPR1A, PPPR1B and PPPR1C are members of a protein family known for its inhibitory effect of protein phosphatase 1 (PP1). Among them, PPPR1B (DARPP-32) is the most studied. Some investigations relate these proteins, especially DARPP-32, to several human diseases such as cancer, schizophrenia and chemical dependence. *In silico* studies about these proteins are scarce and their three-dimensional structure is still unknown. **Justification:** Understand the properties and function of a particular group of proteins, providing an understanding of their biochemistry, physiology and biological effects that may assist future studies on therapeutic alternatives against various diseases in which these proteins are involved. **Goals:** Evaluate the molecular biodiversity of DARPP-32 and its potential counterparts. a) Identify potential homologues of DARPP-32 in humans; b) Characterize their biochemical properties; c) Evaluate the protein architecture of selected molecular targets; d) Construct multiple sequence alignments; e) Identify patterns of conservation and divergence in these sequences. **Methodology:** a) Bibliographic survey; b) Searches in databases and bioinformatics tools; c) Search for sequences and identification of potential homologues; d) Evaluation of protein architecture; e) Classification of protein sequences; f) Biochemical characterization of proteins of interest; g) Construction of multiple sequence alignments. **Results and Discussion:** It has been identified some Potential homologues of DARPP-32. PPP1R1A, PPP1R1B and PPP1R1C share a domain (Pfam: PF05395) that determines their biochemical properties such as structural disorder and molecular interactions. Only PPP1R1B has a glutamate-rich region. The physicochemical properties were analyzed by identifying disorder sub-regions in these proteins. The results show that DARPP-32 and other members of this protein family are Intrinsically Disordered Proteins (PIDs) possessing potentially phosphorylated residues, especially in serine and threonine, and other post-translational modifications. **Conclusion and Perspectives:** This work will help understanding the structure and function of these proteins (especially DARPP-32), assisting research on the molecular evolution and therapies against diseases in which they are involved.

Keywords: darpp-32; homology; biochemical; characterization; protein domain; intrinsically disordered proteins; sequence alignment.

LISTA DE FIGURAS

Figura 1 -	Localização cromossômica do gene de PPP1R1A	16
Figura 2 -	Regulação de PP1 e PKA pela fosforilação de sítios de DARPP-32	20
Figura 3 -	Organização da DARPP-32 e sua isoforma truncada t-DARPP	21
Figura 4 -	Fluxograma das etapas da Metodologia	28
Figura 5 -	Arquitetura proteica de PPP1R1A, PPP1R1B e PPP1R1C	37
Figura 6 -	Representação esquemática de PPP1R1A, PPP1R1B e PPP1R1C	38
Figura 7 -	Estrutura secundária de PPP1R1A, PPP1R1B e PPP1R1C	40
Figura 8 -	Propriedades físico-químicas de PPP1R1A, PPP1R1B e PPP1R1C	42
Figura 9 -	Perfil rico em glutamato de PPP1R1B	44
Figura 10 -	Possíveis resíduos fosforilados e com outras MPTs em PPP1R1A	45
Figura 11 -	Possíveis resíduos fosforilados e com outras MPTs em PPP1R1B	46
Figura 12 -	Possíveis resíduos modificados por fosforilação em PPP1R1C	46
Figura 13 -	Alinhamento das sequências de PPP1R1A, PPP1R1B e PPP1R1C	47
Figura 14 -	Alinhamento de sequências destacando os resíduos de serina e treonina	48

LISTA DE TABELAS

Tabela 1 - Propriedades físico-químicas de PPP1R1A, PPP1R1B e PPP1R1C	40
Tabela 2 - Características bioquímicas de PPP1R1A, PPP1R1B e PPP1R1C	42
Tabela 3 - Matriz de identidade entre PPP1R1A, PPP1R1B e PPP1R1C	48

LISTA DE QUADROS

Quadro 1 - Proteínas com o domínio PF05395 em <i>Homo sapiens</i>	35
Quadro 2 - Sequências de PPP1R1A, PPP1R1B e PPP1R1C	36
Quadro 3 - Arquitetura de PPP1R1A, PPP1R1B e PPP1R1C de <i>Homo sapiens</i>	37
Quadro 4 - Padrão de MPTs de PPP1R1A	43
Quadro 5 - Padrão de MPTs de PPP1R1B	44
Quadro 6 - Padrão de MPTs de PPP1R1C	45

LISTA DE SIGLAS E ABREVIATURAS

AKT	Proteína quinase b
AMPC	Adenosina 3',5' - monofosfato cíclico
BCL2	Regulador de apoptose Bcl2
BLAST	Basic Local Alignment Search Tool
BNDF	Fator neurotrófico derivado do cérebro
DA	Dopamina
DARPP-32	Fosfoproteína regulada por AMPc e dopamina de 32 kDA
EMBL-EBI	European Bioinformatics Institute
FASTA	Formato de sequências moleculares
GLU	Glutamato
L-DOPA	L-3,4-dihidroxifenilalanina
IGF1R	Receptor do fator 1 de crescimento tipo insulina
IPP-1	Família de inibidores de proteína fosfatase 1
MPT	Modificação pós-traducional
NCBI-NLM-NIH	National Center for Biotechnology Information-National Library of Medicine- National Institutes of Health
PANTHER	Protein Analysis Through Evolutionary Relationships
PFAM	Protein Family database
PI	Ponto isoelétrico
PID	Proteínas intrinsecamente desordenadas
PKA	Proteína quinase A
PP1	Proteína fosfatase 1
PPP1R1A	Subunidade reguladora 1 A da proteína fosfatase 1
PPP1R1B	Subunidade reguladora 1 B da proteína fosfatase 1
PPP1R1C	Subunidade reguladora 1 C da proteína fosfatase 1
PSIPRED	PSI-BLASTsecondary structure PRediction
SER	Serina
SIB	Swiss Institute of Bioinformatics
T- DARPP	Isoforma truncada de DARPP-32
THR	Treonina
UNIPROT	Universal Protein Knowledgebase

SUMÁRIO

1 INTRODUÇÃO	14
1.1 A família proteica de inibidores de PP1	14
1.2 Proteína fosfatase 1	15
1.3 PPP1R1A e PPP1R1C: parálogos de DARPP-32	16
1.3.1 PPP1R1A	16
1.3.2 PPP1R1C	17
1.4 DARPP-32	17
1.4.1 História	18
1.4.2 Estrutura, bioquímica e expressão gênica	20
1.4.3 Doenças relacionadas	21
1.5 Proteínas intrinsecamente desordenadas	24
2 JUSTIFICATIVA	26
3 OBJETIVOS	27
3.1 Objetivo geral	27
3.2 Objetivos específicos	27
4 METODOLOGIA	28
4.1 Levantamento bibliográfico	31
4.2 Buscas em bancos de dados e ferramentas	31
4.2.1 Busca por sequências e identificação de potenciais homólogos	31
4.2.2 Avaliação da arquitetura proteica	32
4.2.3 Classificação da sequência proteica	32
4.3 Caracterização bioquímica das proteínas de interesse	33
4.4 Construção dos alinhamentos de múltiplas sequências	34
5 RESULTADOS	35
5.1 Levantamento bibliográfico	35
5.2 Buscas em bancos de dados e ferramentas	35
5.2.1 Sequências e identificação de potenciais homólogos	35
5.2.2 Avaliação da arquitetura proteica	36
5.2.3 Classificação das sequências proteicas	37
5.3 Caracterização bioquímica das proteínas de interesse	39
5.3.1 ProtParam	39
5.3.2 PSIPRED	40

5.3.3 PROSITE.....	SUMÁRIO	43
5.3.4 PhosphoSitePlus		45
5.3.5 Signal P-6.0		47
5.4 Alinhamento das sequências.....		47
6 DISCUSSÃO		49
7 CONCLUSÕES E PERSPECTIVAS.....		54
REFERÊNCIAS.....		55

1 INTRODUÇÃO

1.1 A família proteica de inibidores de PP1

Uma família proteica é um grupo de proteínas que compartilham uma ancestralidade comum e têm certo grau de semelhança em sua estrutura e/ou função bioquímica e biológica. As proteínas que compartilham uma origem comum são definidas como proteínas homólogas e podem se dividir em parálogas e ortólogas. As proteínas parálogas são aquelas originárias de um evento de duplicação gênica seguido por divergência, enquanto as proteínas ortólogas resultam de um processo de diversificação taxonômica (e.g. especiação). As parálogas e ortólogas podem ter a mesma função ou assumir novas (e.g. subfuncionalização) e até mesmo serem inativadas (Nahum, 2012).

Existem três proteínas, que são nomeadas como subunidades reguladoras 1A da proteína fosfatase 1 (PPP1R1A), subunidade reguladora 1b da proteína fosfatase 1 (PPP1R1B) e subunidade reguladora 1C da proteína fosfatase 1 (PPP1R1C). São conhecidas cientificamente como a família de proteínas inibidoras de proteína fosfatase 1 (IPP-1), cuja propriedade mais notória é, como o próprio nome indica, a inibição endógena de atividade da enzima proteína fosfatase 1 (PP1) (MCLAREN, *et al.* 2000).

Em razão dessa propriedade, essas proteínas têm sido associadas, ao longo de décadas de pesquisa, a papéis vitais na transdução de sinal, em especial, na sinalização dos neurotransmissores, dopamina (DA) e glutamato (Glu). Diversos estudos mostram que a fosforilação e a desfosforilação das PPP1R1A, PPP1R1B e PPP1R1C pela proteína quinase A (PKA), principalmente, nos resíduos de serina e treonina, influenciam vias bioquímicas complexas que afetam processos biológicos como angiogênese, resistência a medicamentos e migração celular (AVANES, *et al.*, 2019).

Estudos filogenéticos mostram que a família de inibidores de PP1, provavelmente, teve origem com os ancestrais de Vertebrata, sendo que, o mais provável é que essa família de proteínas tenha surgido com os Gnathostomata (YGER & GIRAULT, 2011).

Atualmente, acredita-se que PPP1R1A, PPP1R1B e PPP1R1C estejam presentes na maioria das espécies de vertebrados terrestres, como anfíbios, répteis, aves e mamíferos, bem como, na grande maioria dos peixes ósseos e cartilagosos. Logo, acredita-se que a presença dessas proteínas possa estar relacionada a maior diversidade das Vertebratas quando comparadas com os Cyclostomata (YGER & GIRAULT, 2011). Embora essas três sejam encontradas em vertebrados, somente a PPP1R1A e a PPP1R1B são amplamente expressas nas células neuronais (GIRAULT & NAIRN, 2021).

Diversas pesquisas têm relacionado estas proteínas, especialmente a PPP1R1B, que também é conhecida como DARPP-32, a inúmeras enfermidades que acometem a sociedade humana. A DARPP-32 tem uma relevante participação em processos biológicos e nas vias bioquímicas complexas, que quando em desequilíbrio podem acarretar uma série de doenças, como vários tipos de cânceres, transtornos neuronais que afetam o sistema motor como o mal de Parkinson e doenças psiquiátricas como esquizofrenia e a dependência química (GIRAULT & NAIRN, 2021).

1.2 Proteína fosfatase 1

A proteína fosfatase 1 (PP1) é uma enzima notória pela ampla participação em reações de desfosforilação nas células dos organismos eucarióticos e, também, são conhecidas mais de duzentas subunidades de proteínas regulatórias de PP1. Sendo assim, o mecanismo mais comum de regulação da atividade de PP1 é a ligação a proteínas fosforiladas, encontrada, em sua maioria, nos resíduos de serina (SER) e treonina (THR). À vista disso, esse tipo de reação bioquímica é conhecida como modificação pós-traducional (MPT). Observa-se que, a fosforilação é um tipo de MPT com ampla ocorrência em organismos eucariotos.

Nota-se que, a enzima PP1, por sua vez, é uma proteína multifuncional, ou seja, destaca-se por catalisar a maior parte das reações de desfosforilação nas células dos organismos, acarretando assim, uma grande influência nos processos biológicos, como a transcrição gênica, a apoptose e a divisão celular (REBELO, *et al.*, 2015).

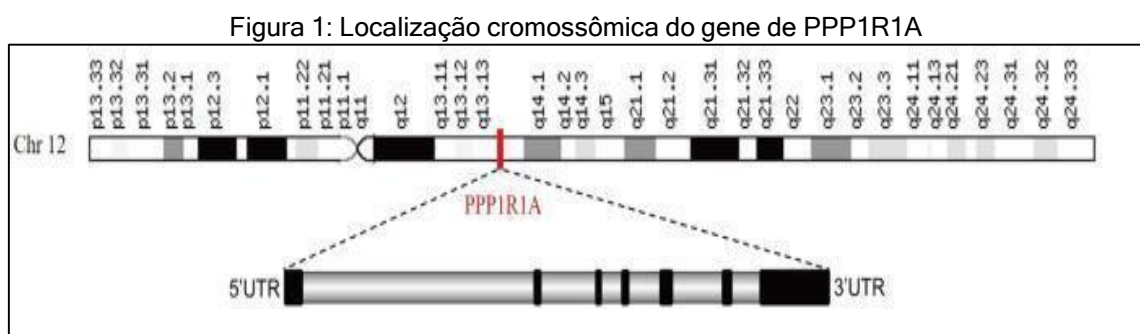
Logo, a compreensão sobre PP1 e as moléculas de interação com a mesma é de grande relevância na biologia e na saúde.

1.3 PPP1R1A e PPP1R1C: parálogos de DARPP-32

Dentre as proteínas integrantes da família IPP-1, a PPP1R1A e a PPP1R1C são as de menor destaque na pesquisa científica. Contudo, elas, também, têm importância na biologia e na saúde. Evolutivamente, estas proteínas e a DARPP-32 são parálogas (GIRAULT, *et al.*, 2021).

1.3.1 PPP1R1A

A proteína PPP1R1A é codificada pelo gene PPP1R1A, cujo *locus* localiza-se no cromossomo 12q13.2, fazem parte da estrutura desse gene sete éxons e seis íntrons (WU, *et al.*, 2022). Esse gene pode formar quatro transcritos diferentes devido as variações por *splicing* que leva a produção de três isoformas. Sabe-se que, a isoforma X1 tem 132 aminoácidos, a isoforma X2 tem 119 aminoácidos e a isoforma mais comum, possui a função de expressar a proteína com um total de 171 aminoácidos na sequência primária, possuindo 516 nucleotídeos na região que contém a sequência codificante. Essa última é a proteína PPP1R1A que predomina nas pesquisas e na literatura científica, sendo que, a localização cromossômica de PPP1R1A está identificada a seguir.



Observa-se que, a PPP1R1A é, preferencialmente, expressa nos tecidos adiposos, nas células renais, no coração, no esôfago e no cérebro. Estudos recentes apontam que a superexpressão de PPP1R1A tem uma participação importante na patogênese

do Sarcoma de Ewing (LUO, *et al.*, 2020), sendo que, essa proteína pode servir como um marcador biológico para o carcinoma hepatocelular (WU ET, *et al.*, 2022) e, também, para a detecção de injúria das células β do pâncreas (JIANG *et al.*, 2013).

1.3.2 PPP1R1C

Sabe-se que, a PPP1R1C é uma proteína integrante da família dos inibidores de PP1 com menor destaque na pesquisa científica, o *locus* dessa proteína localiza-se no cromossomo 2q31.3-q32.1, possui um total de nove éxons, por razão de variações no fenômeno de *splicing*. Existem treze diferentes transcritos para esta proteína, podendo formar duas isoformas. Já a isoforma 1 possui 116 aminoácidos, sendo que, a isoforma mais comum conta com um total de 109 aminoácidos e tem sequência codificante possuindo 330 nucleotídeos.

A proteína isoforma 1 expressa-se, predominantemente, nos tecidos que formam partes do coração, dos testículos, do cérebro, da próstata, da vesícula biliar, do pulmão, do estômago, dos rins e das glândulas salivares. Nota-se que, a PPP1R1C foi provada como um marcador diferencial no prognóstico de câncer de tireoide (TANG, *et al.*, 2018) e, também, para a detecção do glioblastoma (LIU, *et al.*, 2017).

1.4 DARPP-32

Embora vários trabalhos científicos demonstrem a importância da família de inibidores de PP1 na biologia e na saúde, dentre as três integrantes da família, a DARPP-32 tem sido historicamente a mais estudada. As propriedades dela e a relevância bioquímica e fisiológica que possui são mais bem demonstradas por inúmeros estudos desde o descobrimento dela por Paul Greengard e seus colaboradores no início dos anos 80 (HEMMINGS, *et al.*, 1984) e (WALAAS & GREENGARD, 1984). Em razão da abundância de informações e da importância dela na biologia e na saúde, também é chamada de PPP1R1B, sendo o objeto principal dos estudos e informações realizadas e fornecidas neste trabalho.

1.4.1 História

A descoberta de DARPP-32 iniciou-se entre a virada dos anos 70 e começo dos anos 80 com as pesquisas do grupo liderado por Paul Greengard a respeito da neurotransmissão e a participação de segundos mensageiros e das proteíno quinases no mecanismo fisiológico de transdução de sinal (GIRAULT, *et al.*, 2021), (WALAAS & GREENGARD, 1984) e (HEMMINGS, *et al.*, 1984).

O trabalho desse grupo propiciou sustentação ao entendimento sobre a plasticidade sináptica e neuromodulação, sendo que, à medida que as pesquisas progrediram, um grupo de moléculas, inicialmente nomeadas como “terceiros mensageiros”, em virtude do descobrimento da relação dessas com as proteíno quinases, que estão envolvidas na neurotransmissão e estavam sendo identificadas, esse trabalho foi ganhando a atenção dos pesquisadores.

De acordo com os resultados desse trabalho, várias destas proteínas tinham a distribuição restrita a regiões cerebrais específicas, em que a fosforilação sofria grande influência dos segundos mensageiros, como, por exemplo, o AMPc (GIRAULT & NAIRN, 2021).

Com o avanço das pesquisas sobre neurotransmissão dopaminérgica, uma destas proteínas, tidas como “terceiros mensageiros”, ganhou a atenção dos pesquisadores, pois verificou-se que, ela era altamente fosforilada quando estimulada por AMPc, tendo ocorrência abundante no *striatum*, uma região cerebral com grande inervação do neurotransmissor dopamina.

Em razão da fosforilação desta proteína ser intensificada devido a dopamina e AMPc, Walaas e seus colaboradores a nomearam como DARPP-32 ou fosfoproteína regulada por AMPc e como dopamina de massa molecular de 32 kDa (WALAAS & GREENGARD, 1984).

Uma série de pesquisas realizadas por diversos grupos identificaram a presença de DARPP-32 em regiões como os neurônios espinhosos médios do *striatum*, que possui uma grande inervação dopaminérgica e, também, em outros tipos de células com

presença de dopamina, como o córtex cerebral, as células de Purkinje, assim como, as células com presença de outros neurotransmissores, como a serotonina, no caso das células do plexo coroide. A DARPP-32 também teve a presença confirmada nas células de outros órgãos como, por exemplo, nos rins.

No ano de 1984, Hemmings e seus colaboradores foram responsáveis pela descoberta do que, talvez seja, a propriedade bioquímica de maior influência em fisiologia da DARPP-32: a atividade inibitória de PP1. Devido ao efeito inibitório da ação da proteína fosfatase 1, DARPP-32 também foi chamada de PPP1R1B ou subunidade 1 B reguladora da proteína fosfatase 1 (HEMMINGS, *et al.*, 1984).

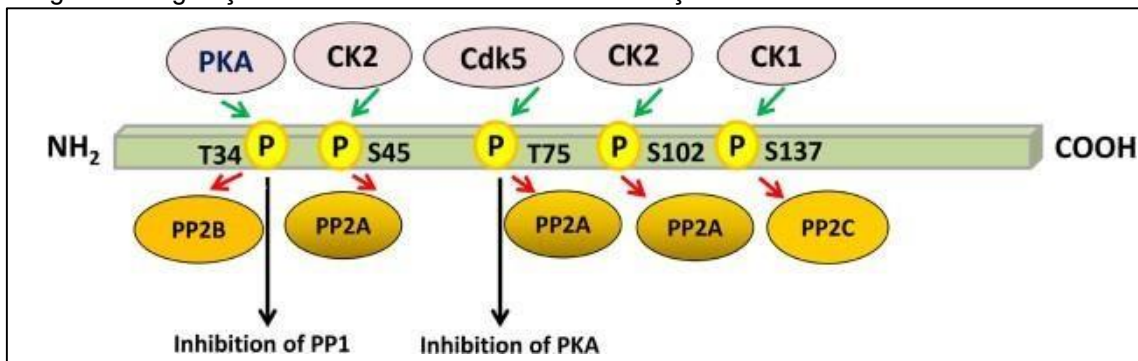
Esse efeito inibitório de PP1 por DARPP-32 deve-se, principalmente, a fosforilação do resíduo de Thr34 da estrutura primária, essa propriedade coloca a PPP1R1B como uma proteína chave na mediação das cascatas bioquímicas complexas, que influenciam em processos biológicos diversos, como a migração celular, a angiogênese, a resistência a fármacos e a proliferação celular (BELKHIRI, *et al.*, 2016).

O mecanismo inibitório de PP1 por DARPP-32 é complexo e bidirecional, enquanto a fosforilação da DARPP-32 no resíduo de Thr34 por uma proteína quinase A (PKA) inibe a PP1, a fosforilação no resíduo Thr75 por uma quinase dependente de ciclina (CDK) inativa PKA e, conseqüentemente, a fosforilação do resíduo de Thr34, levando a uma ativação de PP1.

Desta forma, a DARPP-32 funciona como um centro modulador de PP1 e da atividade bioquímica, que por sua vez está sujeito a influência de substâncias diversas como as proteínas quinases, as AMPc e os neurotransmissores presentes em um dado tipo celular (AVANES, *et al.*, 2019).

A regulação de PP1 por meio dos resíduos fosforilados de DARPP-32 pode ser resumida de acordo com a figura abaixo, observe:

Figura 2: Regulação de PP1 e PKA através da fosforilação em sítios diversos de DARPP-32



Fonte: Belkhir, *et al.*, 2016

Estudos subsequentes revelaram que, os resíduos potencialmente fosforilados de DARPP-32 são diversos e a ação deles não se restringe a neurônios com receptores dopaminérgicos e relacionam-se, também, aos neurônios com receptores glutaminérgicos (SCHEGGI, *et al.*, 2018), serotoninérgicos e adrenérgicos (YGER & GIRAULT, 2011).

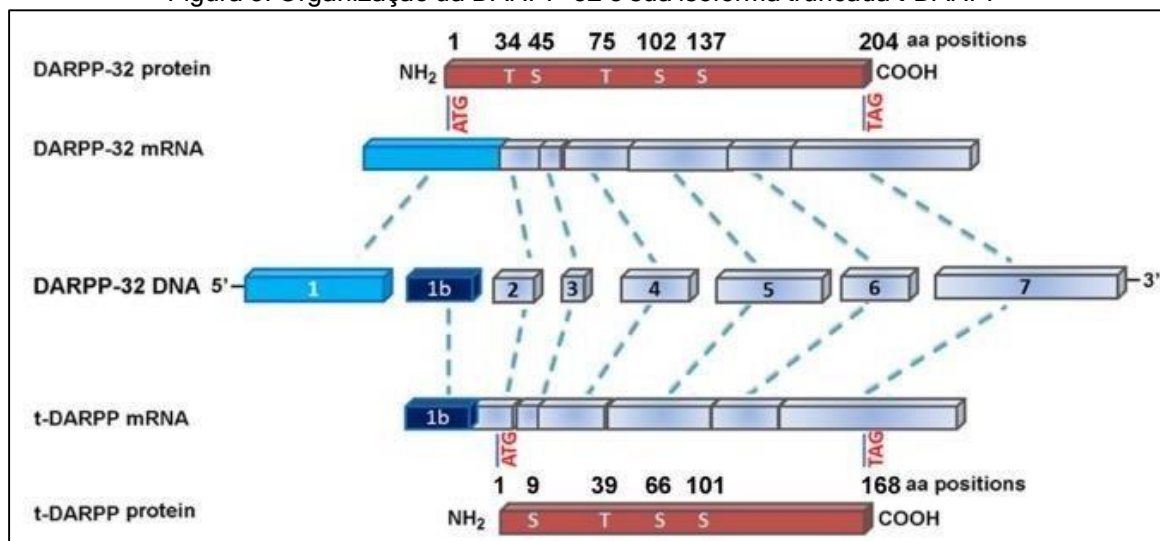
1.4.2 Estrutura, bioquímica e expressão gênica

O *locus* gênico de DARPP-32 está no cromossomo 17q12, sendo que o gene PPP1R1B que codifica a DARPP-32 pode, em razão de variações no *splicing*, gerar dez transcritos diferentes.

Esse gene é responsável tanto pela expressão da proteína DARPP-32, a qual possui 204 aminoácidos e cuja a região gênica codificante possui 615 nucleotídeos, quanto pela expressão de uma isoforma truncada que é denominada de t-DARPP. Essa isoforma foi descoberta em 2002 (AVANES, *et al.*, 2019), tendo apenas 168 aminoácidos na sequência primária e 507 nucleotídeos na sequência gênica contendo a região codificante.

A isoforma t-DARPP possui uma lacuna quanto ao resíduo de Thr34 que possui o efeito inibitório de PP1, sendo que, o códon de iniciação dessa proteína é o que codifica o aminoácido 37 da sequência de DARPP-32, resultando na isoforma t-DARPP que possui uma ausência dos 36 primeiros aminoácidos da forma clássica (AVANES, *et al.*, 2019). A organização dos genes, transcritos e proteínas estão representados na Figura 3.

Figura 3: Organização da DARPP-32 e sua isoforma truncada t-DARPP



DARPP-32 e t-DARPP têm a mesma sequência gênica do éxon 2 ao terminal 3'. Os resíduos fosforilados de DARPP-32 e t-DARPP divergem. O resíduo Thr34 está ausente na isoforma t-DARPP. Fonte: (Modificado de Belkhiri et al. 2016).

Estudos bioquímicos demonstram que a nível estrutural, a DARPP-32 trata-se de uma proteína resistente ao calor e com PH ácido, pois quando submetida a ambas condições não se observou mudança significativa nas propriedades de solubilidade e na fosforilação dos resíduos, especialmente, no resíduo Thr34 (HEMMINGS, *et al.*, 1984).

Trabalhos científicos que analisaram o viés composicional (alteração na frequência dos aminoácidos na sequência primária), aliado a um reduzido número de resíduos hidrofóbicos, sugerem que a DARPP-32 possa ser uma proteína classificada como intrinsecamente desordenada (DANCHECK, *et al.*, 2008). Observa-se que, uma das razões que atribui a DARPP-32 uma classificação como PID é a presença de uma região rica em glutamato no centro do domínio desse elemento (AVANES, *et al.*, 2019).

1.4.3 Doenças relacionadas

Nota-se que, muitos cânceres têm sido associados a forma truncada de DARPP-32 e a superexpressão da DARPP-32 que, também, tem sido relacionada com inúmeras enfermidades. Dentre os exemplos de doenças observadas tendo papel significativo da DARPP-32 ou t-DARPP-32 na patogênese e na evolução, cita-se:

- a- Desordens de movimento: observou-se em pacientes com Mal de Parkinson, que apesar dos níveis de DARPP-32 não estarem alterados significativamente, esta proteína contribui para a ocorrência da discinesia induzida por L-DOPA, que é um tipo estereotipado de movimentos anormais da musculatura, sendo um efeito colateral do tratamento medicamentoso de Parkinson com o medicamento L-DOPA (GIRAULT & NAIRN, 2021);
- b- Doenças psiquiátricas: estudos apontam que DARPP-32 e a forma truncada que possui podem estar envolvidas na etiologia de transtornos neurobiológicos de manifestação psiquiátrica como o TDAH, esquizofrenia e transtorno bipolar. Pacientes com esquizofrenia e transtorno bipolar tiveram detectados altos níveis da isoforma t-DARPP no córtex pré-frontal e dorsolateral, em um amplo estudo de casos. Por outro lado, observou-se que a isoforma, DARPP-32, que participa da inibição de PP1 e PKA de acordo com a fosforilação e desfosforilação, respectivamente, teve uma redução de expressão nesses mesmos tecidos. A redução na expressão de DARPP-32, possivelmente, afeta o funcionamento dos neurônios dopaminérgicos. Todavia, não se acredita que o gene de DARPP-32 seja um gene de risco para esquizofrenia e, provavelmente, a participação na etiologia da doença faz-se com a interação com outros genes. Um amplo estudo realizado em outras partes do mundo com pacientes esquizofrênicos e com tendências suicidas indicou uma possível relação entre o fator neurotrófico derivado do cérebro (BDNF) e a expressão de DARPP-32. Neste sentido, foram identificadas cinco variantes polimórficas de DARPP-32 como descrito a seguir.
- 1- Substituição de uma guanina por uma citosina na posição 2036 do cromossomo 17;
 - 2- No éxon 2 (não traduzido), há ausência de uma guanina;
 - 3- No íntron 2, há uma substituição de uma guanina por uma adenina;
 - 4- No íntron 6, há uma substituição de uma adenina por uma guanina;
 - 5- No éxon 6, ocorre a deleção de uma trinca de pares de bases resultando na ausência de um glutamato.

Desta forma, é possível observar que a expressão de DARPP-32, os polimorfismos e a interação com moléculas como BDNF, tenham alguma relação com o desenvolvimento da esquizofrenia (WANG, *et al.*, 2017).

Acredita-se que, devido a estreita relação entre a DARPP-32 e a dopamina, assim como, a importância deste neurotransmissor na esquizofrenia, é possível que esta proteína e as mutações nas regiões fosforiladas possam interferir na evolução dessa doença (GIRAULT & NAIRN, 2021).

c- Câncer - A descoberta da relação entre a DARPP-32 com o câncer teve início com o descobrimento da t-DARPP, pois, desde então, verificou-se que a isoforma truncada e a superexpressão de DARPP-32 influenciam na formação dos cânceres de estômago, mama, pulmão, cólon, esôfago e próstata. Uma série de pesquisas identificou um aumento no número de cópias de genes no braço q do cromossomo 17 e um dos genes com essa elevação na transcrição é o gene PPP1R1B e as isoformas DARPP-32 e a t-DARPP. Logo, foram constatados níveis anormais em amostras teciduais de câncer no estômago, e observou-se que, o locus gênico de PPP1R1B na região cromossômica 17q12 é, frequentemente, amplificado e as isoformas DARPP-32 e t-DARPP são superexpressas enquanto a frequência de mutações no gene PPP1R1B é baixa, sugerindo, então, que o aumento de níveis de t-DARPP e DARPP-32 é mais relevante para a oncogênese do que a mutação gênica (AVANES, *et al.*, 2019).

A principal via bioquímica afetada por DARPP-32 e a isoforma truncada é a via de ativação de AKT, que é uma proteína relacionada a sobrevivência celular e a inibição de apoptose. Algumas pesquisas revelaram que a fosforilação de AKT no resíduo de serina 472, proporciona a inibição apoptótica e a sobrevivência das células cancerígenas, e isto ocorre em caso de superexpressão de t-DARPP. O mesmo, não se verifica em camundongos nocaute para t-DARPP. Logo, é provável, que a isoforma truncada tenha certa influência na formação de câncer de mama em razão da ativação da via AKT (AVANES, *et al.*, 2019).

Outras vias bioquímicas que levam a formação de tumores afetados por superexpressão de DARPP-32 e t-DARPP são: vias de PKA, IGF1R e Bcl2, pois alguns trabalhos indicam que t-DARPP interfere no funcionamento de PKA formando um complexo em associação com esta molécula, que ativa fatores de transcrição. Nota-se que, a t-DARPP, também, foi identificada formando complexos com o receptor do fator 1 de crescimento tipo insulina em câncer de mama e, também, que a superexpressão de t-DARPP aumenta os níveis da proteína Bcl2, que está envolvida na inibição de apoptose (AVANES, *et al.*, 2019). As vias bioquímicas envolvendo as moléculas NF- κ B, NF κ B, ANGPT2, CXCR4, Her1 e Her2 são outros tipos de vias bioquímicas que têm interação com a superexpressão de t-DARPP, principalmente, na formação do câncer de pulmão, do gástrico e do esôfago.

Um trabalho de Chinnaiyan e colaboradores identificou a formação de um transcrito quimérico, que surge de dois genes localizados em *tandem* no cromossomo 17, já o gene PPP1R1B junto com o gene adjacente expressam a proteína STARD3, em 21% dos cânceres primários do estudo. Nota-se, também, que a proteína quimérica de t-DARPP e STARD3 aumenta a fosforilação de AKT (AVANES, *et al.*, 2019).

1.5 Proteínas intrinsecamente desordenadas

Na última década, foi identificado um tipo especial de proteínas que caracterizam por não terem uma estrutura tridimensional conhecida e não serem possíveis de ser estudadas por técnicas de cristalografia ou serem organizadas em uma estrutura estável em uma solução. Essas proteínas são denominadas proteínas intrinsecamente desordenadas (PID) e as propriedades delas são objetivos de inúmeras pesquisas por bioquímicos, biólogos estruturais e profissionais de bioinformática.

As PIDs são consideradas uma categoria de proteínas, em que a existência de um estágio intermediário que é formado de acordo com a ligação específica a uma dada molécula, pode ser mais vantajoso de manter-se em termos de estabilidade energética

que a manutenção como uma proteína de estrutura definida e integralmente estável (ESPINOSA, *et al.*, 2009).

A plasticidade conformacional, que é a propriedade de uma proteína arranjar-se em diferentes estados conformacionais em intervalos de tempo variáveis e dependendo da molécula ligante a este grupo de proteínas, permite realizar ligações com outras proteínas, ácidos nucleicos e outras moléculas que levam as funções e as atividades biológicas diversas e características (PIERSIMONI, *et al.*, 2022). Como exemplos conhecidos de proteínas desordenadas têm-se as proteínas p27 e p53 que atuam na regulação da atividade de proteíno quinases e no ciclo da divisão celular (NELSON & COX, 2019).

As PIDs têm sido associadas a funções regulatórias de sinalização bioquímica que atuam em processos biológicos variados, como diferenciação celular, transcrição, mitose, meiose e o transporte de proteínas (HABCHI, *et al.*, 2014). Essa diversidade de efeitos biológicos é resultado do conjunto de conformações que estas proteínas podem adotar com ligantes diversos que, por sua vez, acarretam funções bioquímicas variadas a nível molecular (HABCHI, *et al.*, 2014).

Pelo entendimento científico recente, tais proteínas e os estados de desordem estrutural que elas possuem, podem ser reversíveis e variáveis de acordo com o ligante. Estas biomoléculas têm sido associadas a uma variedade de fenômenos biológicos, os quais são resultado de uma grande flexibilidade conformacional e da enorme variabilidade de interação intermolecular (ESPINOSA, *et al.*, 2009).

A descoberta das PIDs foi um achado que colocou em dúvida a teoria do modelo chave-fechadura, consolidado na bioquímica e, que explica a função de grande parte das funções proteicas, especialmente, a função enzimática (HABCHI, *et al.*, 2014). Desde então, várias PIDs têm sido associadas a um grande número de doenças como câncer, diabetes e doenças neurodegenerativas como Parkinson e Alzheimer (HABCHI, *et al.*, 2014). Acredita-se que de 20% a 30% das proteínas existentes nos mamíferos possam pertencer ao grupo das proteínas intrinsecamente desordenadas.

Não há estrutura tridimensional conhecida para DARPP-32 e, nem tampouco, para PPP1R1A e PPP1R1C descritas na literatura e/ou disponíveis em bancos de dados até o presente momento.

2 JUSTIFICATIVA

Este trabalho busca uma compreensão das propriedades e funções de uma família de proteínas, que se mostra como um tema muito relevante na genética molecular e bioquímica evolutiva, as Proteínas Intrinsecamente Desordenadas. Em vista disso, perceber a identificação do domínio proteico e das propriedades físico-químicas dessas proteínas auxiliam a compreensão das razões delas serem classificadas como PIDs.

Por domínio, entende-se uma parte da cadeia de proteína que tem estabilidade e pode manter as propriedades bioquímicas e funcionais, independentemente, do restante da proteína (NELSON & COX, 2019). Sendo assim, este trabalho pretende colaborar para outras pesquisas futuras relacionadas ao entendimento destas proteínas amplamente presentes em mamíferos, incluindo os seres humanos.

Evidências mostram que tais proteínas são dotadas de plasticidade conformacional, bioquímica e fisiológica notáveis resultando em inúmeros efeitos biológicos. Logo, o entendimento sobre DARPP-32, PPP1R1A e PPP1R1C pode propiciar um arcabouço para a pesquisa e desenvolvimento de fármacos e o tratamento de enfermidades que ocorrem na sociedade humana em escala global.

Tais enfermidades causam sofrimento individual, coletivo e grande ônus aos sistemas de saúde como as inúmeras doenças complexas e de amplo espectro como os cânceres, a esquizofrenia, a dependência química e os transtornos neuromotores. Em resumo, este trabalho tem o potencial de contribuir na compreensão a respeito da estrutura e função de um grupo atípico de proteínas, ajudando na seleção e identificação de alvos moleculares e no desenvolvimento de fármacos contra um grande número de doenças humanas.

3 OBJETIVOS

3.1 Objetivo geral

Avaliar a biodiversidade molecular da DARPP-32 e dos potenciais homólogos dessa proteína.

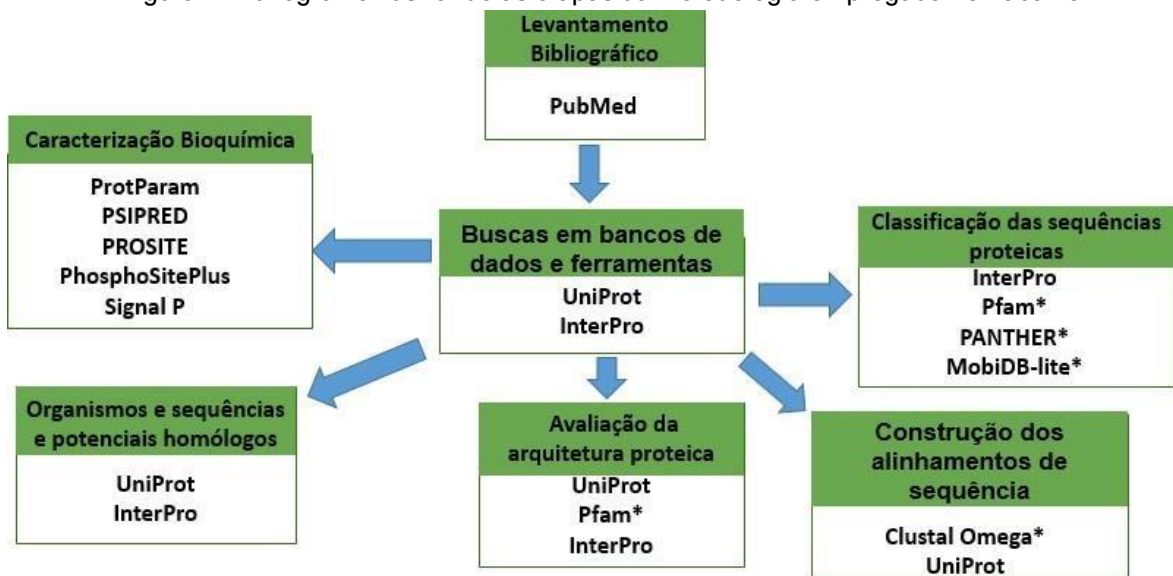
3.2 Objetivos específicos

- Identificar potenciais homólogos da DARPP-32 em humanos;
- Caracterizar as propriedades bioquímicas;
- Avaliar a arquitetura proteica dos alvos moleculares selecionados;
- Construir alinhamentos de múltiplas sequências;
- Identificar padrões de conservação e divergência nas sequências.

4 METODOLOGIA

A Figura 4, abaixo, apresenta a Metodologia usada no presente trabalho.

Figura 4: Fluxograma ilustrando as etapas da metodologia empregada no trabalho.



Autoria própria - 2023

Este trabalho foi desenvolvido utilizando bancos de dados e ferramentas de bioinformática disponíveis na Web. Tem-se a seguir, uma breve descrição a respeito de cada um destes sites:

PUBMED – banco de dados que possibilita a busca de artigos, textos e outros tipos de trabalhos escritos sobre pesquisas biológicas e biomédica, mantido pelo NCBI-NLM-NIH (<https://pubmed.ncbi.nlm.nih.gov>).

UNIPROT – banco de dados universal com informações coletadas e agrupadas de diferentes outros bancos de dados e ferramentas da bioinformática, provendo assim, uma fonte variada de anotação de proteínas e sequências proteicas (CHEN, *et al.*, 2017). Deste modo, fornecendo diversas informações sobre anotação proteica, o UniProt alia o conhecimento coletado a partir da literatura científica com informações

de assinatura proteica geradas através de Técnicas de Aprendizado de Máquina (AM) da Inteligência Artificial (IA).

INTERPRO – plataforma integrada de 13 diferentes bancos de dados, que fornece informações sobre assinatura, estrutura e função de proteínas, permitindo assim a classificação dessas em famílias, as quais possuem um identificador, que as caracteriza e permite o estudo amplo de tais proteínas em tópicos, como a presença de domínios e sítios ativos (CHEN, *et al.*, 2017). A versão 90.0 do InterPro possui 40.597 entradas referentes a um total de 53.784 assinaturas obtidas dos bancos de dados integrantes. Essa versão do InterPro incorporou o banco de dados Pfam e adicionou informações a respeito da predição de estrutura proteica dos bancos de dados AlphaFold e RoseTTAFold. O InterPro em adição as informações presentes nos bancos de dados, membros de assinatura proteica geradas por modelos preditivos, como HMMs e *fingerprint*, agora, também, reúne informações baseadas em métodos de Inteligência Artificial e Aprendizado Profundo (DL) (LAFOSSE, *et al.*, 2022).

BLAST – pacote de ferramentas de bioinformática do NCBI-NLM-NIH para detecção de similaridades entre sequências moleculares, baseado em buscas de alinhamento local básico (BORATYN, *et al.*, 2013) e (SAYERS, *et al.*, 2021). O BLAST está implementado no UniProt.

PANTHER – banco de dados sobre genomas, anotação funcional de genes e vias bioquímicas, que permite o estudo e a classificação de centenas de famílias de genes e a compreensão da evolução deles. Para a identificação de uma família gênica ou proteica, o PANTHER vale-se de assinaturas baseadas em biocuradoria manual e reconstrução filogenética, alinhamentos de múltiplas sequências e família baseada em HMMs (THOMAS, *et al.*, 2021).

PFAM – banco de dados de famílias de proteínas baseado na identificação dos domínios proteicos (CHEN, *et al.*, 2017) e (MISTRY, *et al.*, 2021). Esse banco de dados foi descontinuado no fim de 2022. Sendo que, desde janeiro de 2023, ele está hospedado no InterPro (PAYSAN-LAFOSSE, *et al.*, 2022). As famílias proteicas do Pfam são construídas a partir de um alinhamento de múltiplas sequências e a

construção de um modelo oculto de Markov (HMM), usando um pacote de ferramentas de HMMER (MISTRY, *et al.*, 2021).

MOBIDB-LITE – banco de dados sobre predição de desordem estrutural em proteínas e a caracterização de tais regiões, considerando as propriedades bioquímicas, como a flexibilidade, a basicidade e a acidez (DI DOMENICO, *et al.*, 2012). A apresentação de predição de regiões intrinsecamente desordenadas de proteínas pelo MobiDB-lite segue dois passos: primeiramente, uma região de consenso desordenada obtida pelo mínimo de cinco entre oito diferentes bancos de dados preditores de PIDs é escolhida. Em seguida, ocorre um processo de filtragem para eliminar curtas sequências desordenadas, até que permaneçam sequências com o mínimo de vinte resíduos em regiões desordenadas (NECCI, *et al.*, 2017).

PROTPARAM – ferramenta da bioinformática disponível no portal ExPASy (<https://www.expasy.org>), mantido pelo Swiss Institute of Bioinformatics (SIB), que gera dados sobre parâmetros físico-químicos de uma proteína específica como composição de aminoácidos, peso molecular e hidrofobicidade. Geralmente, as proteínas, cujas informações físico-químicas são fornecidas, estão depositadas no Swiss-Prot ou TrEMBL (WILKINS, *et al.*, 2005).

PSIPRED – ferramenta de bioinformática disponibilizada pela UCL - London's Global University que reúne métodos diversos para a predição estrutural e funcional de proteínas. Nota-se que, para as predições estruturais e funcionais de proteínas, o PSIPRED utiliza métodos baseados em matrizes de pontuação específicas de posição (BUCHAN, *et al.*, 2019).

PHOSPHOSITEPLUS – banco de dados que reúne diversas informações experimentais registradas sobre diversos tipos de modificações pós-traducionais (MPTs), tais como, fosforilação, metilação e glicosilação. O PhosphoSitePlus vale-se de informações sobre MPTs coletadas de estudos de Espectrometria de Massas e da literatura científica com biocuradoria manual, que são revisadas por algoritmos como o Ascore (HORNBECK, *et al.*, 2014).

PROSITE – ferramenta de bioinformática disponível no portal ExPASy, que gera informações sobre domínios proteicos, motivos e sítios. O PROSITE possibilita a identificação e a descrição de perfis e padrões proteicos relacionados às proteínas em estudo. O PROSITE realiza uma busca por padrões e perfis proteicos utilizando sequências conhecidas e, também, uma busca combinatória, que permite reconhecer domínios de acordo com sua funcionalidade ou estrutura (SIGRIST, *et al.*, 2013).

CLUSTAL OMEGA – software para a realização de alinhamento múltiplo de sequências moleculares mantido pelo EMBL-EBI (SIEVERS, *et al.*, 2011). O Clustal Omega está implementado no UniProt, e utiliza o algoritmo mBed e o método HAlign para o alinhamento de perfil HMM (SIEVERS & HIGGINS, 2021).

SIGNAL P 6.0 – ferramenta de bioinformática que prediz a presença e a localização de peptídeo sinal em sequências de proteínas. O Signal P6.0 utiliza uma Técnica de Aprendizado de Máquina para a predição do peptídeo sinal nas proteínas (TEUFEL, *et al.*, 2022).

4.1 Levantamento bibliográfico

O levantamento bibliográfico a respeito das PPP1R1A, PPP1R1B (DARPP-32) e PPP1R1C foi realizado considerando o período a partir de 1983, ano em que, a DARPP-32 foi nomeada como tal. As seguintes palavras e expressões foram usadas como palavras-chave: DARPP-32 and câncer, DARPP-32 and mental diseases, DARPP-32 and Parkinson, PPP1R1A e PPP1R1C. Em relação a família IPP-1 e PIDs e os bancos de dados e ferramentas de bioinformática utilizados neste trabalho, não houve seleção por período e, também, foram empregados, como bibliografia e modelo para o desenvolvimento desta escrita, outros livros, assim como uma dissertação.

4.2 Buscas em bancos de dados e ferramentas

4.2.1 Busca por sequências e identificação de potenciais homólogos

Primeiramente, para que, fossem realizados os trabalhos em diferentes bancos de dados e ferramentas de bioinformática, as sequências em formato FASTA de

PPP1R1A, PPP1R1B e PPP1R1C foram obtidas através do banco de dados UniProt (<https://www.uniprot.org/>), usando como filtro o domínio PF05395 e *Homo sapiens* (taxid:9606). Observa-se que, desde o final de 2022, o Pfam (<http://pfam.xfam.org/>) foi integrado ao InterPro (<https://www.ebi.ac.uk/interpro/>) facilitando o acesso. Assim, junto ao UniProt, esse último fornece grande parte da anotação proteica, especialmente, ao que se refere a sequências, domínios, estruturas, funções e doenças relacionadas às proteínas.

No InterPro, além da pesquisa realizada através da sequência em formato FASTA, foi realizada, também, a busca baseada em texto, a qual buscou-se pelo nome da família proteica, da proteína ou pelo domínio para gerar informação. Para a realização da busca por similaridade e identificação de potenciais homólogos através do BLAST, implementado no UniProt, o grupo taxonômico foi estendido para o táxon Hominidae (taxid:9604), que inclui os primatas filogeneticamente mais próximos ao ser humano, e a filtragem considerou um hit de 100 sequências, nas quais a porcentagem de identidade considerada foi a partir de 80, sendo que, o BLAST foi acessado usando as sequências em formato FASTA obtidas nesta busca.

4.2.2 Avaliação da arquitetura proteica

A arquitetura proteica refere-se a presença e a organização de domínios proteicos, sejam eles estruturais ou funcionais. Para avaliar a arquitetura proteica da DARPP-32 (PPP1R1B) e dos parálogos PPP1R1A e PPP1R1C foi utilizado o UniProt, buscando pelo domínio PF05395 e *Homo sapiens* (taxid:606). Logo, as informações coletadas incluem o início e fim do domínio, o comprimento de cada proteína, as regiões desordenadas, a presença de viés na estrutura proteica, de resíduos modificados e a existência da proteína em nível experimental ou se foi inferida por homologia.

4.2.3 Classificação da sequência proteica

A classificação da sequência proteica foi realizada através do InterPro, que classifica as proteínas em famílias, baseando-se na presença de domínios e outras propriedades estruturais e/ou funcionais das proteínas alvo. O mecanismo de busca

foi feito através da opção *text* na qual o Entry Name da proteína é utilizado para gerar a informação pretendida.

4.3 Caracterização bioquímica das proteínas de interesse

A caracterização bioquímica das proteínas, membros da família IPP-1, em especial da DARPP-32 (PPP1R1B), foi realizada através da sequência FASTA como entrada, utilizando os seguintes bancos de dados e ferramentas de bioinformática: ProtParam, PSIPRED e PhosphoSitePlus, PROSITE e SignalP 6.0.

No ProtParam (<https://web.expasy.org/protparam>), os parâmetros físico-químicos avaliados para cada uma das três proteínas (PPP1R1A, PPP1R1B e PPP1R1C) foram: número de aminoácidos, massa molecular, ponto isoelétrico teórico (pI) e fórmula molecular.

Através do PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred>), a estrutura secundária de cada uma das três proteínas pôde ser classificada, considerando-se critérios como se a proteína é intracelular ou citoplasmática, se possui estrutura em hélice, se contém regiões desordenadas, dentre outras características. Esta ferramenta, também, possibilita a classificação de cada aminoácido constituinte da proteína baseada nas seguintes propriedades físico-químicas: pequeno e apolar, hidrofóbico, polar e aromático com cisteína.

As propriedades físico-químicas dos aminoácidos como a massa molecular, o pH, a polaridade, a hidrofobicidade, dentre outras, são determinantes quanto ao tipo de estrutura secundária, terciária e quaternária de uma proteína. Ou seja, em conjunto, tais propriedades podem determinar o comportamento funcional a nível bioquímico e fisiológico.

O PROSITE (<https://prosite.expasy.org>) foi usado a fim de obter informações sobre o perfil e o padrão proteicos. Assim, por perfil proteico, entende-se a ocorrência de um determinado domínio, motivo ou de uma região rica em determinado aminoácido na

sequência da proteína, enquanto o padrão proteico refere as propriedades bioquímicas da sequência proteica.

Por exemplo, as possíveis modificações pós-traducionais (MPTs), que podem ocorrer na estrutura proteica. Enquanto o domínio descreve uma região com um grau de independência funcional e estrutural dentro de uma sequência de polipeptídeo, o motivo pode ser considerado como uma sequência, geralmente, curta de uma cadeia polipeptídica que tem um padrão característico de enovelamento entre um ou mais aminoácidos (NELSON & COX, 2019).

Buscou-se, também, a identificação dos possíveis resíduos fosforilados, para tanto, empregou-se o PhosphoSitePlus (<https://www.phosphosite.org>), como também, a fim de prever a existência de peptídeo sinal, foi utilizado o SignalP 6.0 disponível no website <https://services.healthtech.dtu.dk/services/SignalP-6.0>.

4.4 Construção dos alinhamentos de múltiplas sequências

A construção do alinhamento das sequências de PPP1R1A, PPP1R1B e PPP1R1C foi realizada no Clustal Omega (Sievers & Higgins 2021), implementado no UniProt (UniProt Consortium 2023), selecionando as sequências em formato FASTA e que tiveram biocuradoria manual registrada no banco de dados.

5 RESULTADOS

5.1 Levantamento bibliográfico

O levantamento bibliográfico realizado no banco de literatura PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), tendo como palavra-chave somente DARPP-32 ou DARPP-32 associado a termos como câncer ou Parkinson, foi executado considerando publicações a partir do ano de 1983. A busca apresentou um total de 1.288 artigos, dos quais 106 tratavam de revisão, dois artigos eram ensaios de controle randomizado e três artigos de ensaios clínicos.

Além dos artigos sobre DARPP-32 e suas parálogas, selecionadas de acordo com o resultado acima, também, foram coletados artigos diversos sobre bancos de dados, ferramentas de bioinformática e PIDs, uma dissertação e um livro também foram inclusos como material bibliográfico para a elaboração deste trabalho.

5.2 Buscas em bancos de dados e ferramentas

5.2.1 Organismos e sequências e identificação de potenciais homólogos

A busca por organismos e sequências realizada no UniProt e InterPro usando as sequências FASTA e/ou o domínio PF05395 aliado aos táxons *Homo sapiens* (taxid:9606) e Hominidae (taxid:9604) retornou com um total de 13 proteínas. Paralelo a isso, quando a busca foi realizada usando o domínio do Pfam aliado ao táxon *Homo sapiens*, como pode ser visto na Tabela 1, houve um total de 53 proteínas e cinco organismos, isso quando a busca utilizou o domínio PF05395 e o taxon Hominidae. Já no resultado da busca usando o domínio PF05395 e o táxon *Homo sapiens*, as três proteínas, que possuem biocuradoria manual correspondem, respectivamente, a PPP1R1A, PPP1R1B e PPP1R1C.

Quadro 1: Proteínas com o domínio PF05395 em *Homo sapiens*

Entry	Entry Name	Gene Names	Length	Protein existence
B3KVQ9	B3KVQ9_HUMAN	PPP1R1B hCG_30016	204	Evidence at transcript level
F8VR76	F8VR76_HUMAN	PPP1R1A	64	Evidence at protein level
H0YIU4	H0YIU4_HUMAN	PPP1R1A	90	Evidence at protein level
J3KSJ8	J3KSJ8_HUMAN	PPP1R1B	144	Evidence at protein level
J3KT77	J3KT77_HUMAN	PPP1R1B	171	Evidence at protein level
Q13522	PPR1A_HUMAN	PPP1R1A IPP1	171	Evidence at protein level
Q6ZVX3	Q6ZVX3_HUMAN		153	Evidence at transcript level
Q8WVI7	PPR1C_HUMAN	PPP1R1C	109	Evidence at protein level
Q9NNW1	Q9NNW1_HUMAN		98	Evidence at transcript level
Q9UD71	PPR1B_HUMAN	PPP1R1B DARPP32	204	Evidence at protein level
R4GMQ8	R4GMQ8_HUMAN	PPP1R1A	153	Evidence at protein level
A0A024R1R3	A0A024R1R3_HUMAN	PPP1R1B,Hcg_30016	168	Inferred from homology
Q53TF4	Q53TF4_HUMAN	LOC 151242	47	Inferred from homology

Fonte: UniProt

As sequências em formato FASTA de PPP1R1A, PPP1R1B e PPP1R1C obtidas a partir do UniProt estão apresentadas no quadro 2.

Quadro 2: Sequências de PPP1R1A, PPP1R1B e PPP1R1C

PPR1A	>sp Q13522 PPR1A_HUMAN Protein phosphatase 1 regulatory subunit 1A OS=Homo sapiens OX=9606 GN=PPP1R1A PE=1 SV=2 MEQDNSPRKIQFTVPLLEPHLDPEAAEQIRRRRPTPATLVLTSDQSSPEIDEDRIPN PHLKSTLAMSPRQRKMTTRITPMKELQMMVEHHLGQQQQGEEPEGAAESTETQ ESRPPGIPDTEVESRLGTSGTAKKTAECIPKTHERGSKEPSTKEPSTHIPPPLDSKGA NSV
PPR1B	>sp Q9UD71 PPR1B_HUMAN Protein phosphatase 1 regulatory subunit 1B OS=Homo sapiens OX=9606 GN=PPP1R1B PE=1 SV=2 MDPKDRKKIQFSVPAPPSQLDPRQVEMIRRRRPTPAMLFRLSEHSSPEEEASPHQ RASGEGHHLKSKRPNPCAYTPPSLKAVQRIAESHLSISNLNENQASEEDELGEL RELGYPREDEEEEDDEEEEDSQAENVLKVIRQSAGQKTTTCGQGLEGPWER PPPLDESERDGGSEDQVEDPALSEPGEERPSPSEPGT
PPR1C	>sp Q8WVI7 PPR1C_HUMAN Protein phosphatase 1 regulatory subunit 1C OS=Homo sapiens OX=9606 GN=PPP1R1C PE=1 SV=1 MEPNPKKIQFAVPVFQSQIAPEAAEQIRRRRPTASLVILNEHNPPEIDDKRGPNT QGELQNASPKQRKQSVYTPPTIKGVKHLKGQNESAFPEEEEGTNEREEQRDH

O cabeçalho contém identificadores, nome da proteína e táxon. Fonte: UniProt

5.2.2 Avaliação da arquitetura proteica

A avaliação da arquitetura proteica de PPP1R1A, PPP1R1B e PPP1R1C realizada no UniProt, utilizando como entrada o domínio PF05395 e o táxon *Homo sapiens* resultou nas informações mostradas abaixo:

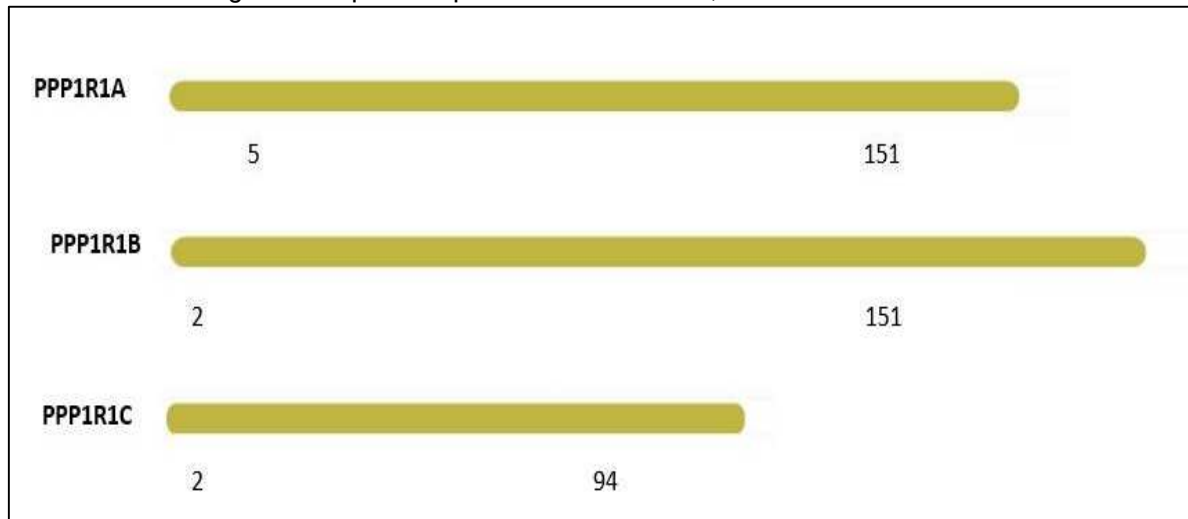
Quadro 3: Arquitetura de PPP1R1A, PPP1R1B e PPP1R1C de *Homo sapiens*

Entry Name	PPR1A_HUMAN	PPR1B_HUMAN	PPR1C_HUMAN
Length	171	204	109
Pfam domain	5-151	2-151	2-94
Disordered region	1-171	1-204	25-109
Modified residue	YES	YES	NO
Compositional bias	YES	YES	YES
Protein existence	Evidence at protein level	Evidence at protein level	Evidence at protein level

Arquitetura baseada em sequências proteicas. Fonte: UniProt

A próxima figura mostra a arquitetura proteica de PPP1R1A, PPP1R1B e PPP1R1C.

Figura 5: Arquitetura proteica de PPP1R1A, PPP1R1B e PPP1R1C

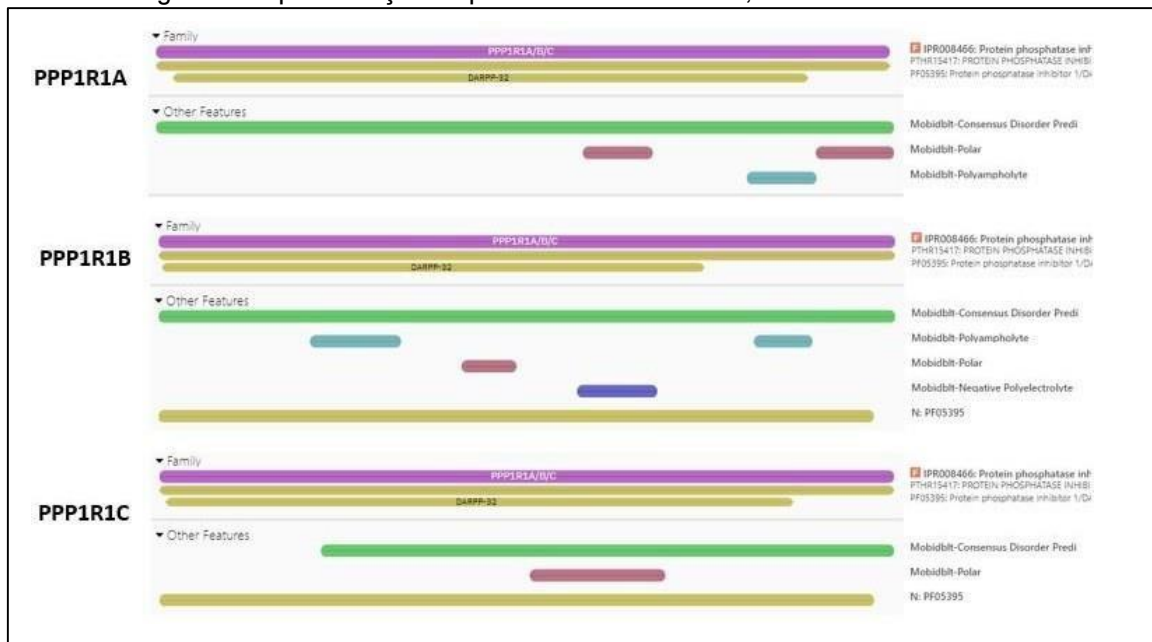


Posições de início e fim do domínio (Pfam: PF05395) estão indicadas. Fonte: InterPro

5.2.3 Classificação das sequências proteicas

As sequências proteicas foram classificadas de acordo com o InterPro em proteínas constituintes da família de Inibidores de PP1 (IPP-1), tendo o seguinte identificador neste banco de dados: IPR008466. Além da classificação em família, o InterPro, no qual o Pfam foi incorporado, por ser uma rede integrada contendo 13 bancos de dados, permitiu obter informações individuais sobre cada uma das três proteínas estudadas, fornecidas pelo PANTHER e MobiDB-lite, além do Pfam, conforme é mostrado na próxima figura.

Figura 6: Representação esquemática de PPP1R1A, PPP1R1B e PPP1R1C



Os identificadores do InterPro, PANTHER, Pfam e MobiDB-lite estão indicados. Fonte: InterPro

O resultado sobre a classificação da sequência proteica realizado no InterPro, gera, de acordo com cada banco de dados de saída resultante e selecionado, as seguintes informações sobre PPP1R1A:

- a) **INTERPRO** - PPP1R1A faz parte da família de inibidores de PP1 (IPP-1) que tem como identificador IPR008466;
- b) **PANTHER** - subunidade 1A da fosfoproteína neuronal regulada por Dopamina e AMPc, inibidora de proteína fosfatase, tendo como identificador PTHR15417;
- c) **PFAM** - proteína inibidora de proteína fosfatase 1 com identificador PF05395;
- d) **MOBIDB-LITE** - proteína com predição de desordem consenso, contendo as seguintes regiões:
 - Região polar: 100 ao 115 aminoácido e 154 ao 171 aminoácido;
 - Região de polianfólito (ácido e básico): 138 ao 153 aminoácido.

Para PPP1R1B, o resultado sobre a classificação da sequência proteica, gerado no InterPro, trouxe para cada banco de dados de saída resultante e selecionado, as seguintes informações:

- a) **INTERPRO** - PPP1R1B integra a família de proteínas inibidoras de PP1 (IPP-1) tendo como identificador IPR008466;
- b) **PANTHER** - subunidade 1B da fosfoproteína neuronal regulada por dopamina e AMPc inibidora de proteína fosfatase, tendo como identificador PTHR15417;
- c) **PFAM** - proteína inibidora de proteína fosfatase, tendo como identificador PF05395;
- d) **MOBIDB-LITE** - proteína com predição de desordem consenso, contendo as seguintes regiões:
 - Região polar: do 85 ao 99 aminoácido;
 - Região de polianfólito: do 43 ao 67 aminoácido e 166 ao 181 aminoácido;
 - Região de polieletrólito negativa: 117 ao aminoácido 138.

De acordo com os resultados sobre a classificação da sequência proteica de PPP1R1C originados no InterPro, as informações relativas a cada banco de dados resultantes e selecionados são:

- a) **INTERPRO** - proteína membro da família de inibidores de PP1 tendo como identificador IPR008466;
- b) **PANTHER** - subunidade 1C da fosfoproteína neuronal regulada por dopamina e AMPc, inibidora de proteína fosfatase, tendo como identificador PTHR15417;
- c) **PFAM** - proteína inibidora de proteína fosfatase, tendo como identificador PF05395;
- d) **MOBIDB-LITE** - proteína com predição de desordem consenso, contendo a seguinte região: Região polar: 56 ao 75 aminoácido.

5.3 Caracterização bioquímica das proteínas de interesse

5.3.1 ProtParam

Inicialmente, a caracterização bioquímica da DARPP-32 (PPP1R1B), PPP1R1A e PPP1R1C foi realizada no ProtParam, que é uma ferramenta de predição de

parâmetros físico-químicos, podendo ser encontrada no portal ExPASy, acessível em <https://web.expasy.org/protparam/>. Os parâmetros avaliados e coletados para cada uma das três proteínas de interesse neste trabalho são os mostrados na tabela a seguir:

Tabela 1: Propriedades físico-químicas de PPP1R1A, PPP1R1B e PPP1R1C

Proteína	Aminoácidos*	Massa molecular	pI	Fórmula Molecular
PPP1R1A	171	22962.77	4.48	C ₉₆₇ H ₁₅₂₆ N ₂₈₈ O ₃₅₁ S
PPP1R1B	204	19011.29	5.9	C ₈₀₆ H ₁₃₂₆ N ₂₄₄ O ₂₇₂ S
PPP1R1C	109	12345.71	6.31	C ₅₃₄ H ₈₅₆ N ₁₆₂ O ₁₇₃ S ₁

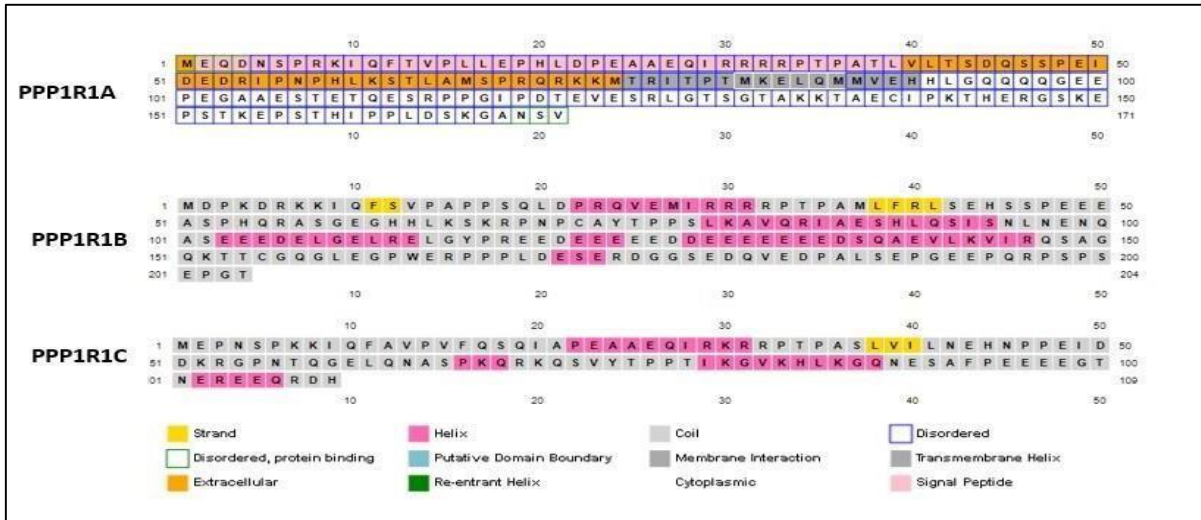
(*) Total de resíduos de aminoácidos na sequência. Fonte: ProtParam

5.3.2 PSIPRED

A caracterização bioquímica feita através do PSIPRED possibilita a caracterização da estrutura secundária das proteínas em questão, distinguindo regiões da sequência e baseando-se em propriedades como a localização extracelular ou citoplasmática, também, se a região é possuidora de estrutura em hélice, se contém regiões desordenadas ou de interação de membranas, dentre outras características.

Outra forma de caracterização bioquímica utilizada pelo PSIPRED, classifica, individualmente, cada aminoácido constituinte da sequência polipeptídica de acordo com quatro de propriedades físico-químicas: pequeno e apolar, hidrofóbico, polar e aromático com a cisteína.

Figura 7: Estrutura secundária de PPP1R1A, PPP1R1B e PPP1R1C



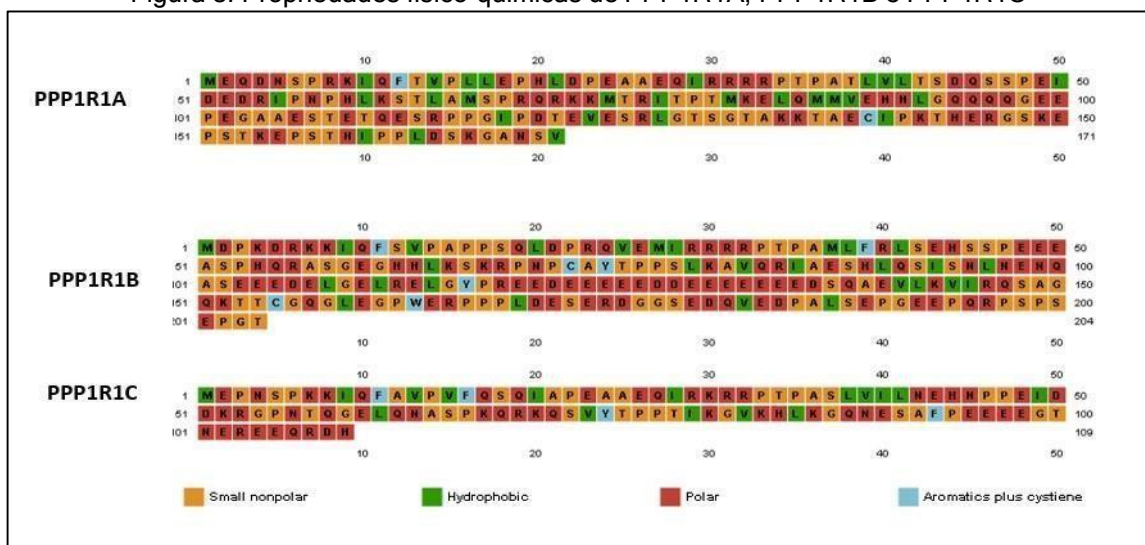
As propriedades bioquímicas das proteínas estão indicadas. Fonte: PSIPRED
De acordo com a classificação da estrutura secundária e as propriedades bioquímicas, o resultado para cada uma das três proteínas foi:

PPP1R1A: O 1º aminoácido, metionina, e a sequência que vai do 40º aminoácido ao 74º aminoácido, caracterizam-se por serem extracelulares e desordenados;
Do 2º aminoácido ao 4º aminoácido, a região caracteriza-se por ser desordenada e ligante a proteína;
Do 5º aminoácido ao 39º aminoácido, a região caracteriza-se como desordenada;
Do 75º aminoácido ao 80º aminoácido e do 87º aminoácido ao 90º aminoácido, a região caracteriza-se como desordenada e formadora de hélice transmembrana;
Do 81º aminoácido ao 86º aminoácido, a região caracteriza-se como helicoidal e transmembrana;
Do 91º aminoácido ao 168º aminoácido, a região caracteriza-se como desordenada e citoplasmática;
Do 169º aminoácido ao 171º aminoácido, a região caracteriza-se como desordenada e ligante a proteína e ao citoplasmática.

PPP1R1B: Do 1º ao 10º aminoácido, a região caracteriza-se como enovelada;
Do 11º ao 12º aminoácido, a região caracteriza-se como extracelular;
Do 13º ao 21º aminoácido, a região caracteriza-se como enovelada;
Do 22º ao 31º aminoácido, a região caracteriza-se como helicoidal;
Do 32º ao 37º aminoácido, a região caracteriza-se como enovelada;
Do 38º ao 41º aminoácido, a região caracteriza-se como extracelular;
Do 42º ao 78º aminoácido, a região caracteriza-se como enovelada;

Do 79º ao 94º aminoácido, a região caracteriza-se como helicoidal;
 Do 95º ao 102º aminoácido, a região caracteriza-se como enovelada;
 Do 103º ao 113º aminoácido, a região caracteriza-se como helicoidal;
 Do 114º ao 121º aminoácido, a região caracteriza-se como enovelada;
 DO 122º ao 124º aminoácido, a região caracteriza-se como helicoidal;
 Do 125º ao 127º aminoácido, a região caracteriza-se como enovelada;
 Do 128º ao 146º aminoácido, a região caracteriza-se como helicoidal;
 Do 147º ao 170º aminoácido, a região caracteriza-se como enovelada;
 Do 171º ao 173º aminoácido, a região caracteriza-se como helicoidal;
 Do 174º ao 204º aminoácido, a região caracteriza-se como enovelada;
PPP1R1C: Do 1º ao 21º aminoácido, a região é caracterizada como enovelada.
 Do 22º ao 31º aminoácido, a região caracteriza-se como helicoidal,
 Do 32º ao 37º aminoácido, a região caracteriza-se como enovelada,
 DO 38º ao 40º aminoácido, a região caracteriza-se como extracelular,
 Do 41º ao 65º aminoácido, a região caracteriza-se como enovelada,
 Do 66º ao 68º aminoácido, a região caracteriza-se como helicoidal,
 Do 69º ao 78º aminoácido, a região caracteriza-se como enovelada,
 Do 79º ao 88º aminoácido, a região caracteriza-se como helicoidal,
 Do 89º ao 101º aminoácido, a região caracteriza-se como enovelada,
 Do 102º ao 106º aminoácido, a região caracteriza-se como helicoidal,
 Do 107º ao 109º aminoácido, a região caracteriza-se como enovelada.

Figura 8: Propriedades físico-químicas de PPP1R1A, PPP1R1B e PPP1R1C



Sequências de aminoácidos e suas propriedades. Fonte: PSIPRED

A caracterização individual das propriedades físico-químicas de cada aminoácido e a posição na sequência primária das três proteínas estudadas foram a seguinte:

Tabela 2: Características bioquímicas de PPP1R1A, PPP1R1B e PPP1R1C

PROTEÍNA	PPP1R1A	PPP1R1B	PPP1R1C	
Aminoácido pequeno e apolar	6,7,13,15,19,23,25	3,12,14,15,16,17	3,5,6,12,14,18,21	
	,26,33,34,35,36	,18,22,33,34,35	,22,24,25,33,34	
	,37,38,42,43,46	,36,42,45,46,47	,35,36,37,46,47	
	,47,48,56,58,62	,51,52,53,57,58	,54,55,57,59,64	
	,63,65,67,68,75	,59,61,66,69,71	,65,66,72,75,76	
	,78,79,80,93,98	,73,75,76,77,78	,77,78,81,87,91	
	,101,103,104,105	,81,86,88,92,94	,92,94,99,100	
	,107,108,113,115	,101,102,109,115		
	Continuação	,116,117,119,121	,117,137,139,148	
		,125,128,129,130	,149,150,153,154	
	,131,132,133,136	,158,161,162,166		
	,137,141,147,148	,167,168,172,176		
	,151,152,153,156	,177,178,185,186		
	,157,158,161,162	,188,190,191,194		
	,165,167,168,170.	,197,198,199,200		
		,202,203,204.		
Aminoácido hidrofóbico	1,10,14,16,17,21	1,9,13,20,25,27	1,9,13,15,20,28	
	,29,39,40,41,50	,28,37,38,41,64	,38,39,40,41,49	
	,55,60,64,66,74	,79,82,85,90,93	,61,73,79,82,85	
	,77,81,84,86,87	,96,108,111,114		
	,88,92,118,123	,141,142,144,145		
	,127,140,160,163,171	,159,169,182,187.		
Aminoácido polar	2,3,4,5,8,9,11,18	2,4,5,6,7,8,10,19	2,4,7,8,10,17,19	
	,20,22,24,27,28	,21,23,24,26,29	,23,26,27,29,30	
	,30,31,32,33,44	,30,31,32,40,43	,31,32,42,43,44	
	,45,49,51,52,53	,44,48,49,50,54	,45,48,50,51,52	
	,54,57,59,61,69	,55,56,60,62,63	,53,56,58,60,62	
	,70,71,72,73,76	,65,67,68,70,80	,63,67,68,69,70	
	,82,83,85,89,90	,83,84,87,89,91	,71,80,83,84,86	
	,91,94,95,96,97	,95,97,98,99,100	,88,89,90,95,96	
	,99,100,102,106	,103,104,105,106	,97,98,101,102	
	,109,111,112,114	,107,110,112,113	,103,104,105,106	
	,120,122,124,126	,118,119,120,121	,107,108,109.	
	,134,135,138,142	,122,123,124,125		
	,144,145,146,149	,126,127,128,129		
	,150,154,155,159	,130,131,132,133		
	,164,166,169.	,134,135,136,138		
		,140,143,146,147		
		,151,152,157,160		
		,164,165,170,171		
		,173,174,175,179		
	,180,181,183,184			
	,189,192,193,195			
	,196,201			
Aminoácido aromático e com cisteína	12 e 139	11,39,72,74,116,155,163	11,16,74,93	

Propriedades físico-químicas dos aminoácidos. Fonte: PSIPRED

5.3.3 PROSITE

A pesquisa por perfil e padrão proteico realizada no PROSITE trouxe as seguintes informações para cada uma das proteínas:

PPP1R1A:

Quadro 4: Padrão de MPTs de PPP1R1A

POSIÇÃO DO RESÍDUO	MPT	ENZIMA
Serina 6	Fosforilação	Proteína quinase C
Serina 67	Fosforilação	Proteína quinase C
Treonina 80	Fosforilação	Proteína quinase C
Treonina 132	Fosforilação	Proteína quinase C
Serina 152	Fosforilação	Proteína quinase C
Arginina 32	Fosforilação	Proteína quinase dependente de cAMP e cGMP
Lisina 75	Fosforilação	Proteína quinase dependente de cAMP e cGMP
Serina 46	Fosforilação	Proteína caseína quinase II
Treonina 80	Fosforilação	Proteína caseína quinase II
Treonina 121	Fosforilação	Proteína caseína quinase II
Serina 152	Fosforilação	Proteína caseína quinase II
Glicina 103	N- miristoilação	N-miristoiltransferase
Glicina 117	N-miristoilação	N-miristoiltransferase
Glicina 128	N-miristoilação	N-miristoiltransferase

Fonte: PROSITE.

Observa-se que para a PPP1R1A, o PROSITE identificou um padrão de 14 MPTs de possíveis ocorrências por quatro enzimas diversas.

PPP1R1B:

Figura 9: Perfil rico em glutamato da sequência de PPP1R1B



Fonte: PROSITE.

De acordo com o resultado obtido no PROSITE, PPP1R1B (DARPP-32) apresenta um perfil de uma região rica em glutamato (GLU), localizada na sequência proteica entre o aminoácido de posição 98 ao de posição 193.

Quadro 5: Padrão de MPTs de PPP1R1B

RESÍDUO	MPT	ENZIMA
Serina 18	Fosforilação	Proteína caseína quinase II
Serina 45	Fosforilação	Proteína caseína quinase II
Serina 46	Fosforilação	Proteína caseína quinase II

Serina 102	Fosforilação	Proteína caseína quinase II
Serina 137	Fosforilação	Proteína caseína quinase II
Serina 172	Fosforilação	Proteína caseína quinase II
Serina 198	Fosforilação	Proteína caseína quinase II
Arginina 31	Fosforilação	Proteína quinase dependente de cAMP e cGMP
Serina 66	Fosforilação	Proteína quinase C
Serina 78	Fosforilação	Proteína quinase C
Serina 172	Fosforilação	Proteína quinase C
Glicina 150	N-miristoilação	N-miristoiltransferase

Fonte: PROSITE

O resultado sobre o padrão proteico obtido no PROSITE para PPP1R1B identificou dois tipos possíveis de MPTs, a fosforilação e a N-miristoilação, com quatro padrões de ocorrência por enzimas diferentes.

PPP1R1C:

Quadro 6: Padrão de MPTs de PPP1R1C

RESÍDUO	MPT	ENZIMA
Serina 5	Fosforilação	Proteína quinase C
Serina 65	Fosforilação	Proteína quinase C
Treonina 78	Fosforilação	Proteína quinase C
Arginina 34	Fosforilação	Proteína quinase dependente de cAMP e cGMP
Arginina 69	Fosforilação	Proteína quinase dependente de cAMP e cGMP
Treonina 57	Fosforilação	Proteína caseína quinase II
Glicina 87	N-miristoilação	N-miristoiltransferase
Asparagina 89	N-glicosilação	Oligossacaril transferase

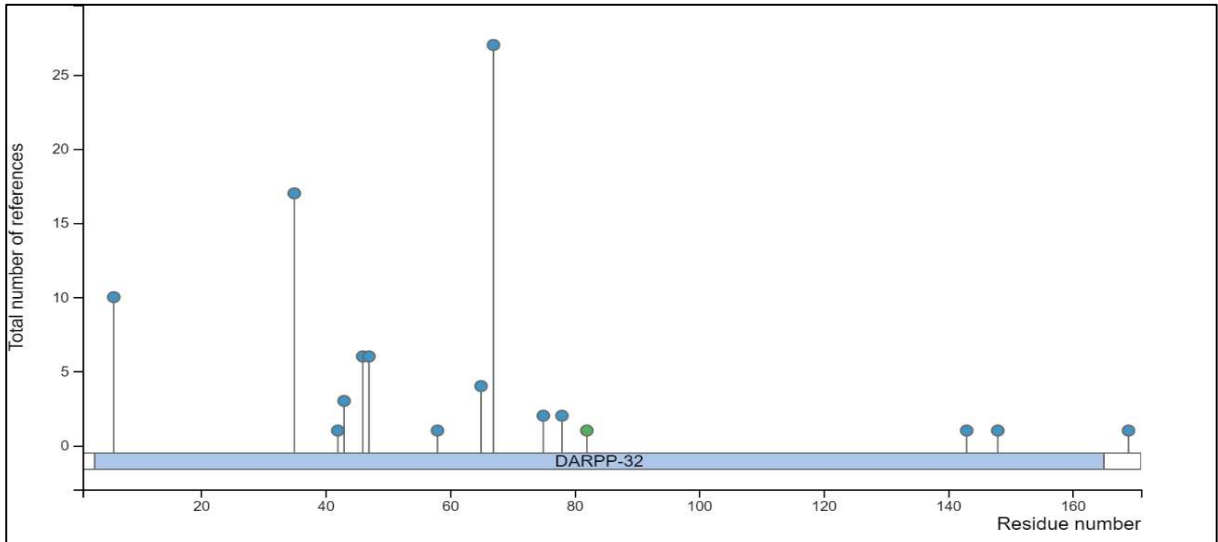
Fonte: PROSITE

O resultado sobre padrão proteico obtido no PROSITE para PPP1R1C encontrou três tipos possíveis de MPTs com um padrão de ocorrência por cinco enzimas diferentes.

5.3.4 PhosphoSitePlus

Observa-se que, a identificação dos possíveis resíduos fosforilados realizada através do PhosphoSitePlus apresentou o seguinte resultado:

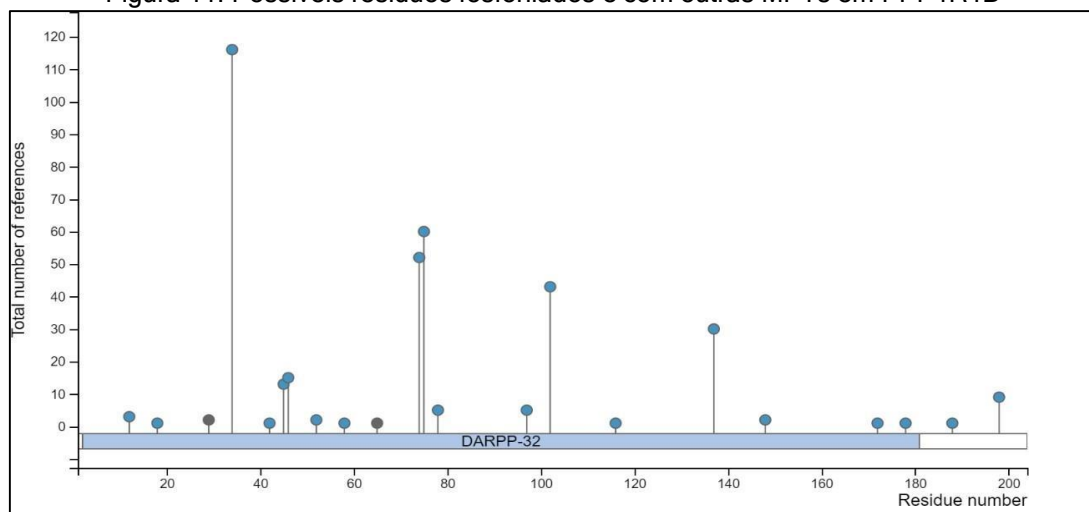
Figura 10: Possíveis resíduos fosforilados e com outras MPTs em PPP1R1A



Resíduos modificados por fosforilação (azul) e outras MPTs (verde). Fonte: PhosphoSitePlus

De acordo com o resultado obtido acima, foram identificados 14 possíveis resíduos fosforilados e um resíduo acetilado em PPP1R1A. Sendo que, as modificações pós-traducionais estão distribuídas na sequência proteica tendo como início o aminoácido 6, serina, e encerrando no aminoácido de posição 169, asparagina, e o resíduo que sofre acetilação é o aminoácido lisina que está na posição 82 da sequência polipeptídica.

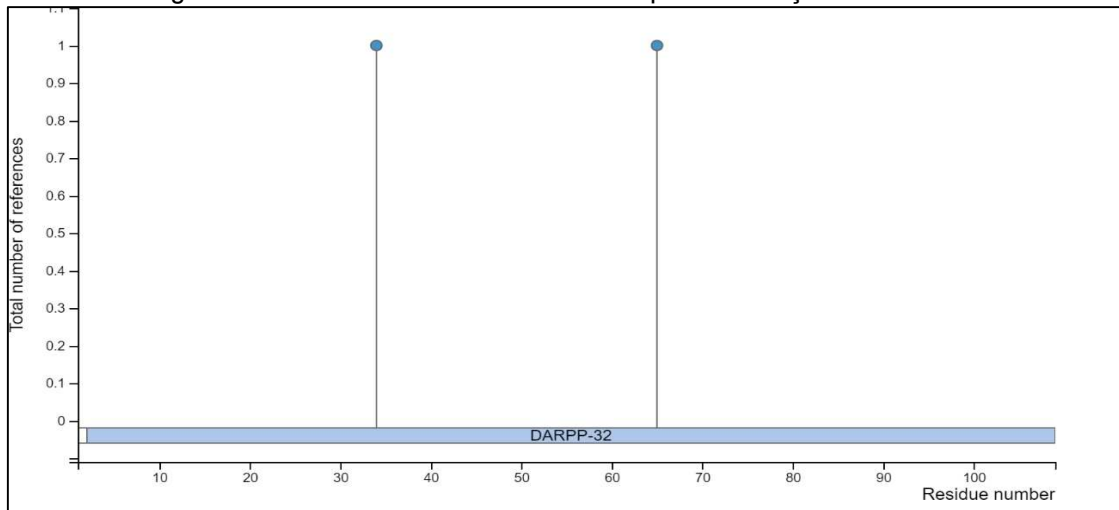
Figura 11: Possíveis resíduos fosforilados e com outras MPTs em PPP1R1B



Resíduos modificados por fosforilação (azul) e outras MPTs (marrom). Fonte: PhosphoSitePlus

De acordo com o resultado obtido, existem 20 possíveis resíduos de fosforilação em PPP1R1B, iniciando-se no aminoácido de posição 12, serina, e terminando no aminoácido de posição 198, serina, podendo ocorrer metilação no resíduo de arginina 29 e no resíduo de lisina na posição 65 da sequência.

Figura 12: Possíveis resíduos modificados por fosforilação em PP1R1C



Resíduos modificados por fosforilação (azul). Fonte: PhosphoSitePlus

De acordo com o resultado obtido no PhosphoSitePlus, os resíduos possíveis de fosforilação em PPP1R1C são os aminoácidos treonina e serina localizados, respectivamente, nas posições 34 e 65 da sequência desta proteína.

5.3.5 Signal P-6.0

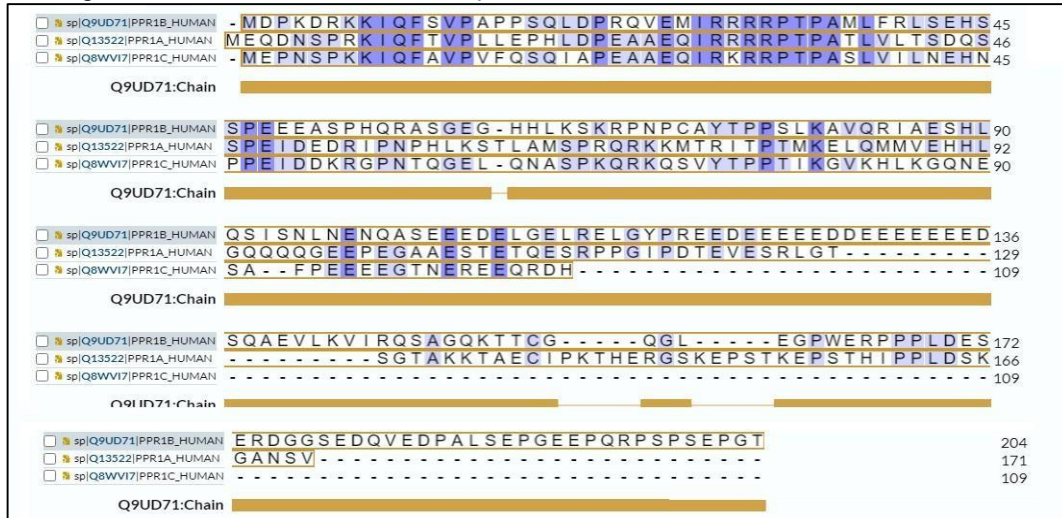
A busca para identificação de peptídeo sinal em PPP1R1A, PPP1R1B e PPP1R1C realizada no Signal P 6.0 revelou que nenhuma das três proteínas possui este tipo de peptídeo na sua estrutura.

5.4 Alinhamento das sequências

Dentre os resultados do alinhamento das sequências de aminoácidos de PPP1R1A, PPP1R1B e PPP1R1C foram selecionados a overview (visão geral) do alinhamento das três sequências e o alinhamento dessas com destaque para os resíduos de serina e treonina.

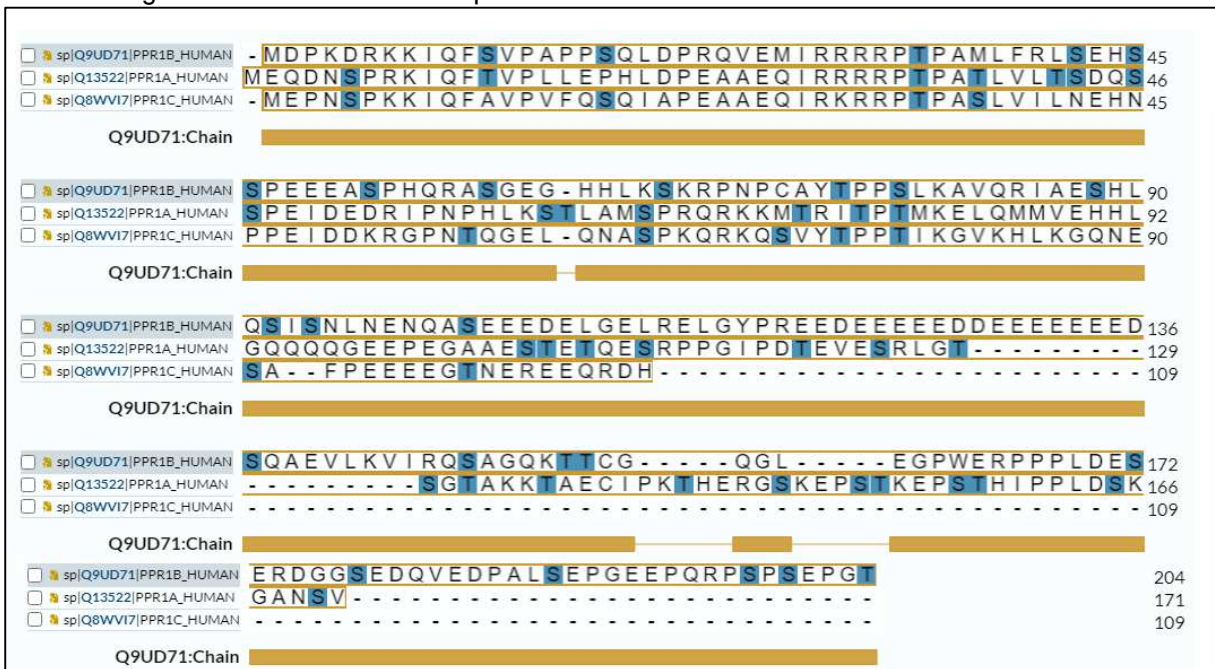
O alinhamento realizado conta com 206 sítios, incluindo sítios modificados, variáveis e *gaps* (lacunas). Considerados como sítios conservados são aqueles em que não há modificação do aminoácido nas três sequências. O alinhamento apresentou 24 sítios conservados, enquanto os sítios variáveis, que são aqueles que tiveram variação ao menos entre duas sequências, excluindo as *gaps*, contou com 126 sítios variáveis.

Figura 13: Alinhamento entre as sequências de PPP1R1A, PPP1R1B e PPP1R1C



Foram identificados 24 sítios conservados entre as sequências. Fonte: Clustal Omega no UniProt
 O alinhamento das três proteínas destacando os resíduos de serina e treonina revelou apenas um sítio de treonina conservado entre as mesmas, como indicado na figura 14.

Figura 14: Alinhamento de sequências destacando os resíduos de serina e treonina



Fonte: Clustal Omega no UniProt

A matriz de identidade resultante do alinhamento de sequências e gerada no UniProt apresenta os valores indicados na tabela abaixo:

Tabela 3: Matriz de identidade entre PPP1R1A, PPP1R1B e PPP1R1C

	PPP1R1A	PPP1R1B	PPP1R1C
PPP1R1A	100%	28.4%	38.53%
PPP1R1B	28.4%	100%	37.61%
PPP1R1C	38.53%	37.61%	100%

Valores baseados no alinhamento de sequências. Fonte: UniProt

Portanto, de acordo com as informações da matriz de identidade, há uma maior identidade entre PPP1RA e PPP1R1C do que entre essas e PPP1R1B.

6 DISCUSSÃO

Este estudo, apesar de intitular-se como um estudo *in silico* de DARPP-32, também, envolveu estudos referentes as duas proteínas parálogas de DARPP-32, PPP1R1A e PPP1R1C. Sendo assim, talvez o título não corresponda exatamente ao trabalho que foi desenvolvido, mas ao foco principal. A inclusão dessas proteínas justifica-se por haver material bibliográfico a respeito de ambas, ainda que pequeno quando comparado a DARPP-32, e o estudo sobre elas em bancos de dados e ferramentas de bioinformática nota-se ter sido possível gerar resultados importantes.

O presente trabalho corrobora a categorização de DARPP-32 (PPP1R1B) e as homólogas PPP1R1A e PPP1R1C como proteínas integrantes da família IPP-1, caracterizadas pelo efeito inibitório de PP1. Tais informações foram coletadas em artigos retirados do PubMed (YGER & GIRAULT, 2011), (DANCHEK, *et al.*, 2008) e (UNG & TEOH, 2014).

Os resultados obtidos no InterPro que agregam além dos específicos resultados desse banco de dados, os resultados do PANTHER, MObiDB-lite e Pfam que confirmam a inclusão destas proteínas como formadoras da família IPP-1, visto que, foi encontrado um identificador próprio do InterPro (IPR008466) e o domínio proteico característico desta família (PF05395).

A busca de potenciais homólogos através do BLAST implementado no UniProt, também, confirmou que a DARPP-32 tem possíveis homólogos em algumas espécies de animais filogeneticamente próximos, como os primatas da família Hominidae. Tais espécies incluem: *Pan troglodytes*, *Gorilla gorila*, *Pan paniscus* e *Pongo abelii*.

Considerando que a seleção de material restringiu a busca a ordem Primata é possível que ampliando ao grupo taxonômico na busca, encontrem um número ainda maior de potenciais homólogos, visto que a DARPP-32, provavelmente, originou-se nos ancestrais dos Gnathostomata e irradiou-se por todos os grandes grupos de Vertebratas (Yger & Girault, 2011).

O estudo sobre a arquitetura proteica feito no UniProt e InterPro identificou a presença de um domínio característico em DARPP-32, PPP1R1A e PPP1R1C, que engloba uma região de aproximadamente 90 a 150 aminoácidos em conformidade com a proteína em questão, apresentando regiões desordenadas como descrito nos trabalhos de Dancheck e colaboradores (2008), Yger e Girault (2011) e Girault e Nairn (2022).

A identificação de grande variedade de possíveis resíduos fosforilados, embora haja divergência na relação ao número predito de resíduos fosforilados, entre o PhosphoSitePlus e PROSITE, demonstra a importância destas proteínas em vários tipos de reações bioquímicas em que a fosforilação tem participação, como a inibição de atividade de PP1.

Neste sentido, a divergência pode estar relacionada ao fato de que o PhosphoSitePlus utiliza como fonte literatura científica relacionada a MPTs, especialmente, a fosforilação, e o PROSITE, que utiliza como fonte algoritmos de busca que encontrarão as mais prováveis combinações de motivos, sítios e domínios em um proteoma, sendo que, a chance de falsos positivos neste último é grande (STROH, *et al.*, 2002).

O grande número de resíduos possivelmente fosforilados, demonstra a importância desta Modificação Pós-Traducional na manutenção da vida, refletida nos vários

processos biológicos sob influência de tais reações, como o crescimento, a migração e a divisão celular (REBELO, *et al.*, 2015).

Apesar do descobrimento da DARPP-32 ter ocorrido há mais de 40 anos (GIRAULT & NAIRN, 2021), este estudo, notabiliza-se por caracterizar DARPP-32 e as irmãs parálogas como PIDs, fato esse, que é sustentado em trabalhos recentes (AVANES, *et al.*, 2019) e (GIRAULT & NAIRN, 2021).

Sendo assim, as estruturas tridimensionais estão ausentes ou possuem baixa confiabilidade em bancos de dados estruturais sobre proteínas como o Protein Data Bank (PDB), (VELANKAR, *et al.*, 2021) e AlphaFold (VARADI, *et al.*, 2021).

O enquadramento de DARPP-32 como PID ganha reforço com a predição da região rica em glutamato identificada no PROSITE, confirmando o fato descrito na literatura (AVANES, *et al.*, 2019). A caracterização de tais proteínas como PIDs e a identificação das regiões de desordem podem ser úteis em pesquisas sobre a relação de tais proteínas com a origem de doenças como cânceres e transtornos mentais, podendo auxiliar na compreensão de como as PIDs, a despeito da plasticidade conformacional que possuem, podem influir em mecanismos bioquímicos e fisiológicos tão complexos e dinâmicos como a plasticidade sináptica, como se constata no caso de DARPP-32, inclusive sendo possível que esta instabilidade seja a razão da influência em tais mecanismos.

Uma constatação interessante deste estudo *in silico* é que o perfil rico em glutamato, considerado como um fator na caracterização de uma PID, na pesquisa feita através do PROSITE, somente é identificado em DARPP-32. Sendo que, esta ferramenta não identifica o mesmo perfil quando a pesquisa é feita com as sequências de PPP1R1A ou PPP1R1C. Além disso, uma pesquisa bibliográfica no PubMed utilizando como palavras chave: profile, rich, glutamic acid, PPP1R1A e PPP1R1C não retorna nenhum resultado que relacione a presença dessa região nas duas proteínas.

Então, pode-se levantar algumas hipóteses: 1) A presença de glutamato não tem influência no fato de que PPP1R1A e PPP1R1C sejam PIDs. 2) Pode ser que, não exista ou haja poucos trabalhos relacionados a região rica em glutamato com essas

proteínas e o fato de serem PIDs. 3) Talvez, devido à pouca atenção na pesquisa científica sobre essas duas proteínas, este perfil, rico em glutamato possa existir, porém tenha sido ignorado nos trabalhos já realizados.

De qualquer forma, a ausência de resultados sobre a região rica em glutamato em PPP1R1A e PPP1R1C é digna de nota, pois tratam-se de proteínas com uma história evolutiva “próxima” e a elevada concentração deste aminoácido, juntamente com outros aminoácidos como, a lisina, a arginina e em algumas proteínas, a prolina, é uma característica marcante das PIDs (NELSON & COX, 2019). Tomadas todas essas informações em conjunto, elas tornam-se relevantes nas áreas de genética molecular e bioquímica evolutiva.

A predição feita no PSIPRED das propriedades bioquímicas da estrutura secundária de DARPP-32, PPP1R1A e PPP1R1C apresenta inconsistências como a identificação de peptídeo sinal e domínio extracelular e transmembrana, os quais não são registrados na literatura científica sobre estas proteínas.

Tais inconsistências podem ocorrer em função de erros na avaliação individual de cada aminoácido na estrutura secundária. Apesar de tais problemas, o PSIPRED na predição individual de cada aminoácido da estrutura primária e de cada uma das proteínas da família IPP-1 confirmou o alto número de aminoácidos polares e, possíveis formadores de carga em relação ao número de aminoácidos hidrofóbicos. Essa relação tem importância no fato de tais proteínas comportarem-se como proteínas intrinsecamente desordenadas (HABCHI, *et al.*, 2014).

Nota-se, interessante, que a identificação de regiões desordenadas, apesar da variação na terminologia empregada na classificação de tais regiões entre MObiDB-lite e PSIPRED e de acordo com os conceitos físico-químicos utilizados têm igual finalidade. Em outras palavras, regiões de polianfólitos, polieletrólitos ou regiões hidrofóbicas, polares e apolares, são classificações baseadas em propriedades físico-químicas das sequências e dos aminoácidos que as constituem.

A maior similaridade encontrada entre PPP1R1A e PPP1R1C pode ser um indicativo, que essas proteínas tiveram uma divergência em um tempo mais recente em comparação a DARPP-32 (PPP1R1B).

O alinhamento de sequências confirmou uma alta conservação de resíduos entre a posição de número 8 a posição de número 38 da sequência das três proteínas. A similaridade tem especial destaque para o motivo KIQF, que tem papel secundário no mecanismo de inibição de PP-1 através da fosforilação no resíduo Thr34 em DARPP-32, como descrito nos trabalhos de Greengard e colaboradores (1999), Yger e Girault (2011) e Girault e Nairn (2022).

Espera-se, também, que este estudo fomente trabalhos sobre a filogenética destas proteínas e a identificação de possíveis alvos moleculares contra as enfermidades nas quais elas têm participação. Lembrando que, um único trabalho *in silico* sobre DARPP-32 é conhecido (UNG & TEOH, 2014).

O presente estudo *in silico* diferencia-se de tais trabalhos por trazer novas informações, em especial, sobre as características bioquímicas e físico-químicas das sequências das proteínas, particularmente no que se refere as regiões preditas como intrinsecamente desordenadas.

Este trabalho foi desenvolvido, mas pode ser complementado, pois apresenta lacunas que trabalhos vindouros possam vir a preencher, como informações referentes as vias metabólicas em que há participação de DARPP-32 e as homólogas dela, e a reconstrução filogenética desta família proteica. O acréscimo de tais informações será valioso para a compreensão da manutenção ao longo do tempo evolutivo de tais proteínas, que demonstram compor uma família bastante flexível quanto a estrutura bioquímica e interações moleculares.

7 CONCLUSÕES E PERSPECTIVAS

O presente trabalho confirma que DARPP-32 e as parálogas PPP1R1A e PPP1R1C são proteínas contendo grande número de resíduos potencialmente fosforilados e que podem sofrer outras MPTs. Esta constatação junto a identificação de uma região desordenada no centro da estrutura destas biomoléculas, corrobora a hipótese de que essas proteínas, especialmente a DARPP-32, tem um papel-chave na mediação de cascatas bioquímicas complexas. Estas reações bioquímicas influenciam fenômenos fisiológicos como a plasticidade sináptica, que é uma propriedade dos tecidos cerebrais em que DARPP-32 e as parálogas se fazem presente em maior grau.

A caracterização bioquímica da estrutura predominantemente desordenada destas proteínas e a identificação de resíduos passíveis de fosforilação e outras MPTs podem vir, em um futuro próximo, a subsidiar pesquisas que tenham como alvo regiões intramoleculares de tais proteínas, por meio de terapias inovadoras como a nanofarmacologia e a terapia gênica.

Embora não tenha sido feita a reconstrução filogenética da família IPP-1, espera-se que pesquisas em filogenética possam vir a utilizar deste trabalho para elucidar a história evolutiva dessa família de proteínas, dos mecanismos biológicos, moleculares

e ambientais que garantiram a manutenção da espécie humana e a importância que tais proteínas demonstram ter na atividade fisiológica, em especial, no que refere as funções do Sistema Nervoso Central.

Pelo entendimento desses mecanismos, grandes avanços podem ocorrer na pesquisa, desenvolvimento e inovação de alternativas terapêuticas para um grande número de enfermidades em que DARPP-32, PPP1R1A e PPP1R1C demonstram ter certa participação, e espera-se que a compreensão sobre o grupo distinto das PIDs, nas quais a DARPP-32 faz parte, auxilie a um melhor entendimento em termos de história evolutiva, propriedades bioquímicas e fenômenos biológicos associados.

REFERÊNCIAS

ANÁLISE de proteínas e proteômica. ICB - Instituto de Ciências Biomédicas - USP. Aula 10 (slides), s/d. Disponível em:

<http://www.icb.usp.br/~biocomp/SLIDES/aula_10.pdf>. Acesso em: 02 mar. 2023.

AVANES, A; Lenz G; Momand J. **Darpp-32 and t-Darpp protein products of PPP1R1B**: Old dogs with new tricks. *Biochem Pharmacol.* 2019. Feb; 160:71-79.

BELKHIRI, A; Zhu S; El-Rifai W. **DARPP-32**: from neurotransmission to cancer. *Oncotarget.* 2016. Apr 5-7 (14).

BORATYN GM; Camacho C; Cooper PS; Coulouris G; Fong A; Ma N; Madden TL; Matten WT; McGinnis SD; Merezhuk Y; Raytselis Y; Sayers EW; Tao T; Ye J; Zaretskaya I. **BLAST**: a more efficient report with usability improvements. *Nucleic Acids Res.* 2013 Jul. 41(Web Server issue): W29-33.

BUCHAN DWA; Jones DT. **Protein Analysis Workbench**: 20 years on. *Nucleic Acids Res.* 2019. Jul 2; 47(W1): W402-W407.

CHEN C; Huang H; Wu CH. **Protein Bioinformatics Databases and Resources**. *Methods Mol Biol.* 2017; 1558:3-39.

MOREIRA, Leonardo Márcio. **CIÊNCIAS genômicas**: Fundamentos e aplicações. Ribeirão Preto: Sociedade Brasileira de Genética, 2015. 403 p.

DI DOMENICO T; Walsh I; Martin AJ; Tosatto SC. **MobiDB**: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics.* 2012 Aug, 1. 28 (15):2080-1.

ESPINOSA Mendoza P; García-González V; Moreno A; Castillo R; Mas-Oliva J. **Disorder to order conformational transitions in protein structure and its relationship to disease**. *Mol Cell Biochem.* 2009 Oct; 330 (1-2): 105-20.

GARG VK, Avashthi H; Tiwari A; Jain PA; Ramkete PW; Kayastha AM; Singh VK; MFPP1. **Bioinformation.** 2016 Apr 10;12 (2):74-77.

GENE e Proteína: PPP1R1A. <https://www.ncbi.nlm.nih.gov/gene/5502>. Acesso: 23 de julho de 2023.

GENE e Proteína: PPP1R1B. <https://www.ncbi.nlm.nih.gov/gene/84152>. Acesso: 23 de julho de 2023.

GENE e Proteína: PPP1R1C. <https://www.ncbi.nlm.nih.gov/gene/151242>. Acesso: 23 de julho de 2023.

GIRAULT JA, Nairn AC. **DARPP-32**: 40 years later. *Adv Pharmacol.* 2021; 90:67-87.

GREENGARD P; Allen PB; Nairn AC. **Beyond the dopamine receptor**: the DARPP-32/protein phosphatase-1 cascade. *Neuron.* 1999. Jul; 23 (3):435-47.

HABCHI J; Tompa P; Longhi S, Uversky VN. **Introducing protein intrinsic disorder**. Chem Rev. 2014 Jul 9; 114 (13):6561-88.

HEMMINGS HC JR; Nairn AC; Aswad DW; Greengard P. **DARPP-32**: A dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus. J Neurosci. 1984 Jan; 4(1):99-110.

HEMMINGS HC Jr; Greengard P; Tung HY; Cohen P. **DARPP-32**: a dopamine-regulated neuronal phosphoprotein is a potent inhibitor of protein phosphatase-1. Nature. 1984 Aug 9-15;310(5977):503-5.

HORNBECK PV; Kornhauser JM; Latham V; Murray B; Nandhikonda V; Nord A; Skrzypek E; Wheeler T; Zhang B; Gnad F. **15 years of PhosphoSitePlus®**: integrating post-translationally modified sites, disease variants and isoforms. Nucleic Acids Res. 2019. Jan 8; 47(D1): D433-D441.

HORNBECK PV; Zhang B; Murray B; Kornhauser JM; Latham V; Skrzypek E; PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015 Jan; 43 (Database issue): D512-20.

LIU L; Zhang X; Nan C; Zhao Z; Ma S; Li W; Hu H; Liang Z. **MicroRNA-182 targets protein phosphatase 1 regulatory inhibitor subunit 1C in glioblastoma**. Oncotarget. 2017 Sep 27; 8 (70):114677-114684.

LUO W; Xu C; Phillips S; Gardenswartz A; Rosenblum JM; Ayello J; Lessnick SL; Hao HX; Cairo MS. **Protein phosphatase 1 regulatory subunit 1A regulates cell cycle progression in Ewing sarcoma**. Oncotarget. 2020 May 12; 11 (19):1691-1704.

MATIOLI, Sergio & Fernandes F M C. **Biologia molecular e evolução**. Segunda edição. Ribeirão Preto: Holos Editora,256 p. 2012.

Maurer-Stroh S; Eisenhaber B; Eisenhaber F. **N-terminal N-myristoylation of proteins**: refinement of the sequence motif and its taxon-specific differences. J Mol Biol. 2002 Apr 5; 317(4): 523-40.

MCLAREN L; Boyle S; Mason JO; Bard JB. **Expression and genomic characterization of protein phosphatase inhibitor-1**: a novel marker for mesothelium in the mouse. Mech Dev. 2000 Sep; 96(2):237-41.

MISTRY J; Chuguransky S; Williams L; Qureshi M; Salazar GA; Sonnhammer ELL; Tosatto SCE; Paladin L; Raj S; Richardson LJ; Finn RD; Bateman A. **Pfam**: The protein families database in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D412-D419.

NECCI M; Piovesan D; Clementel D; Dosztányi Z; Tosatto SCE. **MobiDB-lite 3.0**: fast consensus annotation of intrinsic disorder flavors in proteins. Bioinformatics. 2021 Apr 1;36 (22-23): 5533-5534.

NELSON, David L.; COX, Michael M. **Princípios de bioquímica de Lehninger**. 7ª Edição. Porto Alegre: Artmed, 2019, 1278 p.

PAYSAN-LAFOSSE T; Blum M; Chuguransky S; Grego T; Pinto BL; Salazar GA; Bileschi ML; Bork P; Bridge A; Colwell L; Gough J; Haft DH; Letunić I; Marchler-Bauer A; Mi H; Natale DA; Orengo CA; Pandurangan AP; Rivoire C; Sigrist CJA; Sillitoe I; Thanki N; Thomas PD; Tosatto SCE; Wu CH; Bateman A. **InterPro in 2022**. *Nucleic Acids Res.* 2023 Jan 6; 51(D1): D418-D427.

PIERCE, Benjamin A. **Genética**: um enfoque conceitual. 5ª Edição. Rio De Janeiro: Editora Guanabara Koogan S.A., 2017.

Piersimoni L; Abd El Malek M; Bhatia T; Bender J; Brankatschk C; Calvo Sánchez J; Dayhoff GW; Di Ianni A; Figueroa Parra JO; Garcia-Martinez D; Hesselbarth J; Köppen J; Lauth LM; Lippik L; Machner L; Sachan S; Schmidt L; Selle R; Skolidis I; Sorokin O; Ubbiali D; Voigt B; Wedler A; Wei AAJ; Zorn P; Dunker AK; Köhn M; Sinz A; Uversky VN. **Lighting up Nobel Prize-winning studies with protein intrinsic disorder**. *Cell Mol Life Sci.* 2022 Jul 26;79(8):449.

REBELO S; Santos M; Martins F; da Cruz e Silva EF; Cruz e Silva AO da. **Protein phosphatase 1 is a key player in nuclear events**. *Cell Signal.* 2015 Dec; 27 (12): 2589-98.

RIDLEY, M. **Evolução**. 3ª Edição. Porto Alegre: Artmed,2006,752p.

ROSA, D.V.F. **Expressão de DARPP-32 e NCS-1 em cérebros de ratos submetidos à estimulação eletroconvulsiva**. Dissertação (Mestrado em Farmacologia Bioquímica e Molecular) - Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais. Belo Horizonte, p.51. 2005.

SAYERS EW; Beck J; Bolton EE; Bourexis D; Brister JR; Canese K; Comeau DC; Funk K; Kim S; Klimke W; Marchler-Bauer A; Landrum M; Lathrop S; Lu Z; Madden TL; O'Leary N; Phan L; Rangwala SH; Schneider VA; Skripchenko Y; Wang J, Ye J; Trawick BW; Pruitt KD; Sherry ST. **Database resources of the National Center for Biotechnology Information**. *Nucleic Acids Res.* 2021. Jan 8; 49(D1): D10-D17.

SCHEGGI S; De Montis MG; Gambarana C. **DARPP-32 in the orchestration of responses to positive natural stimuli**. *J Neurochem.* 2018 Nov; 147(4): 439-453.

SIEVERS F; Higgins DG. **The Clustal Omega Multiple Alignment Package**: *Methods Mol Biol.* 2021; 2231:3-16.

SIEVERS F; Wilm A; Dineen D; Gibson TJ; Karplus K; Li W; Lopez R; McWilliam H; Remmert M; Söding J; Thompson JD; Higgins DG. **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega**: *Mol Syst Biol.* 2011 Oct 11; 7: 539.

SIGRIST CJ; de Castro E; Cerutti L; Cucho BA; Hulo N; Bridge A; Bougueleret L; Xenarios I. **New and continuing developments at PROSITE**. *Nucleic Acids Res.* 2013 Jan; 41(Database issue): D344-7.

TANG J; Kong D; Cui Q; Wang K; Zhang D; Yuan Q; Liao X; Gong Y; Wu G. **Bioinformatic analysis and identification of potential prognostic microRNAs and mRNAs in thyroid cancer**. 2018 May 4;6: e4674.

TEUFEL F; Almagro Armenteros JJ; Johansen AR; Gíslason MH; Pihl SI; Tsirigos KD; Winther O; Brunak S; Von Heijne G; Nielsen H. **SignalP 6.0 predicts all five types of signal peptides using protein language models**. Nat Biotechnol. 2022. Jul; 40(7): 1023-1025.

THOMAS PD; Ebert D; Muruganujan A; Mushayahama T; Albou LP; Mi H. **PANTHER: Making genome-scale phylogenetics accessible to all**. Protein Sci. 2022 Jan; 31 (1):8-22.

TIWARI A; Sekhar AK. **Workflow based framework for life science informatics: Comput Biol Chem**. 2007 Oct; 31(5-6): 305-19.

UNG CY; Teoh TC. **The evolutionary strata of DARPP-32 tail implicates hierarchical functional expansion in higher vertebrates**. 2014. Jun; 39(3): 493-504.

UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan 6; 51(D1):D523-D531.

Varadi M; Anyango S; Deshpande M; Nair S; Natassia C; Yordanova G; Yuan D; Stroe O; Wood G; Laydon A; Židek A; Green T; Tunyasuvunakool K; Petersen S; Jumper J; Clancy E; Green R; Vora A; Lutfi M; Figurnov M; Cowie A; Hobbs N; Kohli P; Kleywegt G; Birney E; Hassabis D; Velankar S; Alpha Fold. **Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models**. Nucleic Acids Res. 2022 Jan 7; 50 (D1): D439-D444.

VELANKAR S; Burley SK; Kurisu G; Hoch JC; Markley JL. **The Protein Data Bank Archive: Methods Mol Biol**. 2021; 2305:3-21.

WALAAS SI; Greengard P; DARPP-32. **A dopamine and adenosine 3': 5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions**. I. Regional and cellular distribution in the rat brain. J Neurosci. 1984 Jan;4 (1):84-98.

WILKINS MR; Gasteiger E; Bairoch A; Sanchez JC; Williams KL; Appel RD; Hochstrasser DF. **Protein identification and analysis tools in the ExPASy server**. Methods Mol Biol. 1999. 112: 531-52.

WU X; Wang Y; Yang M; Wang Y; Wang X; Zhang L; Liao L; Li N; Mao M; Guan J; Ye F. **Exploring prognostic value and regulation network of PPP1R1A in hepatocellular carcinoma**. Hum Cell. 2022. Nov; 35 (6):1856-1868.

YGER M; Girault JA. **DARPP-32: Jack of All Trades... Master of Which?** Front Behav Neurosci. 2011 Sep 8; 5:56.