

**Universidade Federal do Rio de Janeiro**  
**Centro de Ciências Matemáticas e da Natureza - CCMN**  
**Instituto de Química - IQ**  
**Departamento de Bioquímica**

**INVESTIGAÇÃO DA ESTABILIDADE ESTRUTURAL E DA ESPECIFICIDADE DE  
LIGAÇÃO DOS DOMÍNIOS PWWP DAS PROTEÍNAS NSD3s E Pdp3**

**Daniel da Silva Costa**

**Rio de Janeiro**  
**2021**

**INVESTIGAÇÃO DA ESTABILIDADE ESTRUTURAL E DA ESPECIFICIDADE DE  
LIGAÇÃO DOS DOMÍNIOS PWWP DAS PROTEÍNAS NSD3s E Pdp3**

**Daniel da Silva Costa**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Bioquímica, do Instituto de Química, da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de mestre em bioquímica.

Orientador: Anderson de Sá Pinheiro

**Rio de Janeiro  
Outubro de 2021**

## CIP - Catalogação na Publicação

CC837i Costa, Daniel da Silva  
Investigação da estabilidade estrutural e da especificidade de ligação dos domínios pwwp das proteínas NSD3s e Pdp3 / Daniel da Silva Costa. -- Rio de Janeiro, 2021.  
57 f.

Orientador: Anderson de Sá Pinheiro.  
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Química, Programa de Pós Graduação em Bioquímica, 2021.

1. Domínios PWWPs. 2. Modelagem molecular. 3. Interação proteína-histonas. I. Pinheiro, Anderson de Sá, orient. II. Título.

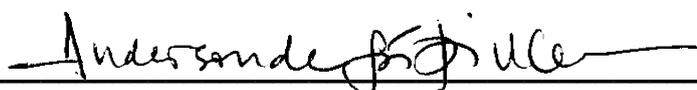
**INVESTIGAÇÃO DA ESTABILIDADE ESTRUTURAL E DA ESPECIFICIDADE DE  
LIGAÇÃO DOS DOMÍNIOS PWWP DAS PROTEÍNAS NSD3s E Pdp3**

**Daniel da Silva Costa**

**Orientador: Anderson de Sá Pinheiro**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Bioquímica, do Instituto de Química, da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de mestre em bioquímica.

**Aprovada por:**



---

**(Anderson de Sá Pinheiro, doutor, UFRJ)**

---

**(Marcos Dias Pereira, doutor, UFRJ)**

---

**(Carolina Alvares da Cunha de Azeredo Braga, doutora, UFRJ)**

**Rio de Janeiro**

**Outubro 2021**

## **AGRADECIMENTOS**

Agradeço a Deus por estar no centro e guiar os propósitos da minha vida, também agradeço aos meus pais Sergio e Aparecida por ser a base de todo amor e carinho e, ainda, o suporte que foram para que eu alcançasse meus objetivos.

Agradeço a doutora Vanessa Joia pelo suporte psicológico prestado durante a escrita de uma dissertação de mestrado em meio a pandemia, sem sua ajuda nenhuma palavra seria escrita.

Agradeço ao meu orientador Anderson de Sá Pinheiro, por acreditar em um aluno de graduação de outra faculdade que pediu uma vaga de iniciação científica no seu laboratório e até aqui foi muito trabalho, muitas conquistas e muitas comemorações. Obrigado por toda orientação, aprendizado e amizade.

Agradeço aos meus colegas de laboratório pela união, companheirismo e amizade, foram essenciais para a conclusão dessa etapa, aprendi um pouco com cada um de vocês.

Agradeço aos professores Bianca Neves, Fabio Almeida, Iraima Cordeiro e Rafael Mesquita, por cederem espaço nos seus laboratórios para eu poder utilizar os equipamentos e softwares necessários para concluir esse trabalho.

## RESUMO

### Daniel da Silva Costa. **Investigação da Estabilidade Estrutural e da Especificidade de Ligação dos Domínios PWWP das Proteínas NSD3s e Pdp3.**

Dissertação (mestrado em bioquímica) – Centro de Ciências Matemáticas e da Natureza, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

NSD3s (*Nuclear Receptor Binding SET-domain containing protein 3*) é a isoforma curta de uma histona metil-transferase humana frequentemente amplificada em 15% dos casos de câncer de mama. A NSD3s possui apenas o domínio PWWP amino-terminal e é suficiente para a transformação de células saudáveis em células tumorais. Nosso grupo de pesquisa mostrou que a superexpressão da NSD3s em células de *Saccharomyces cerevisiae*, bem como a de Pdp3, uma proteína da levedura que contém um domínio PWWP similar ao de NSD3s, promove a aquisição de fenótipos tumorais, como o aumento da fermentação aeróbica. Ainda, nosso grupo mostrou que a expressão de ambas as proteínas regula vias metabólicas similares, porém não idênticas, sugerindo que, apesar da similaridade de sequência entre seus domínios PWWP, as duas proteínas recrutam diferentes complexos de remodelamento da cromatina. Neste sentido, o presente trabalho tem como objetivo comparar estruturalmente os domínios PWWP de NSD3s e Pdp3, tanto no que tange à estabilidade termodinâmica quanto à especificidade de interação com ligantes fisiológicos, como peptídeos derivados de histonas e sequências de DNA.

Os domínios PWWP de NSD3s e Pdp3 recombinantes foram expressos e purificados. As afinidades de interação entre PWWP-NSD3s e PWWP-Pdp3 e nove sequências peptídicas de 13 resíduos de aminoácidos derivados das caudas amino-terminais das histonas H3 e H4 contendo diferentes resíduos de lisinas metiladas e diferentes graus de metilação foram estimadas por espectroscopia de fluorescência. A estabilidade termodinâmica de PWWP-NSD3s e PWWP-Pdp3 foi investigada através de fluorimetria diferencial de varredura (DSF) e dicroísmo circular (CD). As análises de interação entre proteína-DNA foram realizadas através de EMSA (Electrophoretic Mobility Shift Assay). A estrutura tridimensional de PWWP-Pdp3 foi determinada por modelagem comparativa.

Os espectros de fluorescência intrínseca de PWWP-NSD3s e PWWP-Pdp3 apresentaram máximo de emissão em ~340 nm, sugerindo que os resíduos de triptofano são parcialmente expostos ao solvente. A adição de concentrações crescentes de peptídeos promoveu a supressão da fluorescência de modo dose-dependente. As isotermas de ligação derivadas dos dados de fluorescência exibiram duas inflexões distintas, logo os dados foram fitados por uma equação que determina duas constantes de dissociação distintas. A interação de PWWP-NSD3s e PWWP - Pdp3 com os diferentes peptídeos se deu com afinidades semelhantes (~3-10 micromolar), sugerindo que estes domínios são capazes de reconhecer metil-lisinas com alta afinidade, porém com pouca seletividade. PWWP-NSD3s ( $T_m$  ~60°C) se mostrou mais estável que PWWP-Pdp3 ( $T_m$  ~40°C) frente a desnaturação térmica. A diferença de estabilidade parece estar relacionada a ausência de um resíduo de triptofano no motivo Pro-Trp-Trp-Pro de PWWP-NSD3s. A diferença de estabilidade entre os dois domínios foi confirmada por DSF. Ainda, o complexo PWWP-NSD3s:H3K36me3 apresentou  $T_m$ ~73°C, sugerindo que a interação com o peptídeo

estabiliza o domínio PWWP da NSD3s, o mesmo não acontece com o complexo PWWP-Pdp3:H3K36me3. Os ensaios de EMSA revelaram que ambos os domínios interagem com sequências de DNA não específicas na escala de micromolar, porém com diferentes afinidades, sugerindo que o domínio PWWP-NSD3s é mais afim que o PWWP-Pdp3.

## ABSTRACT

Daniel da Silva Costa. **Investigation of Structural Stability and Binding Specificity of the PWWP Domains of NSD3s and Pdp3 Proteins.** Dissertation (master's degree in biochemistry) - Center of Mathematical and Natural Sciences, Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, 2021.

NSD3s (*Nuclear Receptor Binding SET-domain containing protein 3*) it is the short isoform of a human histone methyl transferase frequently amplified in 15% of breast cancer cases. NSD3s has only the amino-terminal PWWP domain and is sufficient for the transformation of healthy cells into tumor cells. Our research group showed that overexpression of NSD3s in *Saccharomyces cerevisiae* cells, as well as that of Pdp3, a yeast protein that contains a PWWP domain similar to that of NSD3s, promotes the acquisition of tumor phenotypes, such as increased aerobic fermentation. In addition, our group showed that the expression of both proteins regulates similar, but not identical, metabolic pathways, suggesting that, despite the sequence similarity between their PWWP domains, the two proteins recruit different chromatin remodeling complexes. The present work aims to structurally compare the PWWP domains of NSD3s and Pdp3, both with regard to thermodynamic stability and the specificity of interaction with physiological ligands, such as peptides derived from histones and DNA sequences.

The recombinant NSD3s and Pdp3 PWWP domains were expressed and purified. The interaction affinities between PWWP-NSD3s and PWWP-Pdp3 and nine peptide sequences of 13 amino acid residues derived from the amino terminal tails of histones H3 and H4 containing different residues of methylated lysines and different degrees of methylation were fluorescence spectroscopy. The thermodynamic stability of PWWP-NSD3s and PWWP-Pdp3 was investigated through differential scanning fluorimetry (DSF) and circular dichroism (CD). Protein-DNA interaction analyzes were performed using EMSA (Electrophoretic Mobility Shift Assay). The three-dimensional structure of PWWP-Pdp3 was provided by comparative modeling.

The intrinsic fluorescence spectra of PWWP-NSD3s and PWWP-Pdp3 showed maximum emission at ~ 340 nm, suggesting that the tryptophan residues are partially exposed to the solvent. The addition of increasing concentrations of peptides promoted suppression of fluorescence in a dose-dependent manner. The binding isotherms derived from the fluorescence data exhibited two distinct inflections, so the data was

viewed by an equation that determines two different dissociation constants. The interaction of PWWP-NSD3s and PWWP-Pdp3 with the different peptides occurred with similar affinities (~ 3-10 micromolar), suggesting that these domains are capable of recognizing methyl lysines with high affinity, but with little selectivity. PWWP-NSD3s ( $T_m \sim 60^\circ\text{C}$ ) proved to be more stable than PWWP-Pdp3 ( $T_m \sim 40^\circ\text{C}$ ) against thermal denaturation. The difference in stability seems to be related to the absence of a tryptophan residue in the PWWP-NSD3s Pro-Trp-Trp-Pro motif. The difference in stability between the two domains was confirmed by DSF. Furthermore, the PWWP-NSD3s: H3K36me3 complex showed  $T_m \sim 73^\circ\text{C}$ , suggesting that the interaction with the peptide stabilizes the PWWP domain of NSD3s, the same does not happen with the PWWP-Pdp3: H3K36me3 complex. The EMSA assays revealed that both domains interact with non-specific DNA sequences on the micromolar scale, but with different affinities, suggesting that the domain PWWP-NSD3s is more similar than the PWWP-Pdp3.

## Sumário

1	INTRODUÇÃO .....	11
1.1	Cromatina.....	11
1.2	Modificações pós-traducionais .....	14
1.3	Nuclear receptor SET Domain-containing Protein (NSD3) .....	15
1.4	PWWP Domain-containing Protein in NUA3B (Pdp3).....	17
1.5	Domínios tipo PWWP .....	18
2	OBJETIVOS.....	20
2.1	Objetivo Geral .....	20
2.2	Objetivos Específicos.....	20
3	METODOLOGIA.....	20
3.1	Modelagem Molecular .....	20
3.2	Meio de Cultivo .....	22
3.3	Cepas Bacterianas .....	22
3.4	Construção dos Plasmídeos.....	22
3.5	Transformação .....	23
3.6	Expressão dos domínios .....	23
3.7	Purificações dos domínios.....	24
3.8	Eletroforese em gel de poliacrilamida sob condições desnaturantes (SDS PAGE) .....	27
3.9	Dosagem de proteínas .....	27
3.10	Differential Scanning Fluorimetry (DSF).....	28
3.11	Dicroísmo circular (CD).....	28
3.12	Peptídeos derivados das Histonas.....	29
3.13	Espectroscopia de Fluorescência .....	30
3.14	Oligonucleotídeos de DNA .....	31
3.15	EMSA .....	32
4	RESULTADOS.....	33
4.1	Modelagem Molecular .....	33
4.2	Expressão e Purificação dos Domínios.....	40
4.3	Purificações do domínio PWWP-NSD3s .....	43
4.4	Differential Scanning Fluorimetry (DSF).....	46
4.5	Dicroísmo Circular .....	48
4.6	Espectroscopia de Fluorescência .....	50
4.7	Electrophoretic Mobility Shift Assay (EMSA) .....	54
5	DISCUSSÃO .....	59
6	CONCLUSÃO .....	61

# INVESTIGAÇÃO DA ESTABILIDADE ESTRUTURAL E DA ESPECIFICIDADE DE LIGAÇÃO DOS DOMÍNIOS PWWP DAS PROTEÍNAS NSD3s E Pdp3

## 1 INTRODUÇÃO

### 1.1 Cromatina

As células eucarióticas possuem em seu núcleo um intrincado polímero de material genético denominado cromatina. A cromatina é constituída por um complexo de DNA genômico e proteínas dispostos em uma fibra com diferentes níveis de enovelamento. Esta é uma estrutura dinâmica que altera entre um estado condensando e compacto (heterocromatina) e um estado mais desenovelado e expandido (eucromatina) (figura 1) (SHAHID *et al.*, 2020).

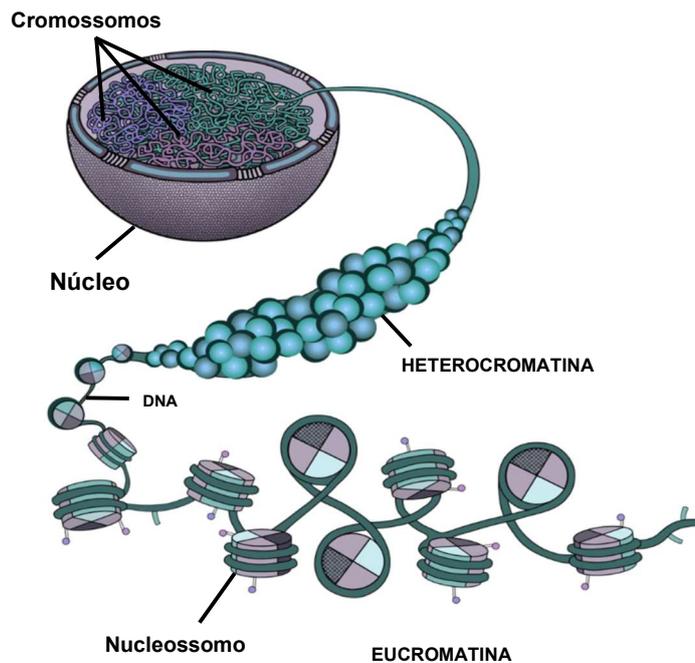


Figura 1: Organização da cromatina em diferentes estágios de enovelamento, evidenciando a diferença entre a forma ativa e inativa do DNA (SHAW, 2013).

Os diferentes graus de enovelamento iniciam-se por meio da interação eletrostática entre o DNA e um complexo proteico positivamente carregado que é circulado por duas voltas de DNA e assim, forma-se o nucleossomo, sendo este a unidade base da organização do material genômico dentro das células eucarióticas. Os nucleossomos podem estar aglutinados formando uma fibra de 30 nanômetros ou mais expostos, sendo a regulação entre esses dois estados gerida por três processos:

metilação ou demetilação do DNA, atuação de enzimas dependentes de ATP ou modificações pós traducionais em histonas (PMTs) (MARKERT & LUGER, 2020).

A dinâmica de conversão entre os diferentes graus de empacotamento da cromatina se relaciona com a regulação de processos biológicos, como: replicação, transcrição, recombinação e reparo do DNA (CRUZ *et al.*, 2015). Consequentemente, estudar o mecanismo que a célula utiliza para promover alterações nos níveis de empacotamento do DNA é de suma importância para o entendimento de doenças que têm como uma das causas a ativação ou inativação de determinados genes.

Ainda que não completamente elucidada, a complexidade estrutural da cromatina tem sido extensivamente estudada. Pesquisas acerca da organização espacial dentro do núcleo identificaram a existência de complexos proteicos capazes de modificar e alternar as formas enovelada e desenovelada da cromatina. Ainda, estas alterações estão relacionadas com processos como a regulação da expressão gênica (CORBETT, 2018).

O nucleossomo, atribuído como a unidade fundamental da cromatina, é composto por 147 pares de bases de DNA envoltos em um octâmero de histonas. As histonas (H3, H4, H2A, H2B E H1), por sua vez, consistem em pequenas proteínas básicas, compostas por um domínio globular e uma cauda N-terminal (figura 2). Enquanto o domínio globular apresenta-se ligado ao filamento de DNA, a região N-terminal é flexível e carregada.

As histonas H3 e H4 se associam em uma estrutura tetramérica, enquanto as histonas H2A e H2B formam dois dímeros. Juntas, as quatro proteínas compõem o octâmero central do nucleossomo (BANNISTER; KOUZARIDES, 2011; IIZUKA; SMITH, 2003). Paralelamente, a histona H1 é responsável por estabilizar o nucleossomo, ligando-se ao DNA que não interage diretamente com o octâmero de histonas (CRUZ *et al.*, 2015; TSAO *et al.*, 2009).

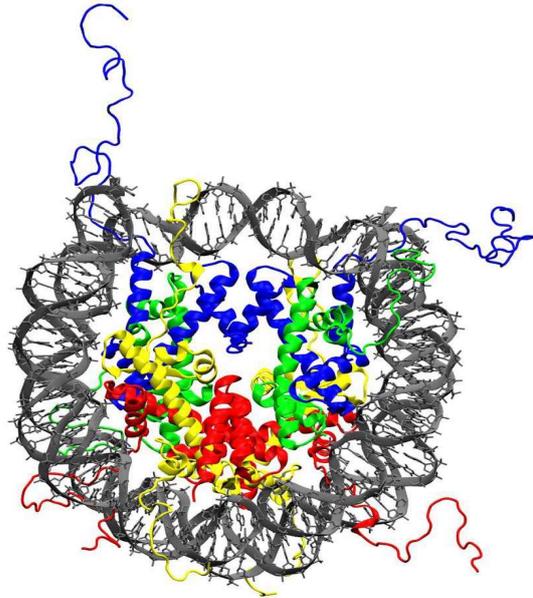


Figura 2: Estrutura cristalina do nucleossomo. Na região superior, o tetrâmero formado pelas histonas H3 (azul) e H4 (verde); na região inferior, os dímeros de histonas H2A (amarelo) e H2B (vermelho) (PDB: 1EQZ).

As histonas H2A, H2B, H3 e H4 apresentam três  $\alpha$ -hélices em seu domínio globular, além de duas alças que atuam na organização da partícula central do nucleossomo. No caso das histonas H3 e H4 observa-se um maior grau de conservação na sequência primária, o que leva a estruturas e funções similares entre diferentes organismos. O oposto ocorre para a histona H1, que apresenta uma maior heterogeneidade - somente em humanos já foram identificadas 11 proteínas diferentes com um domínio globular central e caudas N e C terminais ricas em lisinas (IZZO & SCHNEIDER, 2015).

Os domínios globulares que formam o nucleossomo possuem uma estrutura rígida como consequência da formação de motivos “*hand-shake*” relacionados à dimerização de monômeros através do entrelaçamento das hélices das histonas e contatos eletrostáticos com o DNA, enquanto a região N-terminal das histonas apresenta-se de forma desordenada, projetando-se para fora do nucleossomo (PLACEK, GLOSS, 2002; SUEOKA *et al.*, 2018). Desse modo, as histonas são capazes de atuar diretamente na manutenção contínua da estrutura da cromatina através da possibilidade de diferentes modificações químicas na cauda N-terminal ou, menos frequentes, na cauda C-terminal ou no domínio globular (MARKERT & LUGER, 2020).

## 1.2 Modificações pós-traducionais

Estudos de proteômica quantitativa conduzidos até o momento foram capazes de identificar 22 tipos de modificações pós-traducionais (PTMs - *post-translational modifications*), tais como: fosforilação (Ph), acetilação (Ac), metilação (Me), hidroxilação (OH), succinilação (Succ), butirilação (Bu), crotonilação (Cr), propionilação (Pr), 2 hidroxilisobutirilação (Hib), citrulinação (Cit), glutarilação (Gl), formilação (Fo), ubiquitinação (Ub), malonilação (Ma), sumoilação (Sumo), O-Glcnaçiação (Og), ADP-ribosilação (Ar) e citrulinação (Cit) em 8 resíduos de aminoácidos com 138 posições dentro das 5 histonas, totalizando 550 possibilidades de PTMs (VILLASEÑOR *et al.*, 2020). Ademais, foram identificadas diversas proteínas que atuam na incorporação, reconhecimento e remoção de PTMs em histonas tais proteínas são conhecidas como “escritores”, “apagadores” ou “leitores” da informação presente no padrão de PTMs em histonas (BROWNELL *et al.*, 1996).

As PTMs de histonas são capazes de influenciar múltiplos processos celulares na medida em que podem atuar individualmente ou em combinação com outras PTMs (STRAHL; ALLIS, 2000). Estas diferentes formas de atividade podem acarretar em mudanças estruturais nas próprias histonas, tendo em vista que a função dos cromossomos pode ser afetada pela alteração das cargas eletrostáticas das histonas (IIZUKA; SMITH, 2003). Assim, através das PTMs, domínios da cromatina podem ser expostos, estabilizados, omitidos ou inclusive rompidos através da associação com proteínas regulatórias, que podem ser fatores de transcrição, proteínas relacionadas a condensação e reparo do DNA. Dessa maneira, as PTMs em histonas podem regular a principal forma de controle transcricional epigenético (BANNISTER, KOUZARIDES, 2011; IIZUKA, SMITH, 2003; SANTOS-ROSA, CALDAS, 2005).

No entanto, as PTMs não estão restritas às histonas. Estudos anteriores identificaram sítios específicos de metilação no DNA, como dinucleotídeos citosina-guanina, onde o resíduo de guanina se encontra fosforilado (ilhas de CpG). Neste caso, a metilação inibe a transcrição de genes alvo e favorece o recrutamento de proteínas que auxiliam a condensação da cromatina (RAZIN; RIGGS, 1980).

Dentre as PTMs mais frequentes, destacam-se a acetilação, metilação e fosforilação. A metilação de histonas ocorre em resíduos de lisina e arginina por meio da ação de histonas metil-transferases (BANNISTER; KOUZARIDES, 2011). As PTMs são capazes de provocar alterações no estado de ativação da transcrição gênica e,

para certas PTMs, esse efeito é dependente da posição na qual este evento ocorre. Por exemplo, enquanto a acetilação está associada somente a genes transcricionalmente ativos, a metilação pode estar associada tanto a genes transcricionalmente ativos quanto a genes silenciosos (IIZUKA, SMITH, 2003; MELLOR, 2006). O efeito da metilação sob o estado transcricional de determinado gene é dependente da posição dessa marca epigenética no nucleossomo. Portanto, a metilação das lisinas K4, K36 e K79 na histona H3 está relacionada à ativação da transcrição, ao passo que a metilação das lisinas K9 e K27 na histona H3 associa-se ao silenciamento da transcrição (MELLOR, 2006).

Enquanto a acetilação ocorre somente em resíduos de lisinas, a metilação pode ocorrer tanto em lisinas quanto em argininas. O grupo  $\epsilon$ -amino da lisina pode comportar um máximo de três grupos metila, sendo que cada estado de metilação exerce funções distintas (HAN *et al.*, 2019). O grau de metilação será dependente de um balanço entre a atividade das enzimas responsáveis pela adição do grupo metila (lisinas metil-transferases - KMTs; em inglês: *lysine methyl-transferases*) e daquelas que atuam na remoção deste grupo (demetilases específicas para lisinas - KDMs; em inglês: *lysine demethylases*).

### 1.3 Nuclear receptor SET Domain-containing Protein (NSD3)

Dentre as KMTs, destaca-se uma família de proteínas chamada NSD (*Nuclear Receptor SET Domain-containing protein*), sendo composta pelos três seguintes membros: D1/KMT3B, NSD2/WHSC1/MMSET e NSD3/WHSC1L1. Tais proteínas apresentam grande importância na manutenção da estrutura da cromatina através da metilação da lisina K36 na histona H3 (H3K36me), sendo este um marcador de ativação transcricional conhecido (WAGNER; CARPENTER, 2012).

Em termos estruturais, esta família de proteínas apresenta uma estrutura composta por domínios funcionalmente distintos e comuns às três proteínas, sendo estes: um domínio SET (*Supressor of variegation 3-9, Enhancer of zeste and Thritorax*), dois domínios PWWP (Pro-Trp-Trp-Pro), e de quatro a cinco domínios dedos de zinco PHD (*Plant Homeo Domain*). Enquanto o domínio SET é responsável pela atividade catalítica das NSDs, os domínios PWWP e dedos de zinco PHD estão relacionados ao reconhecimento da cromatina e interações proteína-proteína (VOUGIOUKLAKIS *et al.*, 2015).

As três proteínas desta família apresentam alta taxa de semelhança em certa região, que compreende cerca de 700 aminoácidos, na qual estão inseridos o domínio catalítico SET, um domínio PWWP cinco domínios PHD e um domínio rico em Cys-His específico de NSD. Apesar disso, no caso do gene NDS3, presente na região cromossômica 8p11-12 e também conhecido por WHSC1L1 (*Wolf-Hirschhorn Syndrome Candidate 1-Like 1*), o quinto homeodomínio (PHD5) e o domínio rico em Cys-His adjacente apresentam uma certa preferência por peptídeos da histona H3, como K4 não-metilado e K9 tri-metilado (ANGRAND *et al.*, 2001; HAN *et al.*, 2018). Desta maneira ele difere-se, dos demais membros desta família, na medida em que este mesmo módulo em NSD2 possui alta preferência por H3K9 não metilada (H3K9me0), enquanto que, no caso da proteína NSD1 não há indícios de interação com peptídios derivados de histona H3 (HE *et al.*, 2013).

Estudos de genômica e transcriptômica permitiram observar uma correlação entre o aumento na quantidade de cópias gênicas e a supressão de NSD3 em diferentes tipos de câncer, em sua maioria no câncer de mama. A região cromossômica onde está localizado o gene NSD3 apresenta uma amplificação na faixa de 15-18% em casos de câncer de mama e assim, vem sendo associado à baixa taxa de sucesso nos tratamentos desta variedade da doença (GELSI-BOYER *et al.*, 2005; HAVERTY *et al.*, 2008). Ainda neste contexto, a superexpressão de NSD3s já foi vista como capaz de promover a sobrevivência de células epiteliais de mama e induzir à formação de fenótipos tumorais. Assim, o *knockdown* desta proteína causa redução do crescimento, bem como diminuição de linhagens tumorais, o que reafirma o potencial oncogênico desta enzima no câncer de mama (YANG *et al.*, 2010; KANG *et al.*, 2013). A proteína NSD3 também é aplicada na translocação cromossomal t(8;11) (p11.2;p15) que resulta na proteína de fusão NSD3-NUP98, estando relacionada à leucemia mieloide aguda e à síndrome mielodisplástica (ROSATI *et al.*, 2002; TAKETANI *et al.*, 2009).

Esta terceira proteína da família NSD apresenta três isoformas em consequência ao splicing alternativo do éxon 10, sendo elas: NSD3L, NSD3s e WHISTLE. Enquanto a isoforma mais alongada, NSD3L, codifica a proteína íntegra que comporta 1.437 aminoácidos, ela também contém dois domínios PWWP, além de cinco PHDs, um domínio rico em cisteína associado ao SET (SAC) e, finalmente, um domínio SET (HAN *et al.*, 2018). Por outro lado, a isoforma curta, sendo a NSD3s, codifica uma proteína de apenas 645 aminoácidos não apresentando um domínio

catalítico SET, porém um domínio PWWP N-terminal (HAN *et al.*, 2018). A terceira e última isoforma de NSD3, WHISTLE, apresenta 506 aminoácidos C-terminais, além do domínio SET e o segundo domínio PWWP (VOUGIOUKLAKIS *et al.*, 2015).

Em determinadas linhagens de câncer de mama, a isoforma curta apresenta níveis de expressão mais elevados quando comparada à NSD3L. Ainda, a superexpressão desta isoforma basta para a conversão de células epiteliais de mama MCF10A em células tumorais, o que reforça a relevância do domínio PWWP N-terminal para a atividade oncogênica da NSD3 (YANG *et al.*, 2010).

#### 1.4 PWWP Domain-containing Protein in NUA3B (Pdp3)

A *PWWP Domain-containing Protein in NUA3* (Pdp3) é uma proteína de 35 kDa de *Saccharomyces cerevisiae*, sendo majoritariamente encontrada no núcleo e, até o presente momento, apresenta como único domínio descrito para essa proteína o domínio PWWP N-terminal.

Estudos anteriores do grupo mostraram que a NSD3s é capaz de modificar o perfil metabólico da *saccharomyces cerevisiae* de oxidativo para fermentativo e, no mesmo trabalho, identificou-se que o domínio PWWP-Pdp3 possui 25% de semelhança na sequência primária do domínio PWWP-NSD3s, sendo então a Pdp3 um possível ortólogo da NSD3s na levedura (RONA *et al.*, 2016).

Em outro estudo foi observado através da metabolômica que as duas proteínas regulam vias metabólicas diferentes, sugerindo que, apesar de semelhantes, os domínios PWWP das duas proteínas agem por mecanismos diferentes (RONA *et al.*, 2018), com isso surgem questionamentos se esses domínios conservam semelhanças estruturas e afinidades de interação com ligantes característicos dos domínios PWWP.

Gilbert e colaboradores copurificaram Pdp3 em conjunto com outras proteínas do complexo NUA3 e, neste mesmo trabalho, caracterizaram o complexo em duas formas, sendo uma delas inédita e com a participação da Pdp3. O complexo NUA3a interage com a MPT H3K4me3 através do domínio PHD da proteína Yng1, já o complexo NUA3b reconhece a MPT H3K36me3 através do domínio PWWP presente na Pdp3 (figura 3).

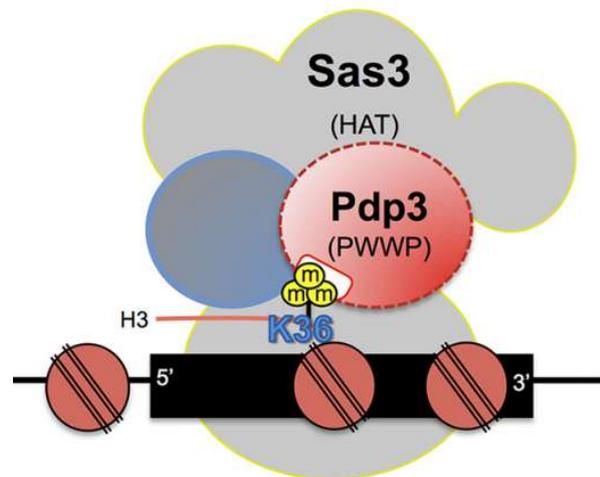


Figura 3: O complexo NUA3b reconhece a trimetilação da lisina 36 da histona H3 através do domínio PWWP-Pdp3 e recruta demais proteínas, entre elas a enzima Sas3, capaz de promover modificações pós traducionais de acetilação em sítios ainda não identificados.

A deleção da Pdp3 resultou em modificações no crescimento celular e mutações no alongamento da transcrição da levedura (GILBERT *et al.*, 2014). Além disso, a Pdp3 transloca do núcleo para o citoplasma em resposta a hipóxia (DASTIDAR *et al.*, 2012). Modificações fenotípicas em situações de hipóxia, também são respostas em células tumorais (VLEUGEL *et al.*, 2005). Logo, as proteínas NSD3s e Pdp3 podem apresentar semelhanças ainda que estejam presentes em organismos diferentes.

### 1.5 Domínios tipo PWWP

Exclusivo de eucariotos, o domínio PWWP está presente em diversas proteínas encontradas no núcleo de células humanas atuando ativamente na divisão, crescimento e diferenciação celular, além de estar relacionado a diferentes funções da cromatina, como modificação e reparo do DNA, bem como a regulação da transcrição. O domínio PWWP apresenta aproximadamente 100-150 resíduos de aminoácidos, além de conter o motivo conservado Pro-Trp-Trp-Pro (RONA *et al.*, 2016). Sua estrutura apresenta-se como um barril  $\beta$  e um feixe helicoidal que possui entre uma e seis  $\alpha$ -hélices, sendo capaz de se ligar tanto a sequências de DNA como a resíduos de lisinas metiladas em histonas (WEAVER *et al.*, 2018). Desta maneira,

tal domínio atua como um motivo de reconhecimento para metil-lisinas no nucleossomo, estando assim intimamente ligado à regulação epigenética (STEC *et al.*, 2000; WU *et al.*, 2011).

Diversos estudos anteriores já foram capazes de elucidar diferentes estruturas tridimensionais de domínios PWWP, bem como seus complexos, o que revelou a alta taxa de conservação no modo de interação entre estes domínios e o DNA (RONA *et al.*, 2016). Vale ressaltar que todos os domínios PWWP investigados até o presente momento apresentaram ligação às caudas de histonas metiladas, com exceção do domínio presente na proteína HDGF (WEAVER *et al.*, 2018).

A ligação entre o domínio PWWP e o DNA se dá através de uma superfície básica, na qual resíduos de arginina e lisina estão presentes adjacentes à região de interação com metil-lisinas, além de incorporar a região do arco  $\beta 1$ -  $\beta 2$  (onde o motivo PWWP está inserido) e parcelas da região helicoidal imediatamente posterior a fita- $\beta$  terminal (WEAVER *et al.*, 2018). Nota-se que para a maioria dos domínios PWWP a ligação ao DNA acontece de forma inespecífica por meio de interações eletrostáticas com o grupo fosfato do esqueleto principal do DNA (RONA *et al.*, 2016), enquanto que no âmbito da interação com metil-lisinas, o domínio PWWP interage por meio de uma cavidade hidrofóbica formada por três resíduos aromáticos dispostos no arco  $\beta 1$ - $\beta 2$ , no motivo Pro-Trp-Trp-Pro e na fita  $\beta 3$ . As cadeias laterais destes aminoácidos estão dispostas perpendicularmente umas às outras, compondo uma gaiola aromática capaz de acomodar o grupo metil (VEZZOLI *et al.*, 2010; WU *et al.*, 2011; QIU *et al.*, 2012; EIDAHL *et al.*, 2013; VAN, NULAND *et al.*, 2013; WEN *et al.*, 2014; WANG *et al.*, 2014).

Tendo em vista que os sítios de ligação a DNA e a histona não se sobrepõem amplamente, sugere-se que estes domínios se liguem simultaneamente ao DNA e às histonas em associação multivalente com o nucleossomo. Corroborando com esta predição, Weaver e colaboradores determinaram que o domínio PWWP de Pdp1 liga-se simultaneamente a H4K20me3 e ao DNA genômico sem afetar a afinidade de ligação ao DNA devido à interação com H4K20me3, o que fundamenta eventos de ligação independentes (WEAVER *et al.*, 2018).

Posteriormente foi mostrado que para o domínio PWWP da proteína DMNT3A as interações do domínio com as modificações de di/trimetilação na lisina 36 da histona H3 e as interações do domínio com o DNA somavam forças para manter a

estabilidade do complexo formado pela proteína DNMT3A e o nucleossomo (DUKATZ *et al.*, 2019).

Tendo em vista a semelhança funcional sugerida para as proteínas NSD3s e Pdp3, ambas contendo apenas os domínios PWWP como domínio enovelado, o presente trabalho tem a finalidade de comparar estruturalmente tais domínios e avaliar (*in vitro*) suas interações com peptídeos derivados das caudas das histonas e oligonucleotídeos de DNA.

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

Este trabalho teve como objetivo comparar a estabilidade termodinâmica e a especificidade de interação com histonas e com DNA dos domínios PWWP das proteínas NSD3s e Pdp3.

### **2.2 Objetivos Específicos**

- Realizar a modelagem comparativa do domínio PWWP-Pdp3
- Expressar e purificar os domínios PWWP-NSD3s e PWWP-Pdp3
- Determinar a estabilidade termodinâmica de PWWP-NSD3s e PWWP-Pdp3
- Determinar a especificidade, afinidade e estequiometria de ligação entre esses domínios e peptídeos derivados das histonas
- Determinar a afinidade e especificidade de ligação dos domínios a oligonucleotídeos de DNA

## **3 METODOLOGIA**

### **3.1 Modelagem Molecular**

Na tentativa de determinar a estrutura do domínio PWWP da proteína Pdp3 o domínio apresentou solubilidade e estabilidade termodinâmica inferiores aos padrões necessários para determinação estrutural por ressonância magnética nuclear (RMN),

com isso optou-se por produzir um modelo estrutural através da modelagem comparativa.

Inicialmente pesquisou-se na plataforma *National Center for Biotechnology Information* (NCBI) a sequência fasta da proteína Ppd3 íntegra e manteve-se somente os resíduos de aminoácidos referentes ao domínio PWWP-Pdp3.

Em seguida, utilizou-se a plataforma *Basic Local Alignment Search Tool* (BLAST) para identificar outros domínios PWWP com maiores percentuais na semelhança da sequência primária ao PWWP-Pdp3, que possuam estrutura determinada e depositada no PDB. Encontrou-se os domínios PWWP-HDGF (PDB: 2B8A), PWWP-p75/LEDGF (PDB: 2M16) e PWWP-PSIP1 (PDB: 4FU6).

Através do programa PyMOL foram observados os resíduos de aminoácidos das sequências fasta dos moldes que não possuem estrutura 3D resolvida. Tais resíduos de aminoácidos foram deletados nas sequências, afim de trabalhar-se somente com os resíduos de aminoácidos dos moldes que possuem suas estruturas determinadas.

As pontes de sulfeto são ligações de suma importância para a modelagem molecular, devido ao potencial dessas ligações de restringir espacialmente as possíveis conformações adotadas pelas proteínas. Nesse sentido, avaliou-se os moldes utilizados quanto a possíveis pontes de sulfeto e, através da plataforma pyMol, observou-se que apenas o molde PWWP-HDGF (pdb:2B8A) apresentava duas cisteínas, no entanto espacialmente distantes entre si. Logo, pode-se concluir que não ocorre a formação de pontes de sulfeto nos moldes.

Utilizou-se o software J pred para a predição de estrutura secundária do domínio PWWP-Pdp3, considerando-se como elementos de estrutura secundária os resíduos de aminoácidos que coincidiram entre os diferentes algoritmos (jnetpred, JNETHMM e JNETPSSM) presentes na plataforma em questão.

O alinhamento múltiplo foi realizado através da plataforma Praline web, utilizando como *input* as sequências em formato fasta do domínio PWWP-Pdp3 e dos moldes PWWP-HDGF, PWWP-p75/LEDGF e PWWP-PSIP1.

O arquivo formato fasta obtido como *output* do alinhamento múltiplo foi transformado em formato pir, com a finalidade de ser usado como *input* da modelagem comparativa junto com os arquivos PDB dos *templates*, através do *software* moddler.

Como output do moddler obteve-se 400 arquivos em formato PDB de possíveis modelos para o domínio PWWP-Pdp3, em seguida, os modelos foram analisados

quanto a menor energia através do algoritmo DOPE (*Discrete Optimized Protein Energy*) fornecido pelo próprio moddler e sendo uma ferramenta extrínseca aos cálculos da modelagem. Essa ferramenta baseia-se na restrição espacial dos átomos do modelo em comparação com proteínas resolvidas estruturalmente, já o DOPE normalizado ocorre quando os valores do DOPE são escalonados quanto ao número de restrições em cada átomo. Os 3 modelos com menor DOPE e menor DOPE normalizado foram selecionados para validação através da plataforma SAVES 5.0.

### 3.2 Meio de Cultivo

O meio Luria-Bertani (LB) (Sigma) [triptona 10 g/L, extrato de levedura 5 g/L e NaCl 5 g/L] foi utilizado para o crescimento das cepas bacterianas, enquanto o meio LB sólido (Sigma) [meio líquido acrescido de ágar 15 g/L] foi utilizado para plaqueamento das células transformadas. O meio LB líquido foi utilizado em uma concentração final de 20 g/L e, para o meio sólido, 35 g/L. No plaqueamento e no cultivo celular os meios foram acrescidos do antibiótico canamicina na concentração final de 100 µg/mL, previamente preparados em soluções de estoque na concentração de 100 mg/mL com a utilização de água ultrapura Milli-Q (Millipore) e esterilizados por filtração em membrana de poro 0,22 µm

### 3.3 Cepas Bacterianas

Foram utilizadas duas cepas de *Escherichia coli*, sendo uma para armazenamento e propagação dos plasmídeos e outra para expressão das proteínas de interesse. A cepa *E. coli* DH5-α [genótipo: F-φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1] foi utilizada para estoque de plasmídeos, enquanto a cepa *E. coli* BL21(DE3) [genótipo: F-ompT hsdSB(rB-,mB-) gal dcm (DE3)] foi utilizada para a expressão das proteínas recombinantes.

### 3.4 Construção dos Plasmídeos

O gene que codifica o domínio PWWP da proteína NSD3s inserido no vetor pETM30-GST-Nhe1 e o gene do domínio PWWP da proteína Pdp3 (nativa e com a

mutação P20W) inserido no vetor RP1B foram comprados da empresa Genscript. Ambas as construções apresentam uma cauda de seis histidinas e um sítio de clivagem para a protease *Tobacco Etch Virus* (TEV) importante para a estratégia de purificação dos domínios. Além disso, o plasmídeo que codifica o domínio PWWP-NSD3s apresenta uma cauda GST fusionada a cauda de histidinas para aumentar a solubilidade do domínio.

### 3.5 Transformação

Uma alíquota de 2  $\mu$ L com aproximadamente 100 ng de plasmídeos foram inseridos em soluções com 200  $\mu$ L de células de *E. coli* BL21(DE3) quimicamente competentes e preparadas segundo a metodologia Hanahan *et al.* (1983). Essas soluções foram submetidas a um tratamento térmico que consistiu em repousar em banho de gelo por 30 minutos, seguido de 45 segundos em banho maria a 42 °C com subsequente banho de gelo ao longo de 2 minutos. Neste momento, adicionou-se 800  $\mu$ L de meio LB líquido e o crescimento bacteriano foi estimulado mantendo as soluções em 37 °C e 200 rpm por 1 hora. Com o objetivo de concentrar as células, as soluções foram centrifugadas a 8000 g por 15 minutos a 4 °C e o precipitado foi ressuspenso em 40  $\mu$ L de meio LB.

As soluções concentradas contendo as células transformadas foram adicionadas sobre as placas de petri contendo 15 mL do meio LB ágar com o agente seletivo (canamicina) a 100  $\mu$ g/mL, em seguida, as placas foram incubadas em estufa a 37 °C por 20 horas para formação de colônias macroscópicas.

### 3.6 Expressão dos domínios

No dia seguinte iniciou-se a etapa de crescimento celular com a retirada de uma colônia isolada de cada placa de petri a ser adicionada em 5 mL de meio LB contendo canamicina a 100  $\mu$ g/mL. O cultivo foi incubado a 37 °C sob constante agitação a 200 rpm por um período de 6 horas. Em seguida, transferiu-se uma alíquota de 250  $\mu$ L para um erlenmeyer contendo 25 mL de meio LB líquido acrescido de canamicina a 100  $\mu$ g/mL, tendo, o cultivo, permanecido a 37 °C e 200 rpm durante 20 h. O terceiro dia do processo iniciou-se com a transferência de uma alíquota de 5 mL para um erlenmeyer contendo 1 L de meio LB e 100  $\mu$ g/mL de canamicina, tendo permanecido

a 37 °C e 200 rpm de agitação até atingir a densidade óptica (D.O.<sub>600nm</sub>) de 0,6 a 600 nm, medida no espectrofotômetro (Shimadzu UV-1601). Neste momento, uma vez atingida a D.O.<sub>600nm</sub> de 0,6 induziu-se a expressão dos domínios de interesse com a adição de 1 mM de IPTG por 18 horas a 18 °C em agitação constante a 200 rpm. Ao final da indução o extrato celular foi centrifugado a 8000 g em temperatura de 4 °C por 30 minutos. Foi feito o descarte do sobrenadante, enquanto o pellet foi congelado a -20 °C e guardado para as etapas de purificação.

### 3.7 Purificações dos domínios

O processo de purificação de ambos os domínios trabalhados (domínio PWWP das proteínas Pdp3 nativa e mutante P20W e PWWP da proteína NSD3s) foram conduzidos ao longo de três etapas, separadamente descritas abaixo:

#### **Etapa 1 - Cromatografia de afinidade a níquel (Ni<sup>2+</sup>)**

A primeira etapa de purificação consistiu em uma cromatografia de afinidade a níquel utilizando a coluna HisTag 5mL (GE Healthcare), acoplada a um sistema de purificação FPLC Äkta Prime Plus (GE Healthcare). Este processo se deu com base na afinidade entre os cátions divalentes (Ni<sup>2+</sup>) e os anéis imidazólicos presentes na cauda de histidinas dos domínios a serem purificados.

A purificação iniciou-se solubilizando o extrato celular obtido na etapa de expressão dos domínios em 50 mL do tampão A [500 mM NaCl, 50 mM TRIS, 5 mM imidazol, pH 8,0] acrescido de 250 µM de fluoreto de fenilmetilsulfonila (PMSF) e a ressuspensão foi lisada no sonicador de ponta (Sonica Vibracell) a 20% de amplitude em ciclos de 20 segundos com intervalos de 1 minuto por um período total de 30 minutos em sonicação. Então, centrifugou-se as amostras a 8000 g e 4 °C por 40 minutos. Descartou-se o *debris* celular e aproveitou-se o sobrenadante para as etapas seguintes.

Após a centrifugação as amostras foram filtradas em filtros 0,22 µm e injetadas na coluna cromatográfica previamente equilibrada com a passagem de 5 vezes o volume de coluna de tampão A. A injeção da amostra foi conduzida a 1,0 mL/min e as demais etapas da primeira purificação ocorreram em fluxo de 3 mL/min.

Passada a injeção da amostra realizou-se uma primeira lavagem de 5 vezes o volume de coluna com o tampão A e em seguida uma lavagem com tampão C [2 M NaCl, 50 mM TRIS, 5 mM imidazol, pH 8,0] foi conduzida para a retirada de ácidos nucleicos ligados ao domínio. Com isso, fez-se uma segunda lavagem com tampão A para retirada do excesso de sal presente no tampão C.

Ao final das lavagens iniciou-se a etapa de eluição, na qual utilizou-se um gradiente de aumento da concentração de imidazol partindo de 5 mM até 500 mM através da combinação dos tampões A e B [500 mM NaCl, 50 mM TRIS, 500 mM imidazol, pH 8,0] mantendo-se um fluxo de 3 mL/min durante 180 mL de eluição.

Cada etapa de lavagem foi coletada em tubos de centrífuga de 50 mL e a eluição foi coletada em microtubos de centrífuga de 1,5 mL, tendo sido retiradas alíquotas de 15 µL de cada uma das lavagens e também da eluição para análise por eletroforese.

As frações referentes ao pico de eluição do domínio foram identificadas por SDS PAGE, unidas e acrescidas de protease TEV em razão molar 1:50 de proteína para clivagem da cauda de histidinas. Tal clivagem foi realizada a 4 °C para o domínio PWWP-Pdp3 e a temperatura ambiente para o domínio PWWP-NSD3s, ambas ao longo de 18 horas. A solução de proteína e TEV foi transferida para uma membrana de diálise (SnakeSkin 3000 Da; Thermo Scientific) imersa em 2 litros de tampão D [500 mM NaCl, 50 mM TRIS, pH 8,0] com o objetivo de difundir por osmose o imidazol presente na amostra eluída dessa primeira etapa de purificação, afim de se evitar a precipitação proteica induzida pela presença de imidazol.

A diálise em questão foi realizada durante 20 horas a 4 °C sem agitação, e em seguida filtrou-se a amostra em filtro de 0,22 µM para dar início a segunda etapa do processo de purificação.

## **Etapa 2 - Cromatografia de afinidade a níquel (Ni<sup>2+</sup>)**

Como essa segunda etapa ocorre após a clivagem da cauda de histidinas, a proteína a ser purificada perde a afinidade pela coluna e é eluída na primeira lavagem, sendo importante para a separação de outras proteínas que também possuem afinidade a níquel.

Para tal, retirou-se a amostra da membrana de diálise e filtrou-se a mesma em filtro de 0,22 µM antes da injeção no cromatógrafo. A injeção ocorreu a 1 mL/min e a

amostra foi lavada com 5 vezes o volume da coluna para eluição do domínio PWWP. Em seguida, iniciou-se um gradiente de imidazol para eluição da protease TEV (que também possui uma cauda de seis histidinas não clivável) e demais proteínas do extrato celular. Retirou-se alíquotas de 15  $\mu$ L da amostra previamente à injeção, da lavagem da proteína de interesse e dos picos encontrados no gradiente para análise por SDS PAGE.

### **Etapa 3 - Cromatografia por exclusão molecular**

A terceira etapa de purificação consiste em uma purificação via exclusão molecular, na qual as proteínas são separadas pelo volume de eluição. Para tanto, utilizou-se uma coluna Superdex 75 PG HiLoad 16/60 (GE Healthcare) de 120 mL de volume total. O princípio de separação acontece devido à diferença de tamanho dos poros da resina que leva partículas menores a percorrerem um caminho maior passando por mais poros e conseqüentemente eluem depois de partículas maiores presentes em solução.

Foi necessário concentrar as amostras eluídas da segunda etapa de purificação previamente à injeção na coluna cromatográfica utilizando-se filtros de centrifuga (Amicon® Ultra 15 mL 3000 NMWL Centrifugal Filters). Com isso, as mesmas foram concentradas a 3000 g por ciclos de 15 minutos a 4 °C até atingir um volume final de 5 mLs, o que corresponde a um valor inferior a 5% do volume total da coluna.

A coluna foi equilibrada a com o tampão E [20 mM  $\text{PO}_4^{2-}$ , 100 mM NaCl, pH 7,4] para a maior parte das análises, exceto para análises de dicroísmo circular, que utilizaram o tampão F [10 mM  $\text{PO}_4^{2-}$ , 50 mM NaCl, pH 7,4] devido à técnica ser sensível a maiores concentrações de sais na solução tampão da amostra.

Após a coluna ser equilibrada, a amostra foi injetada e toda a corrida aconteceu a 1 mL/min (fluxo máximo suportado pela coluna). Os picos eluídos foram coletados em microtubos de centrifuga de 1,5 mL de volume e alíquotas de 15  $\mu$ L e 30  $\mu$ L foram retiradas respectivamente para eletroforese e dosagem de proteínas.

### 3.8 Eletroforese em gel de poliacrilamida sob condições desnaturantes (SDS PAGE)

O procedimento foi realizado segundo o protocolo descrito por Laemmli (1970), utilizando-se Mini-PROTEAN Tetra System 1 mm (Bio-Rad). Foram preparados dois géis separadamente, sendo um chamado de fracionador e o outro, concentrador. Para o gel fracionador 12% preparou-se 5 mL de uma mistura contendo bis-acrilamida 0,4% e acrilamida 11,6% (p/v) acrescido de Tris-HCl 0,39 M (pH 8,8) e dodecil sulfato de sódio (SDS) 0,1% (p/v). Imediatamente antes da aplicação no suporte adicionou-se persulfato de amônio (PSA) 0,1% e tetrametiletilenodiamina (TEMED) 0,07%. Enquanto isso, para o gel concentrador 5% preparou-se 2 mLs de solução contendo bis-acrilamida 0,16% e acrilamida 4,84% (p/v) acrescido de Tris-HCl 0,125 M (pH 6,8) e SDS 0,1% (p/v). Imediatamente antes da aplicação no suporte adicionou-se PSA 0,1% e TEMED 0,07%. Antes da aplicação do gel concentrador aguardou-se 5 minutos para a polimerização do gel fracionador e em seguida adicionou-se o separador de 10 poços curtos (Bio-Rad). Cada amostra proteica analisada por esta metodologia foi acrescida de tampão de amostra 1x [Tris-HCl 50 mM (pH 6,8), SDS 2% (p/v), DTT 100 mM, azul de bromofenol 0,1% (p/v) e glicerol 10% (v/v)] e posteriormente mantida em banho-maria 100 °C por um período de 5 minutos previamente à sua aplicação no gel.

Os géis foram corridos a 30 mA por um período de 40 minutos em tampão de corrida [Tris-HCl 25 mM (pH 8,3), Glicina 250 mM e SDS 0,1% (p/v)]. A visualização apropriada das proteínas foi feita por meio de coloração com Comassie Blue R (Vetec).

### 3.9 Dosagem de proteínas

A determinação da concentração de amostra proteica foi conduzida pelo método colorimétrico no qual fez-se uso do reagente Pierce 660 nm Protein Assay (Thermo Scientific). À uma placa de 96 poços adicionou-se 10 µL da amostra a ser dosada e 150 µL do reagente Pierce. Em seguida, o meio foi incubado ao abrigo da luz por um período de 5 minutos. Foram conduzidas, então, medidas de absorvância a 660 nm das amostras através do emprego do espectrofotômetro SpectraMax M2e (Molecular Devices). Todos os ensaios foram conduzidos em triplicata e a

concentração de cada amostra proteica foi determinada por meio de uma curva padrão elaborada com albumina de soro Bovino (BSA) entre 0,025 e 2 mg/mL.

### 3.10 Differential Scanning Fluorimetry (DSF)

DSF é uma técnica de medição da estabilidade termodinâmica de proteínas através de uma sonda extrínseca, que no caso utilizou-se o corante *Sypro Orange*, devido a sua propriedade de ligação não específica a regiões hidrofóbicas das proteínas. À medida que se aumenta a temperatura do sistema, mede-se a intensidade de fluorescência do corante, indicando o grau de exposição do core hidrofóbico ao solvente.

Inicialmente, adicionou-se 4  $\mu$ L do corante *Sypro Orange* 5000x em 5 mL dos domínios recém purificados em tampão 10 mM fosfato de sódio e 50mM NaCl. Em seguida, alíquotas de 40  $\mu$ L dos domínios foram pipetadas em uma placa de PCR de 96 poços específicas para análise no RT-PCR modelo (Bio-Rad CFX96), seguidas pela adição de 10  $\mu$ L de solução tampão 5x fosfato de sódio (pH 7,0), de forma a manter a concentração final dos domínios a 5  $\mu$ M.

Para a análise dos domínios complexados ao peptídeo derivado da histona H3K36me3 foram adicionadas alíquotas de forma a manter a concentração final tanto de peptídeo quanto de proteína a 5  $\mu$ M.

A variação térmica aferida de 25 a 95°C na velocidade de 1°C/min e, os dados coletados foram processados através do programa Bio-Rad CFX manager 3.1, sendo capaz de determinar o TM das curvas de desnaturação através da primeira derivada.

### 3.11 Dicroísmo circular (CD)

A técnica de dicroísmo circular foi utilizada para analisar o perfil de estrutura secundária dos domínios PWWP e caracteriza-los quanto ao estado enovelado e, estabilidade termodinâmica através de rampas de temperatura, visto que os domínios foram expressos de forma heteróloga. A técnica permiti incidir uma luz circularmente polarizada sobre a amostra, calculando a diferença dos desvios à direita e à esquerda nas componentes vertical e horizontal sofridos pela luz circularmente polarizada. Já a amostra responde ao sinal enviado pelo equipamento através das diferentes angulações nas posições dos carbonos quirais presentes nas estruturas secundárias,

sendo capazes de desviar a luz plano polarizada. Tais desvios reportam a elipticidade segundo a fórmula:

$$\theta = \tan^{-1} (\text{eixo maior}) / (\text{eixo menor})$$

Os domínios PWWP foram analisados no equipamento Jasco-715 (Jasco Corporation) acoplado a um sistema Peltier para controle da temperatura com a utilização de uma cubeta de quartzo 1,0 mm. As concentrações utilizadas foram de 15  $\mu\text{M}$ , 25  $\mu\text{M}$  e 35  $\mu\text{M}$ . Os espectros foram coletados no intervalo de comprimento de onda de 260 nm a 190 nm, com um tempo de resposta de 0,5 segundo, largura de banda de 1,0 nm, resolução espectral de 0,5 nm e número de acumulações igual a 3. Os espectros foram convertidos de elipticidade para elipticidade molar segundo a equação:

$$[\theta] = \theta / (10 \cdot [p] \cdot l \cdot n)$$

Sendo “[ $\theta$ ]” a elipticidade molar em  $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ , “ $\theta$ ” a elipticidade, “[p]” a concentração molar de proteína, “l” o caminho óptico da cubeta em centímetros e “n” o número de aminoácidos dos domínios.

A rampa de temperatura foi realizada fixando-se o comprimento de onda a 208 nm, tendo sido iniciada a 20 °C na velocidade de 2 °C/min até 90 °C, coletando-se os dados a cada 2 °C.

### 3.12 Peptídeos derivados das Histonas

Peptídeos de 13 resíduos de aminoácidos foram comprados da empresa GenScript<sup>®</sup>, cujas sequências descrevem regiões presentes nas caudas das histonas onde já se relataram modificações pós traducionais de metilações em lisinas. Todos os peptídeos possuem uma modificação de trimetilação nas lisinas, exceto para o peptídeo H3K36, que foi analisado desde o seu estado não metilado até trimetilado, como descrito na tabela a seguir.

Tabela 1: Sequência primária dos peptídeos derivados das caudas das histonas.

Peptídeo	Sequência
H3K4me3	Ac-MART <b>Kme3</b> QTARKSTG-Amida
H3K9me3	Ac-TKQTAR <b>Kme3</b> STGGKA-Amida
H3K27me3	Ac-ATKAAR <b>Kme3</b> SAPATG-Amida
H3K36	Ac-PATGGV <b>K</b> KPHRYR-Amida
H3K36me	Ac-PATGGV <b>Kme</b> KPHRYR-Amida
H3K36me2	Ac-PATGGV <b>Kme2</b> KPHRYR-Amida
H3K36me3	Ac-PATGGV <b>Kme3</b> KPHRYR-Amida
H3K79me3	Ac-EIAQDF <b>Kme3</b> TDLRFQ-Amida
H4K20me3	Ac-GAKRHR <b>Kme3</b> VLRDNI-Amida

Os peptídeos foram recebidos na forma liofilizada, ressuspendidos em água ultra pura Milli-Q e armazenados em freezer -20 °C para análises posteriores de interação proteína-peptídeo por espectroscopia de fluorescência.

### 3.13 Espectroscopia de Fluorescência

A espectroscopia de fluorescência foi utilizada para análise do grau de exposição dos triptofanos ao solvente dos domínios livres, além da quantificação da constante de dissociação ( $K_d$ ) entre os domínios PWWP e os peptídeos derivados das histonas.

Os domínios PWWP apresentam triptofanos em posições paralelas e perpendiculares a outros resíduos aromáticos, formando sítios de interação com lisinas metiladas das caudas das histonas. Desta maneira, foi possível a utilização da espectroscopia de fluorescência para determinação do  $K_d$  usando os triptofanos como

sonda intrínseca. Titulou-se os peptídeos derivados das histonas sobre os domínios PWWP, observando-se uma supressão do máximo de emissão dos triptofanos. A área total sob o domínio livre e sob cada ponto de titulação foi normalizada e, em seguida, fitada pela equação:

$$\left( a \frac{(Kd1 + (x^n) + 2) - (((Kd1 + (x^n) + 2)^2) - 4 (x^n) 2)^{0,5}}{4} \right) + b \left( \frac{x^n}{Kd2 + (x^n)} \right)$$

A equação fita os dados por duas constantes de dissociação para dois sítios de ligação diferentes, sendo *a* e *b* máximos de saturação de *y* para cada um dos sítios de ligação (LEITGEB *et al.*, 2009).

Realizou-se os experimentos no equipamento Agilent Technologies (Cary Eclipse Fluorescence Spectrophotometer) a 25 °C sob excitação de luz a 295 nm, enquanto a emissão observada variou de 310 nm a 450 nm.

Para análise do domínio PWWP-NDS3 fixou-se as fendas de emissão e excitação em 5 nm, a voltagem da fotomultiplicadora foi mantida a 880 V, a concentração de amostra a 2 µM e os peptídeos derivados das histonas foram titulados de 0 a 276 µM.

Para análise do domínio PWWP-Pdp3 fixou-se as fendas de emissão em 10 nm, as fendas de excitação em 5 nm, a voltagem da fotomultiplicadora em 800 V, a concentração de amostra a 2,5 µM e os peptídeos derivados das histonas foram titulados de 0 a 416 µM.

### 3.14 Oligonucleotídeos de DNA

Foram adquiridos cinco oligonucleotídeos de DNA da empresa GenScript®, que continham de 15 a 35 nucleotídeos simples fita, descritos na tabela a seguir:

Tabela 2: Sequências dos oligonucleotídeos de DNA.

Oligonucleotídeos	Sequências
Oligo 1	5'/RhoR-XN/GGG GCA TGC CTG CAG GTC GAC TCT AGA GGA TCC CC 3'
Oligo 2	5' GGG GAT CCT CTA GAG TCG ACC TGC AGG CAT GCC CC3'
Oligo 3	5' GGG GAT CCT CTA GAG TTG ACC TGC AGG CAT GCC CC 3'
Oligo 4	5' TCG ACC TGC AGG CAT GCC CC 3'
Oligo 5	5' GGG GAT CCT CTA GAG 3

O oligonucleotídeo 1 é composto por uma sequência randômica de 35 bases fusionado a uma cauda de rodamina, que será a sonda fluorescente dos experimentos de EMSA. Já o oligonucleotídeo 2 consiste na sequência complementar ao oligonucleotídeo 1, enquanto o oligonucleotídeo 3 possui um base não complementar ao oligonucleotídeo 1. Os oligonucleotídeos 4 e 5 são fragmentos do oligonucleotídeo 2, que, quando anelados ao oligonucleotídeo 1, darão origem a uma estrutura de DNA dupla fita com ausência de uma ligação fosfodiéster. O objetivo consistiu em anelar os oligonucleotídeos com o intuito de formar DNA dupla fita com diferenças estruturais.

Para anelar as fitas acima descritas, alíquotas do oligonucleotídeo 1 foram incubadas com os oligonucleotídeos 2 e 3 separadamente e com os oligonucleotídeos 4 e 5 juntos, todos em razão molar 1:1. Assim, obtiveram-se 4 oligonucleotídeos de DNA dupla fita com diferenças estruturais.

O processo de anelamento foi realizado no equipamento CFZ96 (BioRad) através de um ciclo no qual os DNAs foram aquecidos a 95 °C e resfriados em uma taxa de 1 °C/min até a temperatura final de 25 °C.

### 3.15 EMSA

Os ensaios de mudança de mobilidade eletroforética (*EMSA*) utilizam a eletroforese para observar as mudanças no padrão de migração de uma determinada amostra em diferentes condições. Tal técnica foi aplicada com o intuito de avaliar a afinidade de ligação entre os domínios PWWP com oligonucleotídeos de DNA descritos no tópico anterior.

Preparou-se 7 mL de géis de poliacrilamida 10% utilizando o sistema Mini-PROTEAN Tetra System 1 mm (Bio-Rad), com a seguinte composição: 2,66 mL de água ultra pura Milli-Q, bis-acrilamida 0,3%, acrilamida 9,7%, solução tampão TBE 1x [tris (hidroximetil) aminometano 0,023 M, ácido bórico 0,022 M e EDTA 0,83 mM], SDS 0,1% (p/v). Então, adicionou-se persulfato de amônio (PSA) 0,1%, tetrametiletilenodiamina (TEMED) 0,07% e, por último, um espaçador com 10 poços.

Microtubos de centrifuga foram utilizados para promover as reações de ligação entre os domínios PWWP com os oligonucleotídeos de DNA já anelados. Adicionou-se à cada microtubo a solução tampão de ligação [TRIS 50 mM, EDTA 1mM, DTT 1 mM, BSA 0,2 mg/mL e glicerol 12,5%], os domínios PWWP e os DNAs nas razões molares de (1 : 0,25), (1 : 0,5), (1 : 1), (1 : 3), (1 : 5) e (1 : 10) dos domínios em relação ao DNA. As reações ocorreram durante 30 minutos a temperatura ambiente.

O gel correu 30 minutos vazio antes das aplicações das amostras, iniciando-se pelo DNA livre no primeiro poço e as concentrações crescentes de proteína em relação ao DNA nos poços subsequentes. Após as aplicações das amostras nos poços, os géis correram por 20 minutos a 150 V e banho de gelo, respeitando-se o volume de 10  $\mu$ L de amostra em cada poço.

## 4 RESULTADOS

### 4.1 Modelagem Molecular

O domínio PWWP-NSD3s já possuía estrutura determinada e depositada no PDB, por isso a modelagem foi necessária somente para o domínio PWWP-Pdp3. Esta teve como objetivo comparar estruturalmente os dois domínios através dos dados de estabilidade termodinâmica e interação com ligantes.

Inicialmente obteve-se através do BLAST os domínios com maior percentual de identidade de sequência em relação ao domínio PWWP-Pdp3 e que possuem estrutura depositada no PDB. Encontrou-se os domínios PWWP-HDGF (PDB: 2B8A), PWWP-p75/LEDGF (PDB: 2M16) e PWWP-PSIP1 (PDB: 4FU6), os quais apresentaram respectivamente 43,8%, 48% e 48% de semelhança na sequência primária (Tabela 3).

Esses domínios alinhados ao PWWP-Pdp3 apresentaram valores de *max score* entre 29,6 e 35,0, o que indica que o número de resíduos de aminoácidos semelhantes entre os diferentes domínios converge. Além disso, os valores de *max score* e *total*

*score* se repetirem para um mesmo alinhamento indicam que os domínios conservam uma única região de maior semelhança na sequência primária.

O *query cover* reporta de 15% a 16% de porcentagem de resíduos de aminoácidos da sequência alinhados entre os dois domínios. Já o *E. value* é um parâmetro quantitativo de confiança que mostra uma qualidade maior no alinhamento do domínio PWWP-Pdp3 com os domínios PWWP-p75/LEDGF e PSIP1 em comparação com o alinhamento com o domínio PWWP-HDGF.

Tabela 3: Resultados encontrados para os 3 domínios com maior percentual de identidade com o domínio PWWP-Pdp3.

	<i>Max score</i>	<i>Total score</i>	<i>Query Cover</i>	<i>E value</i>	<i>Per. ident</i>	<b>PDB</b>
<b>HDGF PWWP Domain</b>	<b>29,6</b>	<b>29,6</b>	<b>15%</b>	<b>0,72</b>	<b>43,48%</b>	<b>2B8A</b>
<b>P75/LEDGF PWWP Domain</b>	<b>33,5</b>	<b>33,5</b>	<b>16%</b>	<b>0,027</b>	<b>48%</b>	<b>2M16</b>
<b>PSIP1 PWWP Domain</b>	<b>35,0</b>	<b>35,0</b>	<b>16%</b>	<b>0,015</b>	<b>48%</b>	<b>4FU6</b>

Através da plataforma J pred foi possível prever o perfil de estrutura secundária do domínio PWWP-Pdp3, o que gerou como resultado uma sequência com 5 fitas  $\beta$  na região N-terminal e 3  $\alpha$ -hélices na região C-terminal, conforme observado para demais domínios PWWP com estrutura determinada. Além disso, a predição já previa uma cauda desenovelada na região C-terminal de 40 resíduos de aminoácidos (Figura 4).

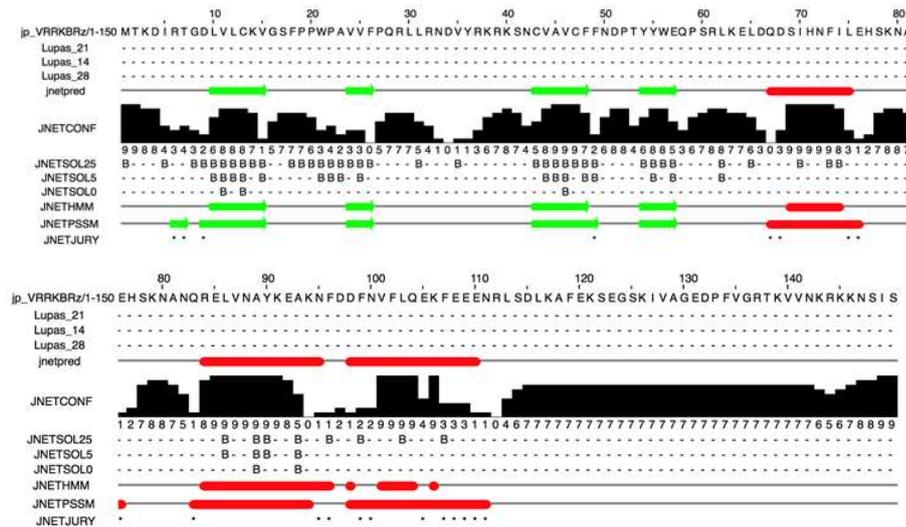


Figura 4: Predição de estrutura secundária do domínio PWWP-NSD3s. Em verde as fitas  $\beta$  e em vermelho as  $\alpha$ -hélices preditas pela plataforma Jpred 4. A presença de cinco fitas  $\beta$  na região N-terminal e de uma a seis  $\alpha$ -hélices na região C-terminal consiste em um perfil comum em domínios PWWP. Observa-se também a não predição de estrutura para os últimos quarenta resíduos de aminoácidos da cauda C-terminal.

Através da plataforma Praline web foi possível determinar o alinhamento múltiplo do domínio PWWP-Pdp3 e dos moldes PWWP-HDGF, PWWP-p75/LEDGF e PWWP-PSIP1, o qual já demonstrava um alinhamento melhor da região N-terminal, considerando o alinhamento global. Entretanto, os dados também foram analisados quanto às posições das estruturas secundárias dos moldes em relação à predição de estrutura secundária do PWWP-Pdp3 e, observa-se no máximo 2 resíduos de aminoácidos não alinhados dentro das fitas  $\beta$  (Fen 31 e Pro 32), já as  $\alpha$ -hélices apresentam de 3 a 8 resíduos de aminoácidos desalinhados exceto para a  $\alpha$ -hélice  $\alpha 2$  (figura 5).

Os domínios PWWP ligam-se a lisinas metiladas das caudas das histonas através do barril  $\beta$  formado na região N-terminal, ou seja, a região N-terminal é mais conservada por sustentar o sítio de interação com o nucleossomo. Já a região C-terminal rica em  $\alpha$  hélices e loops suportam o barril  $\beta$  e variam de domínio para domínio.

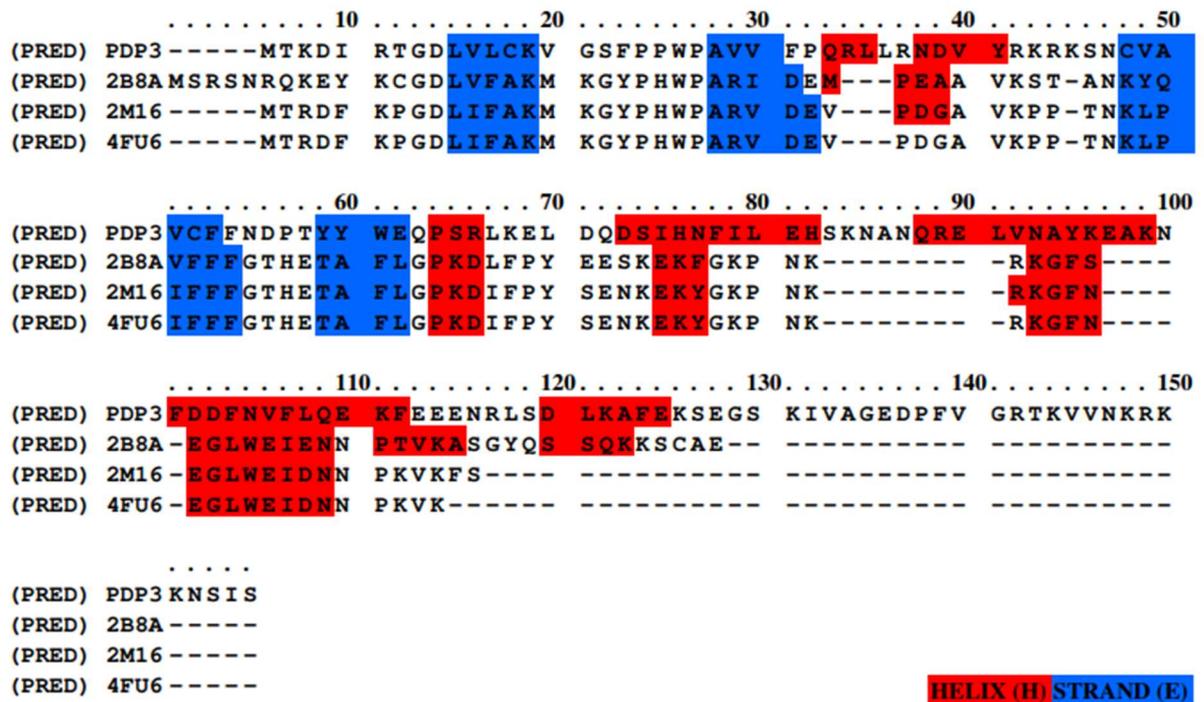


Figura 5: Alinhamento múltiplo do domínio PWWP-Pdp3 com os moldes PWWP-HDGF (PDB: 2B8A), PWWP- p75/LEDGF (PDB: 2M16) e PWWP-PSIP1 (PDB: 4FU6). Em azul as fitas  $\beta$  e em vermelho as  $\alpha$ -hélices. As estruturas secundárias previstas para o domínio PWWP-Pdp3 conservam maiores semelhanças nos 70 primeiros resíduos de aminoácidos em relação aos moldes. Enquanto isso, a região C-terminal rica em  $\alpha$  hélices (vermelho) não se sobrepõe da mesma maneira.

Realizou-se a modelagem comparativa do domínio PWWP-Pdp3 através do programa Modlller, o que resultou em 400 possíveis estruturas para o domínio PWWP-Pdp3. Desses modelos separou-se os três modelos de menor valor de energia absoluto e os três modelos de menor valor de energia normalizado.

Utilizou-se o algoritmo DOPE, fornecido pela própria plataforma de modelagem em questão, para determinar os modelos de menor energia. Neste sentido, o modelo A, representado na figura 6A, aparece tanto entre os três modelos de menor DOPE quanto entre os três modelos de menor DOPE normalizado simultaneamente (figura 6).

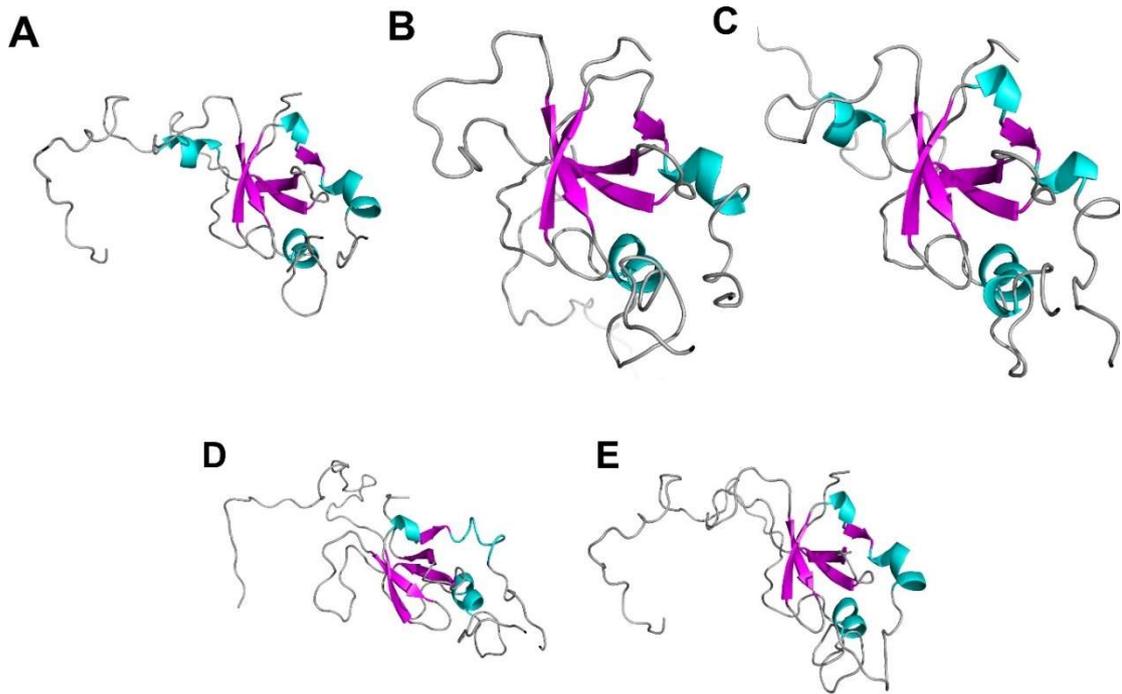


Figura 6: Modelos estruturais do domínio PWWP-Pdp3 obtidos a partir da modelagem comparativa. Encontram-se aqui representadas as cinco estruturas com os menores valores de energia, obtidas através do algoritmo DOPE fornecido pela plataforma Moddler.

Os cinco modelos selecionados foram validados através da plataforma SAVES v5.0. Analisando-se o modelo B (Figura 6B) verifica-se que o mesmo, além de possuir menos  $\alpha$ -hélices que os demais modelos, também apresenta os piores resultados em dois dos quatro algoritmos utilizados para validação (Tabela 4). Logo, tal modelo foi desconsiderado como uma possível estrutura para o domínio PWWP-Pdp3. Dados estatísticos das posições dos átomos dentro de proteínas com estrutura determinada em alta resolução são utilizados para validação de modelos através da ferramenta chamada ERRAT. O modelo C (Figura 6C) apresentou valores de ERRAT menores que todos os demais modelos e moldes (Tabela 4) logo, esse modelo apresenta uma possível desconformidade da posição dos átomos não sequenciais com relação à estrutura do PWWP-Pdp3.

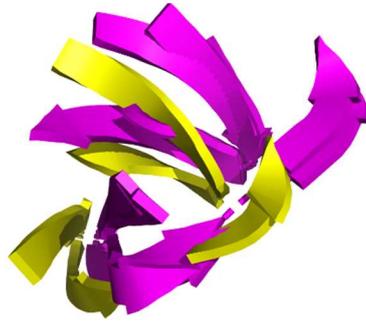


Figura 7: Sobreposição do barril  $\beta$  do modelo D com o barril  $\beta$  dos moldes PWWP-HDGF, PWWP- p75/LEDGF e PWWP-PSIP1. Em rosa os barris  $\beta$  dos moldes e em amarelo o barril  $\beta$  do modelo D.

Ao avaliar o modelo D (figura 6D) nota-se que o mesmo apresenta os melhores valores de validação em comparação com os demais modelos, o que indica uma conformação possivelmente mais estável. No entanto, a região do barril beta do modelo em questão não alinha com nenhum dos moldes utilizados na modelagem (Figura 7). Desta maneira, seria possível concluir que o modelo D seria, em tese, mais estável, porém o mesmo não representa o domínio PWWP-Pdp3, uma vez que sua representação espacial não apresenta a sobreposição esperada. Logo, descarta-se o modelo D como possível representação espacial do domínio PWWP-Pdp3.

Tabela 4: Valores encontrados na validação dos modelos através da plataforma SAVE v5.0.

<b>Nome</b>	<b>Verify</b>	<b>ERRAT</b>	<b>Prove</b>	<b>PROCHECK</b>
<b>Molde 2B8A</b>	87,27%	48,8889	7,4%	4 erros 1 duvidoso 3 acertos
<b>Molde 2M16</b>	77,42%	25,3165	8,3%	3 erros 2 duvidosos 3 acertos
<b>Molde 4FU6</b>	32,58%	100,0000	0,4%	0 erros 3 duvidosos 5 acertos
<b>Modelo A</b>	40,00%	26,1905	6,6%	2 erros 3 duvidosos 3 acertos
<b>Modelo B</b>	37,30%	27,7778	10,0%	4 erros 1 duvidoso 3 acertos
<b>Modelo C</b>	42,67%	16,9492	5,9%	2 erros 3 duvidosos 3 acertos
<b>Modelo D</b>	46,00%	42,5373	7,1%	3 erros 3 duvidosos 2 acertos
<b>Modelo E</b>	40,67%	39,8438	5,3%	5 erros 1 duvidoso 2 acertos

Tanto o modelo A (Figura 6A) quanto o modelo E (Figura 6E) alinham melhor entre si e com os moldes em comparação com os demais modelos, resultando nos dois modelos que melhor representariam o domínio PWWP-Pdp3. No entanto, a diferença entre os dois modelos está na presença de uma  $\alpha$ -hélice entre as fitas  $\beta 2$  e  $\beta 3$  presente no modelo A e ausente no modelo E.

Segundo Wang R. et al (2019) a presença de  $\alpha$ -hélices entre as fitas  $\beta 2$  e  $\beta 3$  é uma exceção aos domínios PWWP presente somente nos domínios PWWP-BRPF1, PWWP-TbTFIIS2-2 e PWWP-Pdp1. Tratando-se assim de uma exceção, diminui-se a

probabilidade de o modelo A representar o domínio PWWP-Pdp3. Além disso, a análise através da plataforma *Predictor of Natural Disordered Regions* (PoNDR) possibilitou prever uma desordem na região entre os resíduos de aminoácidos 25 a 50 do domínio PWWP-Pdp3 referente à região entre as fitas  $\beta 2$  e  $\beta 3$  (Figura 8). Assim, infere-se que possivelmente o modelo E (Figura 6E) é a melhor representação do domínio PWWP-Pdp3. Os dados obtidos pela plataforma PoNDR também indica que a região C-terminal do domínio PWWP-Pdp3 apresenta uma cauda desenovelada de aproximadamente 40 resíduos de aminoácidos, conforme observado nos modelos na figura 6.

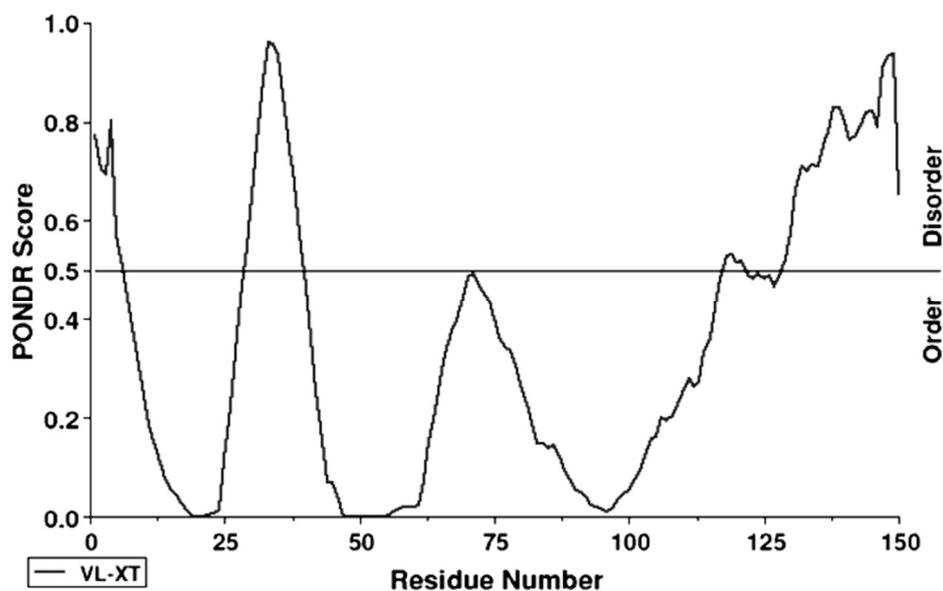


Figura 8: Resultado obtido da plataforma PONDR para o domínio PWWP-Pdp3. Valores acima de 0,5 indicam possíveis regiões desordenadas.

#### 4.2 Expressão e Purificação dos Domínios

A primeira etapa de purificação, tendo sido uma cromatografia de afinidade com coluna de níquel, teve como objetivo eliminar da amostra solúvel o maior teor possível de proteínas do vetor bacteriano utilizado na expressão (*E. coli*). A cromatografia do domínio PWWP-Pdp3 apresentou um platô inicial estabilizado em cerca de 1200 mAu de absorbância a 280 nm referente às proteínas sem interação com a coluna de níquel, seguido de uma elevação menor que 200 mAu durante a lavagem com solução 2 M de NaCl. Tal fenômeno indica assim uma possível eluição de produtos de

interação eletrostática com as proteínas aderidas à coluna. Finalmente, observou-se um pico em aproximadamente 160 mM de imidazol durante o gradiente imposto de 5 mM a 500 mM de imidazol, o que indica a eluição de proteínas passíveis de interação com o níquel (figura 9A).

O gel de eletroforese SDS PAGE contém em cada poço uma fração da amostra colhida em cada etapa da cromatografia exceto pelas duas primeiras colunas. Enquanto a primeira é referente aos padrões de peso molecular (Thermo Unstained Protein MW Marker - Thermo Scientific - 26610), na segunda coluna encontra-se a amostra antes de ser injetada no cromatógrafo. Em seguida, depositou-se a amostra referente às proteínas eluídas sem interagir com a coluna (flow-through) e, por último as proteínas eluídas durante o gradiente de imidazol (figura 9B). É visível uma concentração de proteínas próximas ao padrão de 18 kDa na amostra antes da injeção e nas proteínas eluídas no gradiente de imidazol. Logo, conclui-se que a etapa de purificação foi bem sucedida, visto que a construção a ser purificada PWWP<sub>Pdp3</sub>-HIS<sub>6</sub> apresenta uma massa molecular de 19,7 kDa.

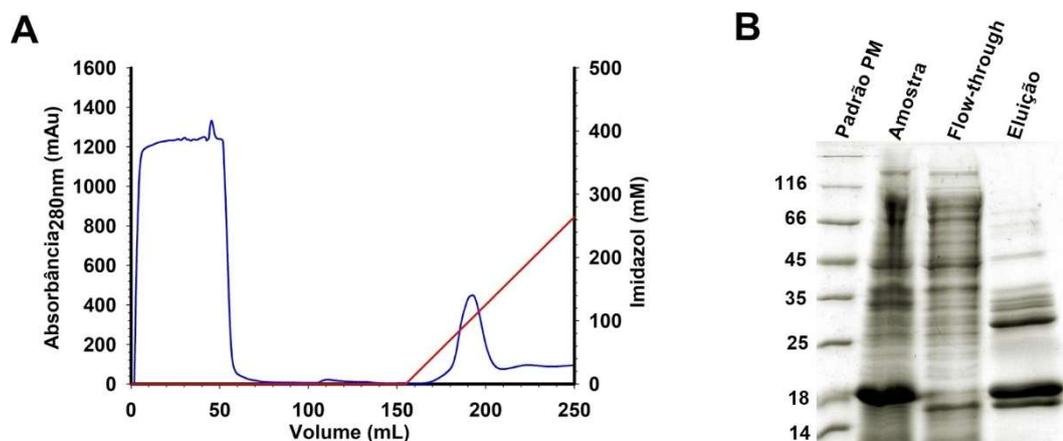


Figura 9: Cromatograma e SDS PAGE da primeira purificação do domínio PWWP-Pdp3. (A) Cromatograma da primeira purificação (em azul) mostrando o platô inicial de eluição (0 - 50 mL) das proteínas celulares do vetor de expressão sem interação com a coluna de Ni<sup>2+</sup>, seguido das lavagens com e sem alta concentração de NaCl (50 - 150 mL) e, por último, o gradiente de imidazol (em vermelho) com eluição da possível proteína de interesse em aproximadamente 160 mM de imidazol (180 a 210 mL). (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteínas que não interagiram com a coluna e proteínas eluídas.

Após a primeira etapa de purificação utilizou-se a protease TEV na razão molar 1:50 para a clivagem da cauda HIS<sub>6</sub>. A obtenção da proteína de interesse passou a ser observada na lavagem da segunda purificação por afinidade a níquel, devido à

perda de afinidade pela coluna após a retirada da cauda de histidinas. As demais proteínas do extrato celular, que apresentam afinidade a níquel, continuam ligadas a coluna durante a eluição da proteína de interesse.

O cromatograma da segunda purificação do domínio PWWP-Pdp3 apresenta um platô inicial (0 - 60 mL), seguido da lavagem da coluna (60 - 85 mL) com a eluição do domínio PWWP-Pdp3 e, por fim, ao longo do gradiente de imidazol eliminam-se as demais proteínas com afinidade a níquel e a protease TEV (28 kDa).

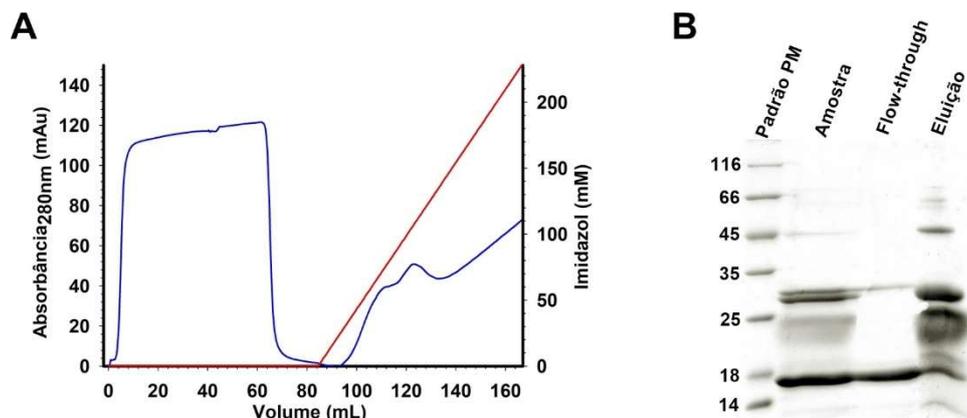


Figura 10: Cromatograma e SDS PAGE da segunda purificação do domínio PWWP-Pdp3. (A) Cromatograma da segunda purificação (em azul), mostrando o platô inicial de eluição (0 - 60 mL) das proteínas sem interação com a coluna de  $\text{Ni}^{2+}$ , seguido das lavagens até a estabilização da linha de base (60 - 85mL) e, por último, o gradiente de imidazol (em vermelho) com eluição de dois picos entre 50 e 100 mM de imidazol. (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteína de interesse (17,4 kDa) e demais proteínas eluídas no gradiente de imidazol.

Como o gel de SDS-PAGE da segunda etapa de eluição não apresentava uma única banda, realizou-se a terceira etapa de purificação, sendo esta uma cromatografia de exclusão molecular. Para tanto, utilizou-se uma coluna de 120 mL, sendo o volume de eluição do domínio PWWP-Pdp3 de 75 mL (figura 11A), que, através da curva de calibração da coluna, corresponde a um domínio globular de 16,8 kDa. Correu-se um gel eletroforese SDS PAGE com alíquotas de antes e depois à cromatografia de exclusão molecular, obtendo-se uma única banda próxima ao padrão de 16 kDa, o que se aproxima da massa molecular do domínio PWWP-Pdp3 (17,4 kDa) (figura 11B).

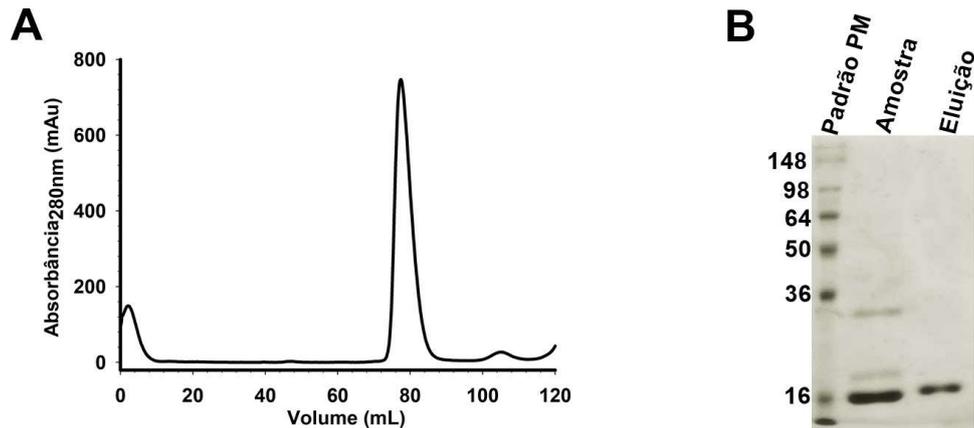


Figura 11: Cromatografia de exclusão molecular e gel de eletroforese SDS PAGE do domínio PWWP-Pdp3. (A) Cromatograma da exclusão molecular do domínio PWWP-Pdp3. Observa-se um pico inicial (0 - 10 mL) de estabilização da linha de base, um pico de maior intensidade (75 - 90 mL) dentro do volume de eluição esperado para o domínio PWWP-Pdp3 e ondulações menores ao final do cromatograma, referentes à eliminação de componente menores presentes na solução tampão. (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteína de interesse (17,4 kDa).

#### 4.3 Purificações do domínio PWWP-NSD3s

A primeira etapa de purificação do domínio PWWP-NSD3s consistiu em uma cromatografia de afinidade a níquel. Esta apresentou um platô inicial em aproximadamente 2000 mAu referente às proteínas que não interagiram com a coluna, seguida por uma elevação em 50 mAu durante a lavagem com 2 M de NaCl, devido à eliminação de compostos que interagem por ligação eletrostática. Finalmente, observou-se um pico em 100 mM de imidazol referente à eluição das proteínas que interagiram com a coluna de níquel (figura 12A). O gel de eletroforese SDS PAGE das alíquotas coletadas de cada etapa da purificação se encontra na figura 12B, iniciando pelo padrão de peso molecular, seguido da amostra antes da injeção, proteínas que não interagiram com a coluna durante a injeção (*flow-through*) e, por último, as proteínas eluidas ao longo do gradiente de imidazol.

A construção PWWP<sub>(NSD3s)</sub>-GST-HIS<sub>6</sub> possui 45 kDa e pode ser observada super expressa na amostra antes da injeção, estando também presente na amostra eluída do gradiente de imidazol.

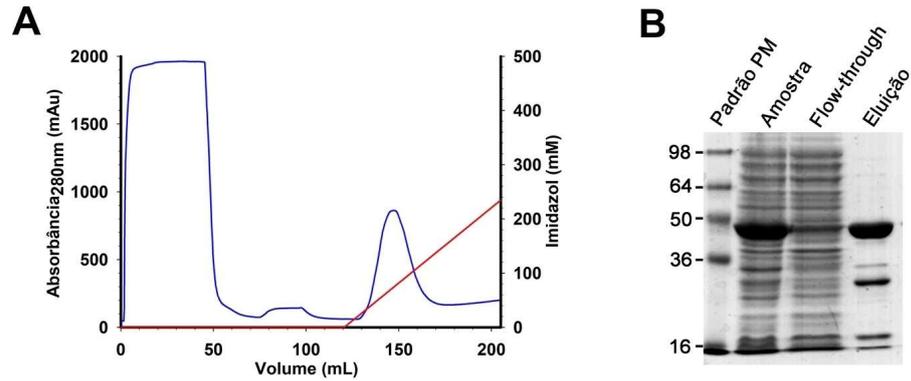


Figura 12: Cromatograma e SDS PAGE da primeira purificação do domínio PWWP-NDS3. (A) Cromatograma da primeira purificação (em azul), mostrando o platô inicial de eluição (0 - 50 mL) das proteínas celulares do vetor de expressão sem interação com a coluna de  $\text{Ni}^{2+}$ , seguido das lavagens com e sem alta concentração de NaCl (50 - 120 mL) e, por último, o gradiente de imidazol (em vermelho) com eluição da possível proteína de interesse em aproximadamente 150 mM de imidazol (130 a 160 mL). (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteínas que não interagiram com a coluna e proteínas eluídas no gradiente de imidazol.

A segunda etapa de cromatografia de afinidade a níquel do domínio PWWP-NSD3s ocorreu após uma etapa de clivagem com a protease TEV, responsável pela separação do domínio PWWP das caudas GST-HIS<sub>6</sub>. A figura 13A apresenta o cromatograma da segunda etapa de purificação com um platô inicial de 0 a 60 mL, onde é eluído o domínio PWWP, uma lavagem de estabilização da linha de base (60 a 80 mL) e um gradiente de imidazol para eliminação da cauda GST-HIS<sub>6</sub> e da protease TEV. Empregou-se um gel de eletroforese SDS PAGE com alíquotas da amostra antes de a injeção com o domínio PWWP-NSD3s (17,5 kDa) clivado da calda GST-HIS<sub>6</sub> (27,5 kDa), seguida por alíquotas do *flow-through* e da lavagem contendo apenas o domínio PWWP-NSD3s e, por último, o gradiente de imidazol com a eluição da cauda GST-HIS<sub>6</sub> e da protease TEV, ambos próximos a 28 kDa (figura 13B).

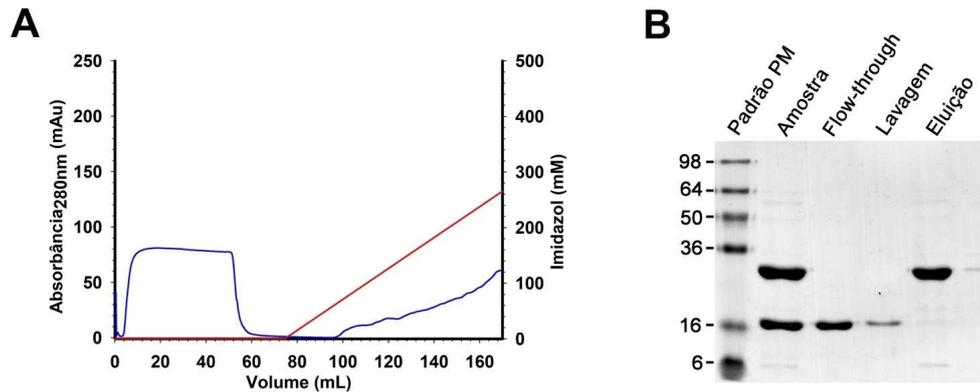


Figura 13: Cromatograma e SDS PAGE da segunda purificação do domínio PWWP-NSD3s. (A) Cromatograma da segunda purificação (em azul), mostrando um platô inicial de eluição (0 - 60 mL) das proteínas sem interação com a coluna de  $\text{Ni}^{2+}$ , seguido das lavagens até a estabilização da linha de base (60 - 80 mL) e, por último, o gradiente de imidazol (em vermelho) com eluição da protease TEV e da cauda GST-HIS<sub>6</sub>. (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteína de interesse (17,4 kDa) eluída no *flow-through* e lavagem e demais proteínas eluídas no gradiente de imidazol.

A última etapa de purificação do domínio PWWP-NSD3s correspondeu a uma cromatografia de exclusão molecular, na qual o volume de eluição do domínio foi de 76,6 mL, correspondendo a um domínio globular de 15,6 kDa, sendo a massa molecular do domínio 17,4 kDa (figura 14A). Correu-se um gel de eletroforese SDS-PAGE com as alíquotas antes de a injeção e do pico eluído da cromatografia de exclusão molecular, mostrando uma única banda no pico eluído e próxima ao padrão de 16 kDa (Figura 14B).

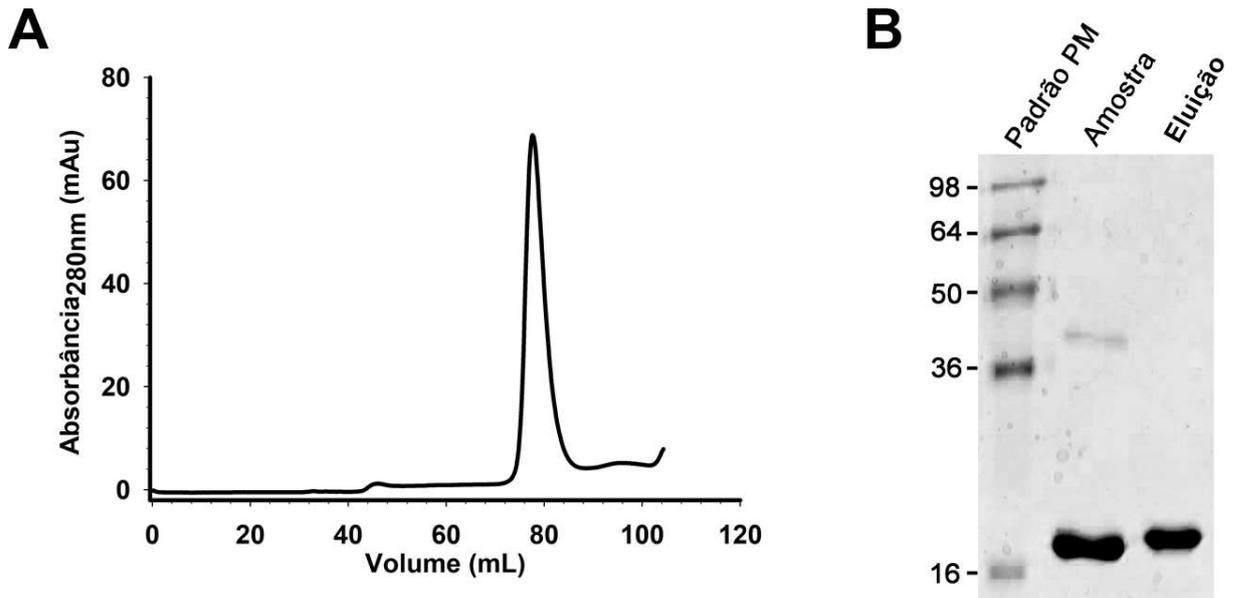


Figura 14: Cromatografia de exclusão molecular em coluna Superdex 75 PG HiLoad 16/60 (GE Healthcare) de 120 mL ao fluxo de 1 mL/min e gel de eletroforese SDS PAGE em sistema Mini-PROTEAN Tetra System 1 mm (Bio-Rad). do domínio PWWP-NSD3s. (A) Cromatograma da exclusão molecular do domínio PWWP-Pdp3 mostrando um pico inicial (0 - 10 mL) de estabilização da linha de base, um pico de maior intensidade (75 - 90 mL) dentro do volume de eluição esperado para o domínio PWWP-Pdp3 e ondulações menores ao final do cromatograma, referentes à eliminação de componentes menores presentes na solução tampão. (B) Gel de eletroforese SDS PAGE 12%. Da esquerda para a direita: padrão de peso molecular, amostra antes da injeção, proteína de interesse (17,4 kDa).

Não foram observadas alterações entre os processos de purificação do domínio PWWP-Pdp3 e PWWP-Pdp3(P20W), assim como não houve diferenças significativas na purificação dos domínios PWWP-Pdp3 e PWWP-NSD3s, ambos purificados com sucesso por duas cromatografias de afinidade e uma cromatografia de exclusão molecular. Os processos cromatográficos utilizados são comuns em purificações de domínios PWWP conforme observado em Lukasik, S.M. et al (2005).

#### 4.4 Differential Scanning Fluorimetry (DSF)

A técnica consiste em expor a amostra a um aumento gradativo de temperatura na presença de um fluoróforo (*Sypro Orange*) que se liga ao *core* hidrofóbico, o equipamento capta a emissão no comprimento do fluoróforo e o sinal aumenta à medida que o domínio perde sua forma enovelada. O experimento tem por finalidade avaliar a estabilidade termodinâmica dos domínios PWWP-Pdp3 e PWWP-NSD3s em

seus estados livre e complexado ao peptídeo H3K36me3, já que é descrito como a modificação pós traducional de maior afinidade aos domínios PWWP.

Uma vez que o domínio quando complexado ao peptídeo é passível de provocar alterações no estado monomérico dos domínios PWWP, previamente à análise via DSF, a amostra foi submetida a uma purificação em cromatografia por exclusão molecular afim de analisar a presença ou não de tal fenômeno. Os domínios PWWP-Pdp3 livre e complexado apresentaram os mesmos volumes de eluição de 75 mL correspondente a uma massa globular de 17 kDa na cromatografia de exclusão molecular, indicando que o domínio complexado não altera o estado monomérico do domínio (Figura 15A).

O resultado do DSF indica um  $T_m$  (*melting temperature*) de 42°C para o domínio PWWP-Pdp3, ou seja, nessa temperatura mais de 50% da amostra desenovela e expõe a sonda antes complexada ao *core* hidrofóbico. O aumento gradativo do sinal de fluorescência em temperaturas maiores é plotado no gráfico e reporta a sensibilidade termodinâmica do domínio. Devido ao ponto de inflexão coincidir sobre a mesma temperatura para forma livre e complexada, conclui-se que o domínio PWWP-Pdp3 não apresenta variações termodinâmicas quando complexado ao peptídeo H3K36me3 (Figura 15B).

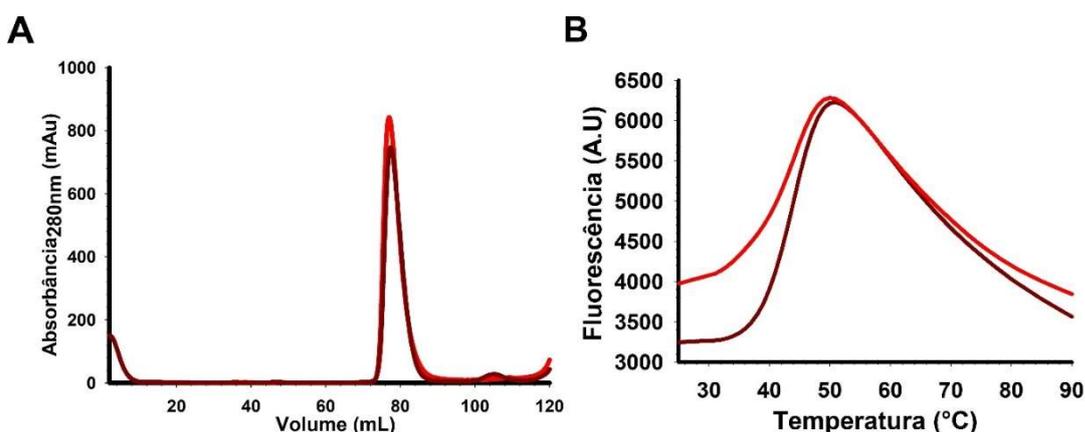


Figura 15: Cromatografia de exclusão e DSF do domínio PWWP-Pdp3 livre e complexado ao peptídeo derivado da histona H3K36me3, (A) sobreposição das cromatografias de exclusão molecular da forma livre e complexada, em vermelho claro o domínio livre e em vermelho escuro o complexado. (B) Rampa de temperatura para a o domínio PWWP-Pdp3 livre (vermelho claro) e complexado (vermelho escuro).

O domínio PWWP-NSD3s foi purificado através da cromatografia de exclusão molecular em sua forma livre e complexada, ambos os cromatogramas estão sobrepostos (Figura 16A) e apresentaram volumes de eluição de aproximadamente 76 mL indicando que o complexo ao peptídeo H3K36me3 não altera o estado monomérico do domínio PWWP-NSD3s.

Os espectros de DSF do domínio PWWP-NSD3s para as formas livre e complexada também se encontram sobrepostos (Figura 16B). A forma livre possui um espectro com uma única inflexão e um TM de aproximadamente 60 °C, enquanto a forma complexada apresenta duas inflexões e um TM a 60 °C e outro a 72 °C indicando a possibilidade de o domínio estar parcialmente complexada ao peptídeo.

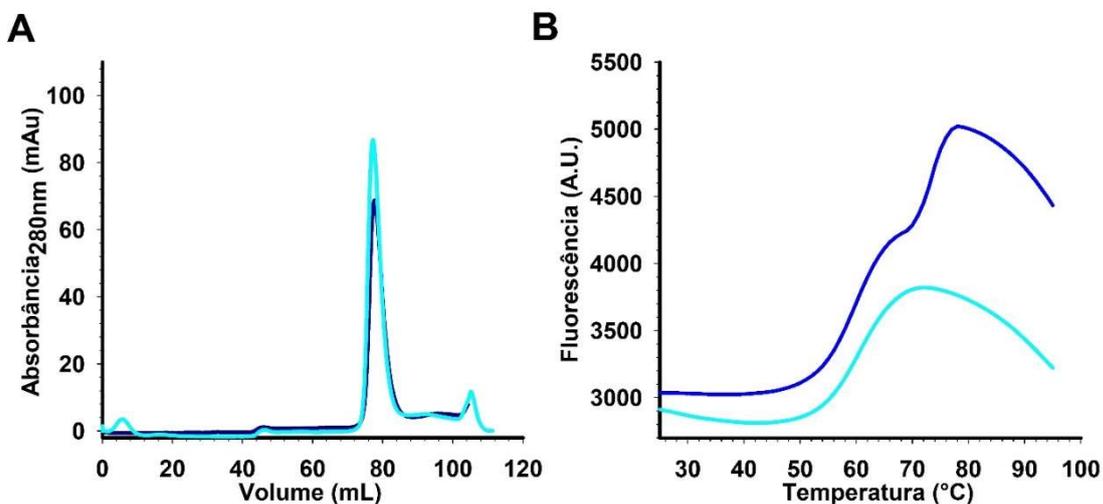


Figura 16: Cromatografia de exclusão e DSF do domínio PWWP-NSD3s livre e complexado ao peptídeo derivado da histona H3K36me3, (A) sobreposição das cromatografias de exclusão molecular da forma livre e complexada, em azul claro o domínio livre e em azul escuro o complexado. (B) Rampa de temperatura para a o domínio PWWP-Pdp3 livre (vermelho claro) e complexado (vermelho escuro).

O ensaio revela que apesar de pertencerem a proteínas ortólogas, há uma diferença de cerca de 20 °C no TM dos domínios estudados, indicando possíveis diferenças estruturais e uma maior estabilidade termodinâmica do domínio PWWP-NSD3s em relação ao PWWP-Pdp3.

#### 4.5 Dicroísmo Circular

O dicroísmo circular foi utilizado com o intuito de comparar o perfil de estrutura secundária, e a estabilidade termodinâmica dos domínios, porém sem a necessidade de utilizar uma sonda extrínseca como no DSF. Analisou-se também, através dessa

técnica, um mutante P20W do domínio PWWP-Pdp3, visto que a posição 20 para os demais domínios PWWP encontram-se um triptofano ou uma histidina, e o domínio PWWP-Pdp3 é o único domínio no qual a posição 20 é ocupada por uma prolina.

Os espectros com variação do comprimento de onda de 200 a 260 nm estão sobrepostos e em ambos os domínios indicam um perfil de contribuição tanto de  $\alpha$ -hélices quanto de fitas  $\beta$ , o que condiz com o perfil de estrutura dos domínios PWWP em geral (Figura 17A).

Mesmo assim existem diferenças significativas no espectro PWWP-NSD3s em relação ao PWWP-Pdp3, O PWWP-NSD3s, de maneira geral, desvia menos a luz circularmente polarizada do que os domínios PWWP-Pdp3 (WT) e mutante, ou seja, observa-se espectros com maiores amplitudes de elipticidade para o PWWP-Pdp3 (WT). Além disso, há uma elevação próximo 230 nm presente no espectro de PWWP-NSD3s e ausente nos demais, tal sinal indica contribuição aromática dos triptofanos (ANDERSON, D. *et al* 2001).

Ambos os domínios apresentam o pico negativo em 208 nm como o ponto mínimo de seus respectivos espectros, por isso as rampas de temperatura foram realizadas fixando o comprimento de onda em 208 nm (Figura 17B).

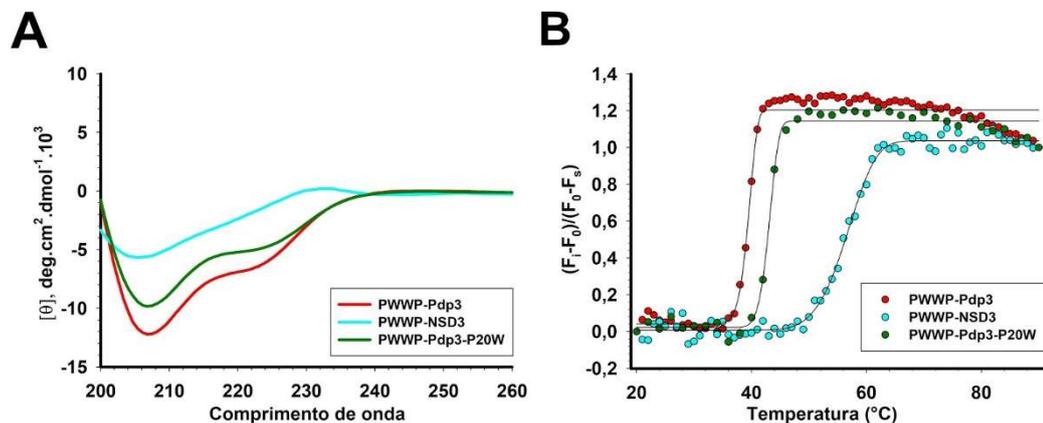


Figura 17: Dicroísmo circular dos domínios PWWP-Pdp3, PWWP-NSD3s e PWWP-Pdp3-P20W. (A) sobreposição dos espectros de CD dos domínios PWWP-Pdp3 (vermelho), PWWP-NSD3s (azul) e PWWP-Pdp3-P20W (verde). (B) Sobreposição das rampas de temperatura dos domínios PWWP-Pdp3 (vermelho), PWWP-NSD3s (azul) e PWWP-Pdp3-P20W (verde).

Avaliou-se também se entre as concentrações de 15  $\mu$ M a 35  $\mu$ M os domínios apresentavam curvas de TM dependentes da concentração, e observou-se que independente das concentrações testadas nenhum dos domínios variou os TM's de

38° C, 42° C e 56° C para respectivamente os domínios PWWP-Pdp3 (WT), PWWP-Pdp3 (P20W) e PWWP-NSD3 nas diferentes concentrações, indicando que a estabilidade termodinâmica não é dependente para as concentrações analisadas (Figura 18).

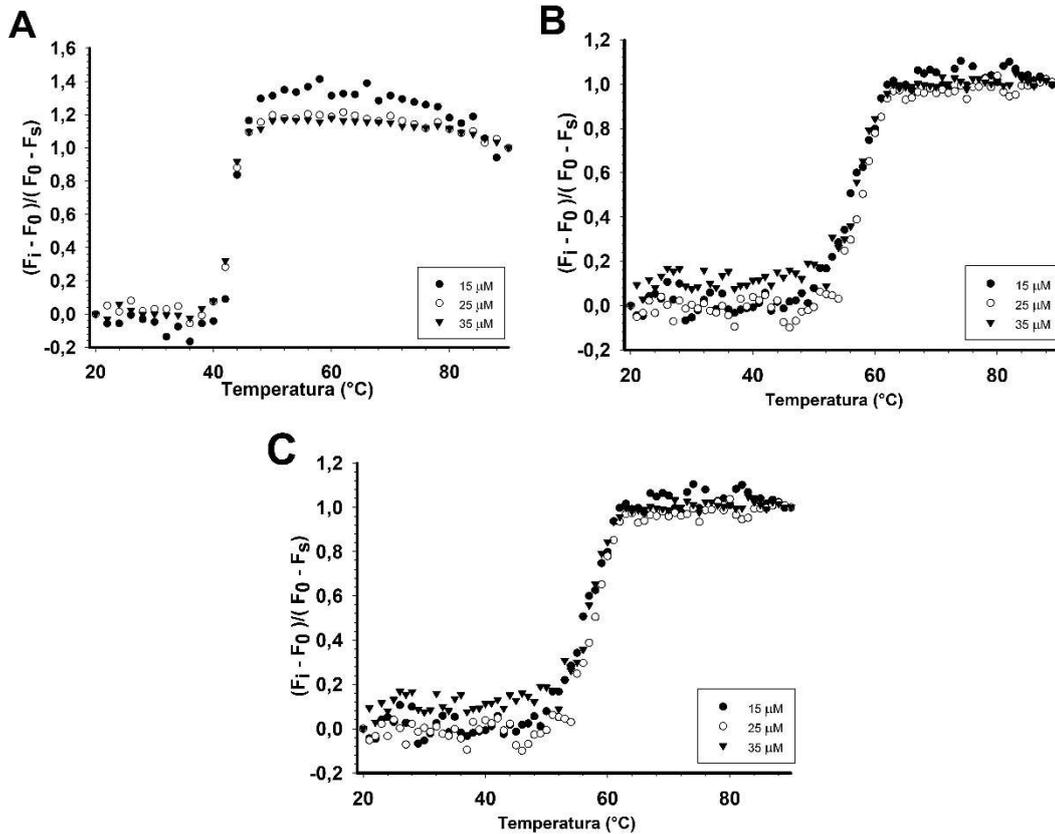


Figura 18: Rampas de temperatura dos domínios PWWP-Pdp3, PWWP-NSD3s e PWWP-Pdp3-P20W a 15, 25 e 35  $\mu\text{M}$ . (A) Sobreposição das rampas de temperatura a 15, 25 e a 35  $\mu\text{M}$  do domínio PWWP-Pdp3. (B) Sobreposição das rampas de temperatura a 15, 25 e a 35  $\mu\text{M}$  do domínio PWWP-NSD3s. (C) Sobreposição das rampas de temperatura a 15, 25 e a 35  $\mu\text{M}$  do domínio PWWP-Pdp3-P20W.

#### 4.6 Espectroscopia de Fluorescência

A espectroscopia de fluorescência foi utilizada para avaliação da intensidade e da especificidade de interação dos domínios PWWP-NSD3s e PWWP-Pdp3 em relação a 9 peptídeos derivados das caudas das histonas. Inicialmente os espectros de varredura de 300 a 450 nm partindo dos espectros dos domínios livres apresentaram máximos de emissão próximos a 340 nm indicando que os triptofanos dos domínios se encontram parcialmente expostos ao solvente (Figura 19A e Figura 20A).

Como os triptofanos dos domínios PWWP concentram-se no sitio interação com as caudas das histonas, escolheu-se os triptofanos como sonda intrínseca, e os peptídeos derivados das histonas foram titulados sobre os domínios PWWP, já os dados plotados são referentes a supressão do sinal de emissão do triptofano em relação concentração de peptídeo titulada (Figura 19B e Figura 20B).

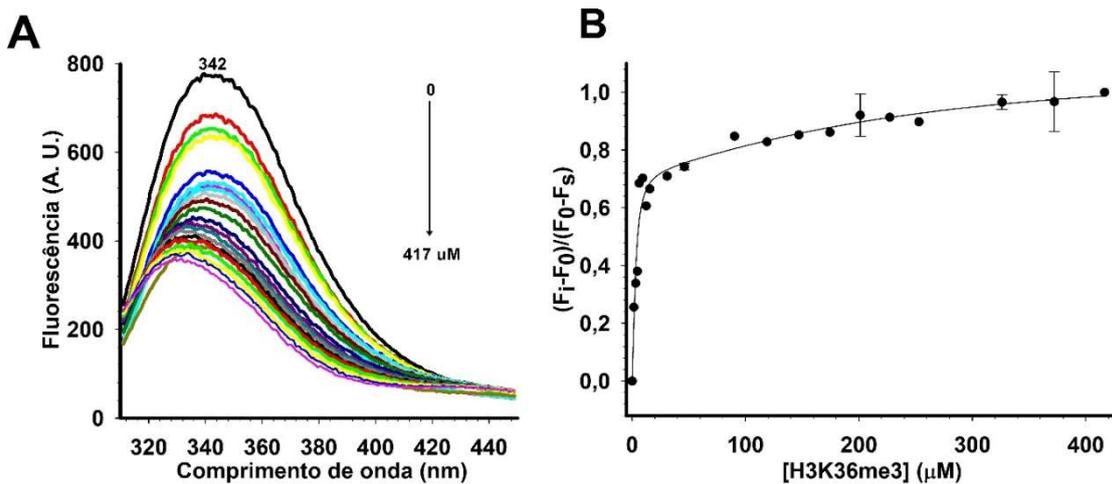


Figura 19: Titulação dos peptídeos sobre o domínio PWWP-Pdp3 na espectroscopia de fluorescência. (A) Supressão do sinal de emissão dos triptofanos a medida que o peptídeo H3K36me3 foi titulado. (B) *Fitting* da curva referente a intensidade total de cada ponto da titulação.

Para o domínio PWWP-Pdp3 os peptídeos foram titulados até 417  $\mu\text{M}$  (Figura 19A) e os dados foram fitados através de uma equação com duas componentes, e com isso, obteve-se duas constantes de dissociação para cada titulação conforme observado na tabela 5. Entretanto não existe descrito na literatura mais de um sitio de ligação entre PWWP e cauda das histonas em ensaios *in vivo*, e há especificidade entre diferentes posições nas caudas das histonas, logo o tamanho dos peptídeos analisado pode ser um diferencial nos resultados obtidos.

É possível concluir que a especificidade entre diferentes sítios de ligações nas caudas das histonas não aconteça exclusivamente no sitio de interação com os domínios, indicando a possibilidade de toda a cauda desenovelada das histonas desempenharem um papel fundamental na especificidade dessas proteínas *in vivo*.

Tabela 5: Constante de dissociação do domínio PWWP-Pdp3 em relação aos peptídeos derivados das histonas titulados.

Peptídeo	Kd 1(μM)	Kd 2 (μM)
H3K4me3	187 ± 500	0,2 ± 1
H3K9me3	-	26 ± 10
H3K27me3	-	31 ± 18
H3K36	5352 ± 8574	6 ± 5
H3K36me	2861 ± 3263	2,5 ± 0,7
H3K36me2	4249 ± 4684	12 ± 6
H3K36me3	2823 ± 4003	4 ± 2
H3K79me3	-	4 ± 3
H4K20me3	-	-

O domínio PWWP-NSD3s apresentou uma maior intensidade do sinal de emissão dos triptofanos, conseqüentemente necessitou-se de menores concentrações da amostra para realização dos experimentos, então foi possível realizar a titulação dos peptídeos até a concentração de 270 μM. O domínio apresentou um maximo de emissão em 343 nm indicando que os triptofanos também se encontram expostos ao solvente(Figura 20A).

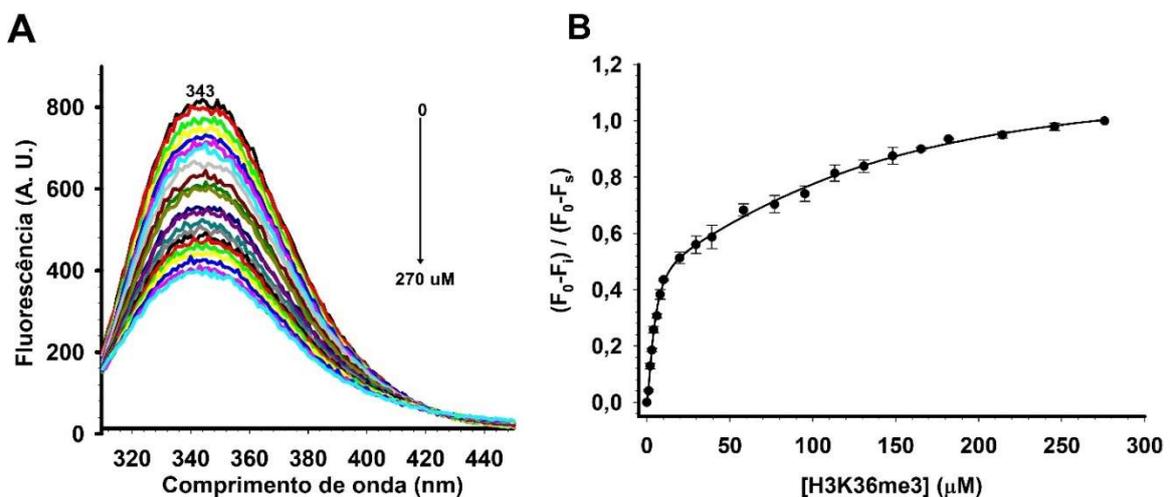


Figura 20: Titulação dos peptídeos sobre o domínio PWWP-NSD3s na espectroscopia de fluorescência. (A) Supressão do sinal de emissão dos triptofanos a medida que o peptídeo H3K36me3 foi titulado. (B) *Fitting* da curva referente a intensidade total de cada ponto da titulação.

O domínio PWWP-Pdp3 forma aglomerados visíveis a olho nu entre dois a 5 dias após a purificação do domínio, como os experimentos necessitam de mais de um dia para serem realizados, a baixa estabilidade termodinâmica do domínio PWWP-Pdp3 pode atrapalhar a reprodutibilidade da amostra, com isso observa-se que as titulações do domínio PWWP-NSD3s oscilam menos e apresentam menores valores para os erros entre as triplicatas, embora manteve-se o padrão de afinidade da ordem de micromolar e seletividade insignificantes entre os diferentes peptídeos titulados (Tabela 6).

Tabela 6: Constante de dissociação do domínio PWWP-NSD3s em relação aos peptídeos derivados das histonas

Peptídeo	Kd 1	Kd 2
H3K4me3	473 ± 149	3,8 ± 0,5
H3K9me3	-	6,9 ± 0,9
H3K27me3	1265 ± 329	5,6 ± 0,6
H3K36	1512 ± 615	7 ± 1
H3K36me	577 ± 243	4,9 ± 0,7
H3K36me2	1901 ± 1117	9 ± 2
H3K36me3	1871 ± 663	10 ± 1
H3K79me3	2358 ± 1174	8 ± 1
H4K20me3	1597 ± 725	5 ± 1

A interação dos domínios PWWP analisados com os peptídeos derivados das histonas concordaram quanto a alta afinidade e baixa seletividade, porém isso não significa que a nível atômico as ligações entre os resíduos aromáticos presentes no barril  $\beta$  e as lisinas metiladas das caudas das histonas sejam idênticas. Analisando a supressão do sinal de emissão dos triptofanos foi possível perceber que o domínio PWWP-Pdp3 desloca o máximo de emissão para comprimentos de onda menores, quando titulados os peptídeos H3K36 com diferentes graus de metilações, fato não observado para o domínio PWWP-NSD3s.

O deslocamento no máximo de emissão indica uma possível mudança de conformação dos resíduos de aminoácidos que participam da ligação à medida que interagem com lisinas metiladas, e que essa mudança de conformação seja induzida

pelas metilas presentes na lisina, visto que o deslocamento é maior para lisinas trimetiladas (Figura 21).

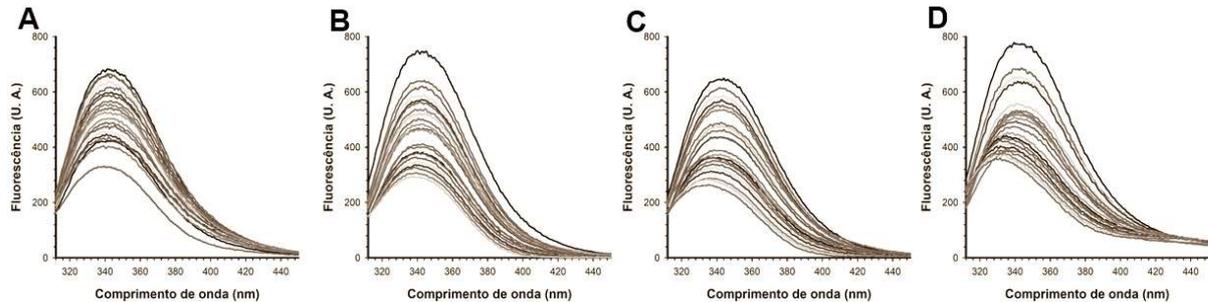


Figura 21: Supressão de emissão dos triptofanos nas interações entre PWWP-Pdp3 e H3K36. (A) Titulação por espectroscopia de fluorescência entre PWWP-Pdp3 e H3K36. (B) Titulação por espectroscopia de fluorescência entre PWWP-Pdp3 e H3K36me1. (C) Titulação por espectroscopia de fluorescência entre PWWP-Pdp3 e H3K36me2. (D) Titulação por espectroscopia de fluorescência entre PWWP-Pdp3 e H3K36me3.

#### 4.7 Electrophoretic Mobility Shift Assay (EMSA)

Domínios PWWP podem ou não ligar a DNA concomitante com ligações com as caudas das histonas, e as ligações de domínios PWWP com DNA são descritas como eletrostáticas, logo a análise das interações com DNA reportam acerca da carga de superfície desses domínios. Visto que domínios com regiões positivas nas superfícies são mais propensos à interação com o DNA que são carregados negativamente.

O ensaio de mudança de mobilidade eletroforética foi utilizado com o objetivo de caracterizar a capacidade de interação dos domínios PWWP-NSD3s e PWWP-Pdp3 com sequencias duplas fitas de DNA com diferentes estruturas. O desenho experimental foi adaptado do modelo utilizado por Laguri, *et al.* (2015).

O domínio PWWP-NSD3s apresenta mudança de mobilidade crescente em estequiometrias superiores a 1:1, indicando que há mais de um sitio no DNA onde o domínio se liga. Além disso, o padrão de migração não muda independente das modificações estruturais do DNA, logo o domínio PWWP-NSD3s se liga ao DNA independente das modificações presentes na estrutura de DNA analisadas (Figura 22).

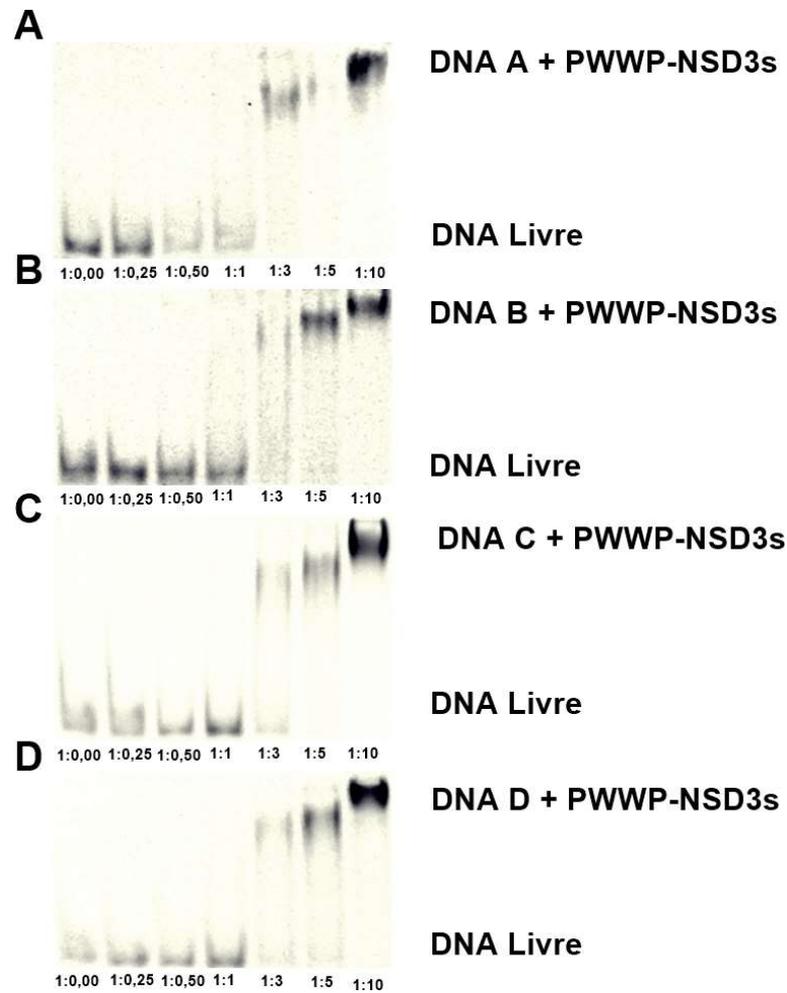


Figura 22: Géis de EMSA da interação do domínio PWWP-NSD3s com DNA. (A) Gel de EMSA do domínio PWWP-NSD3s com o DNA A. (B) Gel de EMSA do domínio PWWP-NSD3s com o DNA B. (C) Gel de EMSA do domínio PWWP-NSD3s com o DNA C. (D) Gel de EMSA do domínio PWWP-NSD3s com o DNA D.

O domínio PWWP-Pdp3 apresenta um padrão de migração com maior arraste em comparação com o domínio PWWP-NSD3s, indicando que possivelmente o complexo PWWP-Pdp3/DNA é menos instável que o complexo PWWP-NSD3s/DNA. No entanto, na razão molar de 1:10 a migração do complexo PWWP-Pdp3/DNA é menor do que o complexo PWWP-NSD3s/DNA; logo apesar de mais instável, o complexo PWWP-Pdp3/DNA é maior em excesso de proteína em relação a DNA.

Outra diferença é a variação da quantidade de complexo formado em relação a diferentes estruturas de DNA, o domínio PWWP-Pdp3 varia as razões molares dos complexos quando o DNA apresenta mudanças estruturais (Figura 23), o que não foi observado para o domínio PWWP-NSD3s (Figura 22). O complexo PWWP-Pdp3/DNA dupla fita completamente anelado (Figura 23A) só pode ser observado em razões molares superiores a 1:5.

O complexo PWWP-Pdp3/DNA não anelado ao centro (Figura 23B) e o complexo não anelado em uma das extremidades (Figura 23C) são visíveis a partir de razões molares de 1:3. Curiosamente o DNA completamente anelado, porém sem uma das ligações fosfodiester demonstrou complexos formados a partir da estequiometria de 1:1 (Figura 23D). Além disso, nas concentrações testadas não foram observadas saturações do DNA livre, indicando que a afinidade do domínio PWWP-Pdp3 pelo DNA é menor do que do domínio PWWP-NSD3s independente das modificações estruturais do DNA.

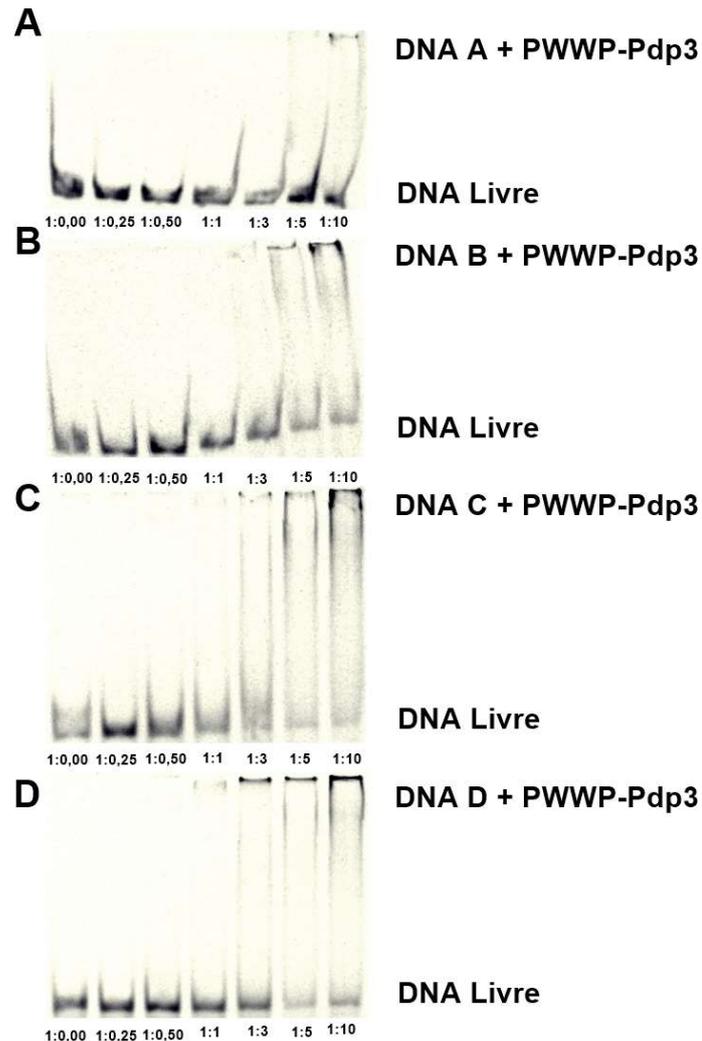


Figura 23: Géis de SDS PAGE da interação do domínio PWWP-Pdp3 com DNA. (A) Gel de EMSA do domínio PWWP-Pdp3 com o DNA A. (B) Gel de EMSA do domínio PWWP-Pdp3 com o DNA B. (C) Gel de EMSA do domínio PWWP-Pdp3 com o DNA C. (D) Gel de EMSA do domínio PWWP-Pdp3 com o DNA D.

Comparando a superfície do modelo estrutural do domínio PWWP-Pdp3 e da estrutura do domínio PWWP-NSD3s é possível perceber que a concentração das cargas positivas nas superfícies dos domínios não são sobreponíveis, então, provavelmente, os domínios não conservam a mesma região de interação com DNA, visto que a interação PWWP-DNA é majoritariamente eletrostática (Figuras 24;25).

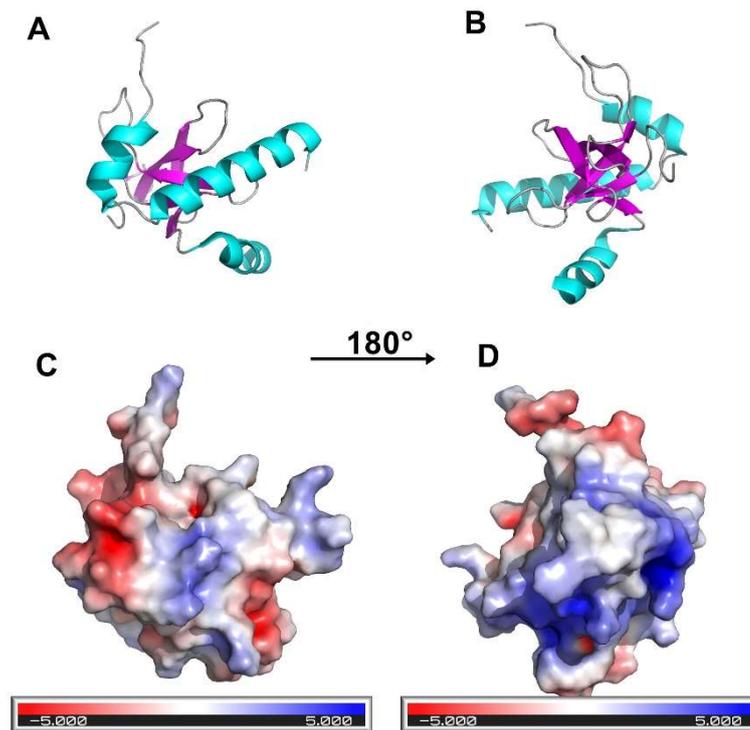


Figura 24: Superfície eletrostática do domínio PWWP-NSD3s. (A) Estrutura do domínio PWWP-NSD3s (PDB: 6G3T). (B) rotação em 180° da estrutura representada em A. (C) carga de superfície da estrutura representada em A. (D) carga de superfície da estrutura representada em B.

Devido a ligação dos domínios PWWP com o DNA ser caracterizada como majoritariamente de natureza eletrostática, a análise das cargas de superfícies dos domínios PWWP-NSD3s e PWWP-Pdp3 podem ajudar a entender como esses domínios se comportam na presença do DNA.

O domínio PWWP-NSD3s concentra as cargas positivas de superfície em regiões adjacentes ao barril  $\beta$  no centro do domínio indicando que provavelmente esse tipo de interação ocorra sem influência da dinâmica do domínio em solução (Figura 24). Já o domínio PWWP-Pdp3 concentra as cargas positivas na extremidade C-terminal desenovelada indicando que provavelmente a ligação entre o PWWP-Pdp3 e o DNA dependa da dinâmica da proteína em solução devido a maior flexibilidade dessa região de ligação (Figura 25).

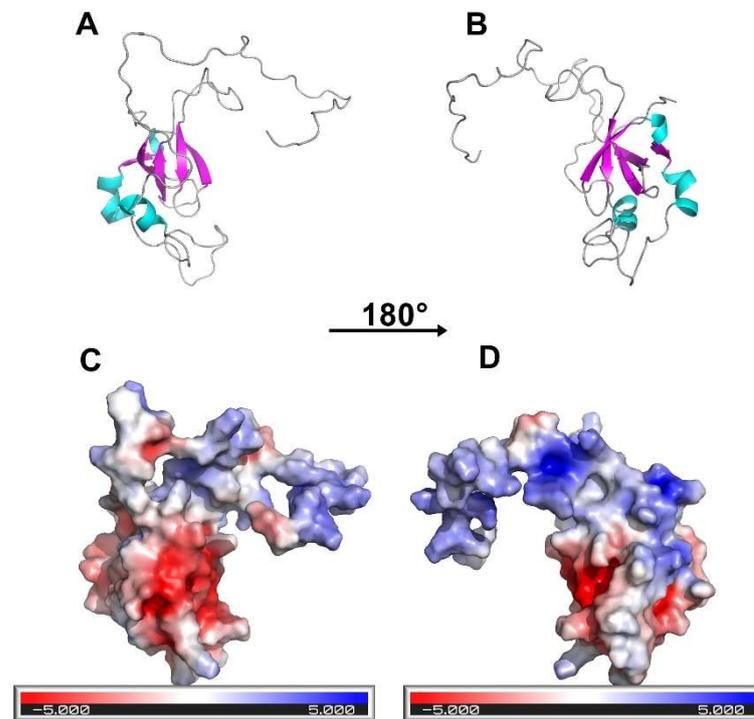


Figura 25: Superfície eletrostática do domínio PWWP-Pdp3. (A) Modelo estrutural do domínio PWWP-Pdp3. (B) rotação em 180° do modelo estrutural representado em A. (C) carga de superfície do modelo estrutural representado em A. (D) carga de superfície do modelo estrutural representado em B.

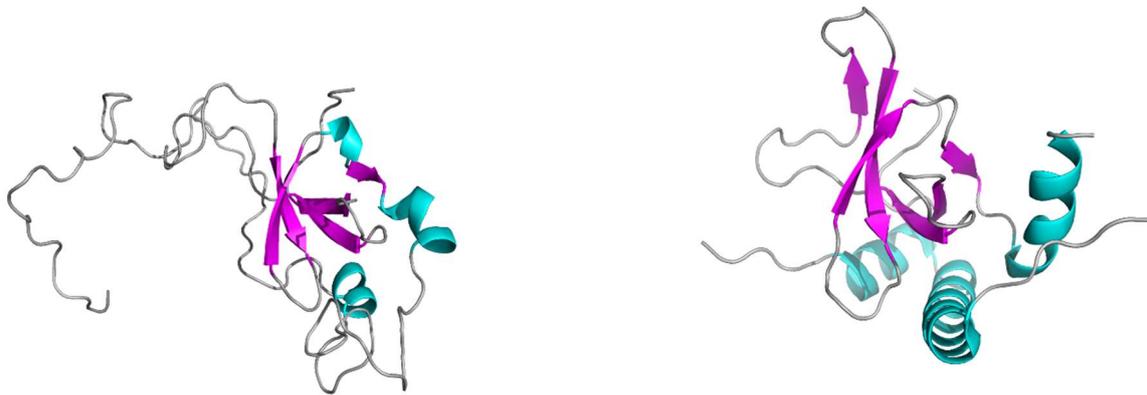
## 5 DISCUSSÃO

Do ponto de vista estrutural o modelo obtido para o domínio PWWP-Pdp3 conserva características estruturais do domínio PWWP-NSD3s, ambos apresentam de 5 a 6 fitas  $\beta$  formando um barril  $\beta$  na região N-terminal e 3  $\alpha$  hélices na região C-terminal. Entretanto os domínios diferem-se no percentual de estrutura enovelada, visto que o domínio PWWP-Pdp3 possui em média 6 resíduos de aminoácidos por  $\alpha$  hélices enquanto o domínio PWWP-NSD3s 13 resíduos de aminoácidos por  $\alpha$  hélices.

Outras diferenças evidentes entre os domínios ocorrem entre as fitas  $\beta$ 2 e  $\beta$ 3, onde no domínio PWWP-Pdp3 observa-se um loop no domínio PWWP-NSD3s observa-se uma fita  $\beta$  a mais. Além disso, a cauda C-terminal desenovelada do domínio PWWP-Pdp3 não é observada para o domínio PWWP-NSD3s conforme observado na figura a seguir (figura 26).

Já havia sido descrito que construções inferiores aos 150 resíduos para o domínio PWWP-Pdp3 não são capazes de interagir com as caudas das histonas (GILBERT; *et al.*, 2014). Portanto apesar do sitio de interação com histonas ser a

gaiola aromática no centro do barril  $\beta$  da região N-terminal, a cauda desordenada presente na região C-terminal possivelmente participa dessa interação.



**Modelo E Pdp3-PWWP**

**NSD3s-PWWP PDB: 6g3t**

Figura 26: Comparação estrutural entre o modelo do domínio PWWP-Pdp3 e a estrutura cristalina do domínio PWWP-NSD3s.

Tais diferenças estruturais suportam os resultados de estabilidade termodinâmica que indicam uma diferença de aproximadamente 20°C no TM dos domínios, e provavelmente a menor reprodutibilidade nos experimentos do domínio PWWP-Pdp3.

A primeira prolina do motivo de sequência dos domínios PWWP já foi descrita como uma posição crucial para a estabilidade desses domínios, visto que esse resíduo está localizado no loop entre as fitas  $\beta 1$  e  $\beta 2$  e promove a flexibilidade necessária para a interação do loop com as  $\alpha$  hélices que suportam o barril  $\beta$  (HUNG; *et al.*, 2015). Curiosamente, o segundo resíduo de aminoácido do PWWP pode ser um triptofano ou uma histidina, esse resíduo está presente no loop  $\beta 1$ - $\beta 2$  e adjacente a fita  $\beta 2$ . O domínio PWWP-Pdp3 difere dos demais por apresentar uma prolina nessa posição, logo a ausência de um resíduo de aminoácido com uma cadeia lateral que interaja com as  $\alpha$  hélices pode contribuir para a menor estabilidade do domínio (figura 27).

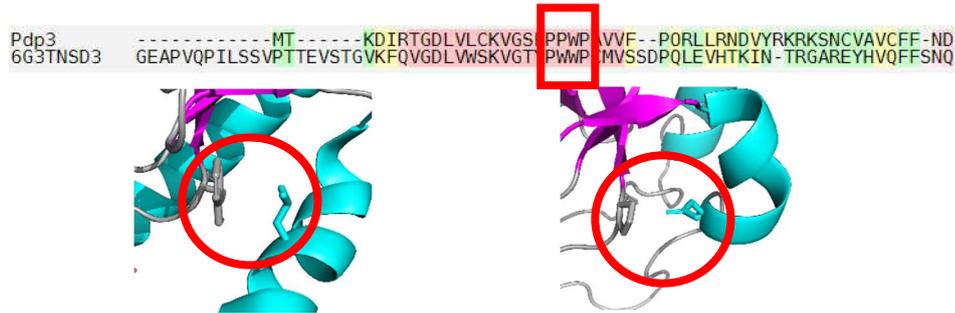


Figura 27: Alinhamento de sequência da região N-terminal dos domínios PWWP-Pdp3 e PWWP-NSD3s ressaltando a diferença dentro do motivo de sequência PWWP entre os dois domínios. A baixo e a esquerda destacou-se a proximidade entre a cadeia lateral do triptofano presente no loop  $\beta$ 1- $\beta$ 2 com a  $\alpha$  hélice no domínio PWWP-NSD3s. A baixo e a direita destacas-se a proximidade entre a prolina na posição 20 com a  $\alpha$  hélice do modelo PWWP-Pdp3.

Do ponto de vista da interação com ligantes, os resultados obtidos por espectroscopia de fluorescência mostram ausência de uma seletividade dos domínios PWWP-Pdp3 e PWWP-NSD3s pelos peptídeos derivados das histonas. Sugere-se a possibilidade da proteína íntegra possuir outros sítios de interação com as histonas. Visto que estudos anteriores demonstraram a especificidade dos domínios tipo PWWP pela posição H3K36me3 através de ensaios de imunoprecipitação (GILBERT *et al.*, 2014) ou na presença do nucleossomo íntegro (RONDELET; *et al.*, 2016).

Sabe-se que domínios PWWP interagem simultaneamente às histonas e ao DNA, de modo que essas interações cooperam para manter a estabilidade do PWWP à cromatina (DUKATZ; *et al.*, 2019). Tanto o PWWP-Pdp3 quanto PWWP-NSD3s mostraram-se capazes de ligar a oligonucleotídeos de DNA independente da estrutura adotada pelo mesmo.

Apesar de ambos os domínios concentrarem mais resíduos positivos na região C-terminal, os mesmos diferencem quanto a afinidade de ligação, provavelmente o fato da cauda C-terminal do domínio PWWP-Pdp3 não ser enovelada diminui a interação com o DNA.

## 6 CONCLUSÃO

A modelagem comparativa de um possível modelo que melhor representa a estrutura do domínio PWWP-Pdp3 foi capaz de chegar em um único modelo que melhor representaria, ainda que os valores encontrados na validação sejam inferiores

aos moldes utilizados, cabe ressaltar que o domínio apresenta regiões desordenadas entre as fitas  $\beta 2$ - $\beta 3$  e no C-terminal que dificultam a validação.

Os domínios PWWP-Pdp3 e PWWP-NSD3s foram expressados e purificados com sucesso, e possibilitaram a realização dos ensaios em vitro para comparar os domínios quanto a estabilidade termodinâmica e interações com ligantes.

A estabilidade termodinâmica foi um dos pontos onde os domínios PWWP-Pdp3 e PWWP-NSD3s destoam, os domínios foram analisados através das técnicas de fluorimetria diferencial de varredura (DSF) e dicroísmo circular (CD). Através do DSF foi possível perceber que além de mais estável a forma complexada do domínio PWWP-NSD3s aumenta ainda mais sua estabilidade na presença do peptídeo derivado da cauda da histona, já o CD foi capaz de reportar mudanças sutis no perfil de estrutura secundária os dados das duas técnicas convergem para que a diferença do TM entre os domínios seja de aproximadamente 20 °C.

O mutante PWWP-Pdp3-P20W aumentou a estabilidade do domínio PWWP-Pdp3, logo a falta de um resíduo de aminoácido com cadeia lateral, na posição 20, é um dos fatores que diminuem a estabilidade do domínio PWWP-Pdp3.

A interação com ligantes foi analisada através da espectroscopia de fluorescência para avaliar a intensidade e especificidade aos peptídeos derivados das caudas das histonas, e concluiu-se que a especificidade da ligação com as histonas não pode ser determinada somente avaliando o sitio de ligação, visto peptídeos de 13 resíduos de aminoácidos não são capazes de modificarem o kD dos domínios.

A espectroscopia de fluorescência também mostrou que o domínio PWWP-Pdp3 blinda os triptofanos à medida que se aumenta o grau de metilação do peptídeo H3K36.

Os ensaios de EMSA permitiram diferenciar os domínios quanto a interação com DNA, o domínio PWWP-NSD3s tem maior afinidade e maior deslocamento do complexo PWWP/DNA e não é sensível a modificações estruturais do DNA, já o domínio PWWP-Pdp3 além de interagir menos, varia o perfil de corrida do EMSA com DNA de diferentes estruturas, mas forma complexos maiores e com maior variedade de tamanho.

## REFERÊNCIAS

- ALBERT, M.; HELIN, K. Histone methyltransferases in cancer. **Cell**, v. 21, n. 2, p. 209-220, 2010.
- ANGRAND, P. O.; APIOU, F.; STEWART, A. F.; DUTRILLAUX, B.; LOSSON, R.; CHAMBON, P. NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. **Genomics**, v. 74, n. 1, p. 79-88, 2001.
- BANNISTER, A. J.; KOUZARIDES, T. Regulation of chromatin by stone modifications. **Cell**, v. 21, n. 3, p. 381-395, 2011.
- BROWNELL, J. E.; ZHOU, J.; RANALLI, T.; KOBAYASHI, R.; EDMONDSON, D. G. ROTH, S. Y.; ALLIS, C. D. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. **Cell**, v. 84, n. 6, p. 843-851, 1996.
- CORBETT, A. H. Post-transcriptional Regulation of Gene Expression and Human Disease. **Cell**, v. 52, p. 96-104, 2018.
- DASTIDAR, R. G.; HOODA, J.; SHAH, A.; CAO, T. M.; HENKE, R. M.; ZHANG, L. The nuclear localization of SWI/SNF proteins is subjected to oxygen regulation. **Cell and Bioscience**, v. 2, n. 1, 2012.
- DE LA CRUZ, X.; LOIS, S.; SÁNCHEZ-MOLINA, S.; MARTÍNEZ-BALBÁS, M. A. Do protein motifs read the histone code? **BioEssays**, v. 27, n. 2, p. 164-175, 2015.
- ANDERSSON, D.; CARLSSON, U.; FRESKGAARD, P. O. Contribution of tryptophan residues to the CD spectrum of the extracellular domain of human tissue factor Application in folding studies and prediction of secondary structure. **FEBS**. v. 268, p. 1118-1128, 2001.
- DUKATZ, M.; HOLZER, K.; CHOUDALAKIS M.; EMPERLE, M. LUNGU, C.; BASHTRYKOV, P.; JELTSCH, A.; H3K36me2/3 binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. **Molecular Biology**, 2019.
- EIDAH, J. O.; CROWE, B. L.; NORTH, J. A.; MCKEE, C. J.; SHKRIABAI, N.; FENG, L.; PLUMB, M.; GRAHAM, R. L.; GORELICK, R. J.; HESS, S.; POIRIER, M. G.; FOSTER, M. P.; KVARATSKHELIA, M. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. **Nucleic Acids Research**, v. 41, n. 6, p. 3924–3936, 2013.
- EGGER, G.; LIANG, G.; APARICIO, A.; JONES, P. A. Epigenetics in human disease and prospects for epigenetic therapy. **Nature**, v. 429, n. 6990, p. 457-463, 2004.

FASMAN, G.D. Circular Dichroism and the Conformational Analysis of Biomolecules, 1996.

GELSI-BOYER, V.; ORSETTI, B.; CERVERA, N.; FINETTI, P.; SIRCOULOMB, P.; ROUGÉ, C.; LASORSA, L.; LETESSIER, A.; GINESTIER, C.; MONVILLE, F.; ESTEYRIÈS, S.; ADÉLAÏDE, J.; ESTERNI, B.; HENRY, C.; ETHIER, S. P.; BIBEAU, F.; MOZZICONACCI, M.; CHARAFE-JAUFFRET, E.; JACQUEMIER, J.; BERTUCCI, F.; BIRNBAUM, D.; THEILLET, C.; ZHAFFANET, M. Comprehensive profiling of 8p11-12 amplification in breast cancer. **Molecular Cancer Research**, v. 3, n. 12, p. 655-667, 2005.

GREER, E.L.; SHI, Y. Histone methylation: a dynamic mark in health, disease and inheritance. **Nature Reviews Genetics**, v. 13, n. 5, p. 343-357, 2012.

HAN, X.; PIAO, L.; ZHUANG, Q.; YUAN, X.; LIU, Z.; HE, X. The role of histone lysine methyltransferase NSD3 in cancer. **Onco Targets Therapy**, v. 11, p. 3847-3852, 2018.

HANAHAHAN, D. J. Studies on transformation of Escherichia coli with plasmids. **Molecular Biology**, v. 166, n. 4, p. 557-580, 1983.

HAVERTY, P. M.; FRIDLAND, J.; LI, L.; GETZ, G.; BEROUKHIM, R.; LOHR, S.; WU, T. D.; CAVET, G.; ZHANG, Z.; CHANT, J. High-resolution genomic and expression analyses of copy number alterations in breast tumors. **Genes Chromosomes Cancer**, v. 47, n. 6, p. 530-542, 2008.

HE, C.; LI, F.; ZHANG, J.; WU, J.; SHI, Y. The Methyltransferase NSD3 Has Chromatin-binding Motifs, PHD5-C5HCH, That Are Distinct from Other NSD (Nuclear Receptor SET Domain) Family Members in Their Histone H3 Recognition. **Journal of Biological Chemistry**, v. 288, n. 7, p. 4692–4703, 2012.

HUANG, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. **Chemical Reviews**, v. 115, n. 6, p. 2376–2418, 2015.

IIZUKA, M.; SMITH, M.M. Functional consequences of histone modifications. **Current Opinion in Genetics & Development**, v.13, n. 2, p.154-160, 2003.

IZZO, A.; SCHNEIDER, R. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. **Biochimica et Biophysica Acta**, v. 1859, n. 3, p. 486-495, 2016.

KONBERG, R. D.; LORCH, Y. Chromatin-modifying and remodelling complexes. **Current Opinion Genetics & Development**, v. 9, n. 2, p. 148–151, 1999.

LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v.15, n. 227, p. 680–685, 1970.

LEITGEB S.; STRAGANZ G.D.; NIDETZKY B. Biochemical characterization and mutational analysis of the mononuclear non-haem Fe<sup>2+</sup> site in Dke1, a cupin-type

dioxygenase from *Acinetobacter johnsonii*. **Biochemistry**, v. 418, n. 2, p. 403-411, 2009.

LOPEZ, J.; PERCHARDE, M.; COLEY, H. M.; WEBB, A.; CROOK, T. The context and potential of epigenetics in oncology. **British Journal of Cancer**, v. 100, n. 4, p. 571-577, 2009.

LUKASIK, S. M.; CIERPICKI, T.; BORLOZ, M.; GREMBECKA, J.; EVERETT, A.; BUSHWELLER, J. H. High resolution structure of the HDGF PWWP domain: A potential DNA binding domain. **Protein Science**, v. 15, n. 2, p. 314–323, 2006.

MARGUERON, R.; TROJER, P.; REINBERG, D. The key to development: Interpreting the histone code? **Current Opinion Genetics & Development**, v.15, n. 2, p.163-176, 2005.

MELLOR, J. Dynamic nucleosomes and gene transcription. **Cell**, v.22, n. 6, p. 320-329, 2006.

PLACEK, B. J.; GLOSS, L. M. The N-Terminal Tails of the H2A–H2B Histones Affect Dimer Structure and Stability. **Biochemistry**, v. 41, n. 50, p. 14960–14968, 2002.

RAZIN, A.; RIGGS, A. D. DNA methylation and gene function. **Science**, v. 210, p. 604–610, 1980.

RONA, G. B.; ELEUTHERIO, E. C. A.; PINHEIRO, A. S. PWWP domains and their modes of sensing DNA and histone methylated lysines. **Biophysical Reviews**, v. 8, p. 63-74, 2016.

RONA, G. B.; ALMEIDA, D. S. G.; PINHEIRO, A. S.; ELEUTHERIO, E. C. A. The PWWP domain of the human oncogene WHSC1L1/NSD3 induces a metabolic shift toward fermentation. **Oncotarget**, v. 12, n. 33, p. 54068-54081, 2017.

RONA, G. B. ALMEIDA, N. P.; SANTOS, G. C. J.; FIDALGO, T. K.; ALMEIDA, F. C.; ELEUTHERIO, E. C.; PINHEIRO, A. S. <sup>1</sup>H NMR metabolomics reveals increased glutaminolysis upon overexpression of NSD3s or Pdp3 in *Saccharomyces cerevisiae*. **Cellular Biochemistry**, v. 120, n. 4, p. 5377-5385, 2018.

ROSATI, R.; STARZA, R. L.; VERONESE, A.; AVENTIN, A.; SCHWIENBACHER, C.; VALLESPI, T.; NEGRINI, M.; MARTELLI, M. F.; MECUCCI, C. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t (8;11) (p11.2;p15). **Blood**, v. 99, n. 10, p. 3857-3860, 2002.

SANTOS-ROSA, H.; CALDAS, C. Chromatin modifier enzymes, the histone code and cancer. **European Journal of Cancer**, v. 41, n.16, p. 2381-2402, 2005.

SHAHID, Z.; SIMPSON, B.; KATHLEEN, H.; SINGH, G. Genetics, Histone Code. StatPearls, Sep 2, 2020.

STEC, I.; NAGL, S.B.; OMMEN, G. J. V.; DUNNEN, J. T. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? **FEBS Letters**, v. 473, n. 1, p. 1-5, 2000.

STRAHL, B. D.; ALLIS, C. D. The language of covalent histone modifications. **Nature**, v. 403, n. 6765, p. 41-45, 2000.

SUEOKA, T.; KOYAMA, K.; HAYASHI, G.; OKAMOTO, A. Chemistry-driven Epigenetic Investigation of Histone and DNA Modifications. **The Chemical Record**, v.18, n. 12, p. 1727-1744, 2018.

TSAO, H.S.; SPINELLA, S. A.; LEE, A, T.; ELMORE, D. E. Design of novel histone-derived antimicrobial peptides. **Peptides Science**, v. 30, n. 12, p. 2168-2173, 2009.

VLEUGEL, M. M.; GREIJER, A. E.; SHVARTS, A.; GROEP, P. V.; BERKEL, M. V.; AARBODEM, Y.; TINTEREN, H. V.; HARRIS, A. L.; DIEST, P. J. V.; WALL, E. V. Differential prognostic impact of hypoxia induced and diffuse HIF-1alpha expression in invasive breast cancer. **Clinical Pathology**, v. 58, n. 2, p. 172-177, 2005.

VEZZOLI, A.; BONADIES, N.; ALLEN, M. D.; FREUND, S. M. V.; SANTIVERI, C. M.; KVINLAUG, B. T.; HUNTLY, B. J. P.; GÖTTGENS, B.; BYCROFT, M. Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. **Nature Structure Molecular Biology**, v. 17, n. 5, p. 617-619, 2010.

VOUGIOUKLAKIS, T.; HAMAMOTO, R.; NAKAMURA, Y.; SALOURA, V. The NSD family of protein methyltransferases in human cancer. **Epigenomics**, v. 7, n. 5, p. 863-874, 2015.

WAGNER, E.J.; CARPENTER, P.B. Understanding the language of Lys36 methylation at histone H3. **Nature Reviews Molecular Cell Biology**, v. 13, n. 2, p. 115-126, 2012.

WANG, R.; GAO, J.; ZHANG, J.; ZHANG, X.; XU, C.; LIAO, S.; TU, X. Solution structure of TbTFIIS2-2 PWWP domain from *Trypanosoma brucei* and its binding to H4K17me3 and H3K32me3. **Biochemical Journal**, v. 476, n. 2, p. 421-431, 2019.

WEAVER, T.; MORRISON, E.; MUSSELMAN, C. Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. **Molecules**, v. 23, n. 10, p. 2614, 2018.

WU, H.; ZENG, H.; LAM, R.; TEMPEL, W.; AMAYA, M. F.; XU, C.; DOMBROVSKI, L.; QIU, W.; WANG, Y.; MIN, J. Structural and histone binding ability characterizations of human PWWP domains. **PLoS One**, v. 6, n. 6, p. 18919, 2011.

YANG, Z.Q. Liu, G.; Bollig-Fischer, A.; Giroux, C. N.; Ethier, S. P. Transforming properties of 8p11-12 amplified genes in human breast cancer. **Cancer Research**, v. 70, n. 21, p. 8487-8497, 2010.

**ANEXOS**

Tabela da composição dos tampões das purificações

	Composição
Solução A	500 mM NaCl, 50 mM TRIS, 5 mM imidazol, pH 8,0
Solução B	500 mM NaCl, 50 mM TRIS, 500 mM imidazol, pH 8,0
Solução C	2 M NaCl, 50 mM TRIS, 5 mM imidazol, pH 8,0
Solução D	500 mM NaCl, 50 mM TRIS, pH 8,0
Solução E	20 mM $\text{PO}_4^{2-}$ , 100 mM NaCl, pH 7,4
Solução F	10 mM $\text{PO}_4^{2-}$ , 50 mM NaCl, pH 7,4