

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM
BIOINFORMÁTICA

Bruno Marques Silva

**Domínio HIRAN: variações estruturais e suas implicações
na interação com ssDNA**

Belo Horizonte

2020

BRUNO MARQUES SILVA

**Domínio HIRAN: variações estruturais e suas implicações
na interação com ssDNA**

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

Orientadora: Profa. Dra. Mariana Torquato Quezado de Magalhães

Coorientadora: Dra. Lucianna Helene Silva dos Santos

Belo Horizonte

2020

043

Silva, Bruno Marques.

Domínio HIRAN: variações estruturais e suas implicações na interação com ssDNA [manuscrito] / Bruno Marques Silva. – 2020.

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

Orientadora: Profa. Dra. Mariana Torquato Quezado de Magalhães.

Coorientadora: Dra. Lucianna Helene Silva dos Santos.

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

1. Replicação do DNA. 2. DNA de Cadeia Simples. 3. Estrutura Terciária de Proteína. 4. Domínios e Motivos de Interação entre Proteínas. I. Magalhães, Mariana Torquato Quezado de. II. Santos, Lucianna Helene Silva dos. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



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

PARECER Nº 13/2020/ICB-SECCPGBIN
PROCESSO Nº 23072.237527/2020-00

FOLHA DE APROVAÇÃO

Bruno Marques Silva

"Domínio HIRAN: variações estruturais e suas implicações na interação com ssDNA"

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Dra. Mariana Torquato Quezado de Magalhaes - Orientado
ICB/UFMG

Dra. Lucianna Helene Silva dos Santos - Coorientadora
ICB/UFMG

Dra. Déborah Antunes dos Santos
FioCruz-MG

Dra. Rafaela Salgado Ferreira
UFMG

Belo Horizonte, 29 de outubro de 2020.

Documento assinado eletronicamente por **Deborah Antunes dos Santos, Usuário Externo**, em 29/10/2020, às 11:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Lucianna Helene Silva dos Santos, Usuário Externo**, em 29/10/2020, às 11:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 29/10/2020, às 11:06, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rafaela Salgado Ferreira, Professora do Magistério Superior**, em 29/10/2020, às 11:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site

[https://sei.ufmg.br/sei/controlador_externo.php?](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0338724** e o código CRC **1FC8455F**.

AGRADECIMENTOS

Primeiramente gostaria de agradecer a minha orientadora **Mariana Quezado** pela oportunidade de realizar esse trabalho, me acolher em seu laboratório e me guiar durante todo este período do mestrado. Principalmente por ter me apresentado o círculo cromático e xingar minhas imagens em preto e branco, apesar d'eu ter engordado o Chimbau.

Agradecer também a minha coorientadora **Lucianna dos Santos** que foi de extrema importância na elaboração deste trabalho, respondendo todos os meus questionamentos, que não foram poucos, e não se incomodando com as minhas ideias de experimentos, digamos, diferenciados.

Gostaria de agradecer a **Universidade Federal de Minas Gerais**, ao **programa de pós-graduação de Bioinformática**, principalmente ao **Tiago Silva** e a **Sheila Santana** por todo o suporte durante este período. Agradecer ao **Laboratório de Genética Bioquímica** e o **Laboratório de Biologia Estrutural** pelo suporte instrumentário e pessoal para a realização dos experimentos, principalmente a **Verônica**, ao **Arthur**, ao **Wesley** e a **Débora**. Agradecer também à bolsa concedida pela **FAPEMIG**, que foi de extrema importância.

Agradecer ao técnico **Jamil** pela incrível atenção e carinho com todos que passam pelo seu laboratório, não medindo esforços para ajudar e orientar quem o procurava. Tudo isso levando em consideração a situação do Cruzeiro.

Ao longo das disciplinas foi possível conhecer diversos alunos do programa, mas gostaria de deixar um agradecimento grandioso ao **Renato Oliveira**, **João Paulo Almeida** e ao **Marcos Viana** pela composição do melhor grupo de EGTP. Deixar minhas desculpas pela questão do códon do triptofano, que pensando melhor, contribuiu para a evolução científica de cada um e eles quem deviam em agradecer.

Agradecer ao pessoal do *clan UFSJ CCO campus Pampulha*: **Bruno**, **Diego**, **Thiago**, **Rodrigo**, **Pedro** e **Vidy** (infelizmente vai o apelido mesmo). As melhores resenhas pós-almoço que abrangiam desde a dúvida se a avião plana até as probabilidades da física quântica.

Ao pessoal do **Laboratório de Biofísica de Macromoléculas**, ao prof. **Adriano Pimenta**, e os membros do **Laboratório de Veneno e Toxinas de Animais** por me aguentarem e pelo incrível suporte que todos ofereceram. Ninguém trabalha sozinho e precisamos sempre de pessoas assim ao nosso lado.

Por fim, mas não menos importante, agradecer aos meus pais pelo incrível suporte que me deram ao longo desse tempo no mestrado de infor... biologia... bio o quê mesmo Bruno?

RESUMO

O processo de replicação do DNA é um mecanismo que deve garantir a sua fidelidade e eficiência de forma a evitar a propagação de erros no material genético e manter a integridade do mesmo. Na presença de erros no DNA, a forquilha de replicação é impedida de continuar e, conseqüentemente, é reconhecida por complexos proteicos que irão iniciar 2 vias de sinalização: a Síntese Translesão ou a Forquilha de Replicação Reversa. Enquanto em leveduras uma única proteína, a Rad5, reconhece a forquilha de replicação parada, em humanos esta função é desempenhada por 2 proteínas distintas, a HLTF e SHPRH. Tanto a Rad5 quanto a HLTF têm a função de estabilizar o complexo na forquilha por meio de um domínio conhecido como HIRAN. Há uma carência de estudos a respeito dos mecanismos moleculares deste domínio em interação com a ssDNA e, devido a sua importância no início da via de sinalização, hipotetiza-se que variações estruturais sejam capazes de interferir na interação do domínio HIRAN com a ssDNA. Portanto, o objetivo deste trabalho é estudar o domínio HIRAN das proteínas HLTF e Rad5 a fim de encontrar possíveis diferenças estruturais entre as mesmas que sejam relacionados com o mecanismo molecular de interação com seus alvos. Além disso, tentar correlacionar estes resultados com a necessidade em humanos da existência de uma segunda proteína no reconhecimento da forquilha parada, a SHPRH, juntamente com a HLTF, uma vez que em leveduras somente uma proteína realiza esta função. Por ser uma estrutura de difícil resolução por métodos experimentais, neste trabalho tentamos elucidar a estrutura terciária e o comportamento dinâmico do HIRAN da Rad5 por métodos de modelagem computacional. O uso da modelagem comparativa e da dinâmica molecular permitiu acesso a informações estruturais do domínio HIRAN da Rad5 e da HLTF. Os resultados encontrados demonstram que o domínio HIRAN da proteína Rad5 possui um maior grau de liberdade e apresenta desnaturação de estruturas secundárias, sendo mais instável que o domínio HIRAN da HLTF. Importante ressaltar que, quando há a presença de ssDNA, ambos os domínios se tornam mais estáveis. As regiões de flexibilidade estrutural nas estruturas terciárias estão correlacionadas, sugerindo um mecanismo similar de interação com a fita simples e indução de conformação por parte da presença do ligante. As análises dos contatos atômicos sugerem um número maior de interações eletrostáticas com os nucleotídeos, por parte do domínio da Rad5. Conclui-se que os domínios podem apresentar mecanismos moleculares similares de interação com a ssDNA e que o domínio HIRAN da Rad5 possivelmente abrange diferentes conformações da ssDNA, uma vez que contém um número maior de resíduos polares na região do sítio de interação.

Palavras Chaves: Forquilha de Replicação Reversa; Domínio HIRAN; ssDNA; Dinâmica Molecular.

ABSTRACT

The DNA replication process is a mechanism that must guarantee its fidelity and efficiency to avoid the propagation of errors in the genetic material and maintain its integrity. In the presence of errors in the DNA, the replication fork is stalled and is recognized by protein complexes that will start two signaling pathways: the Translesion Synthesis or the Reverse Replication Fork. While in yeasts a single protein, Rad5, recognizes the stalled fork, in humans this function is performed by two distinct proteins, HLTF and SHPRH. Both Rad5 and HLTF function stabilize the fork complex through a domain known as HIRAN. There is a lack of molecular mechanisms studies of this domain in interaction with ssDNA and, given its importance at the beginning of the signaling pathway, it is hypothesized that structural variations can interfere in its interaction with the ssDNA. Therefore, the aim of this work is to study the HIRAN domain of the HLTF and Rad5 proteins to find possible structural differences between them that are related to the molecular mechanism of interaction with their targets. In addition, trying to correlate these results with the humans' need for the existence of a second protein in the recognition of the stalled fork, SHPRH, together with HLTF, since in yeasts only one protein performs this function. As a difficult structure to solve by experimental methods, in this work we try to elucidate the tertiary structure and the dynamic behavior of Rad5 HIRAN domain by computational modeling methods. Using of comparative modeling and molecular dynamics allowed access to structural information from the HIRAN domain of Rad5 and HLTF. Our results show that the HIRAN domain of the Rad5 protein has a greater degree of freedom and denaturation of secondary structures, being more unstable than the HIRAN domain of HLTF. It is important to note that, when in the presence of ssDNA, both domains become more stable. The dynamic regions in the tertiary structures are correlated, suggesting a similar interaction with ssDNA and a conformational change induced by the ligand. The analysis of atomic contacts suggests a greater number of electrostatic interactions with nucleotides by the Rad5 domain. We concluded that the domains have a similar mechanism of interaction with the ssDNA and that the HIRAN domain of Rad5 can cover different conformations of the ssDNA, since it contains a greater number of polar residues in the interaction pocket.

Key words: Fork Reversal; HIRAN domain; ssDNA; Molecular Dynamics.

ÍNDICE DE FIGURAS

Figura 1: Processos envolvidos na formação da forquilha de replicação do DNA.....	16
Figura 2: Estrutura terciária do domínio HIRAN da HLTF em solução.....	18
Figura 3: Resíduos responsáveis pela formação da cavidade de interação e estabilização da ssDNA do domínio HIRAN da proteína HLTF.....	19
Figura 4: Distribuição de domínios proteicos ao longo das sequências primárias das proteínas de levedura Rad5 e Rad54, e de humanos HLTF, SHPRH e Rad54.....	20
Figura 5: Representação dos motivos de interação ao DNA da família SWI2/SNF2 na estrutura da proteína Rad54 de <i>zebrafish</i>	21
Figura 6: Domínio PHD da proteína SHPRH.....	24
Figura 7: Exemplos de parâmetros e cálculos utilizados por um Campo de Força.....	29
Figura 8: Cálculos utilizados para o desenvolvimento da trajetória na etapa de produção da dinâmica molecular.....	30
Figura 9: Estrutura e nomenclatura atômica da ssDNA.	42
Figura 10: Predição de estrutura secundária do domínio HIRAN da Rad5 e alinhamento com a sequência do domínio da proteína HLTF.	44
Figura 11: Análises angulares <i>phi</i> e <i>psi</i> do modelo scRad5.....	45
Figura 12: Análises das interações inter-residuais do modelo scRad5.....	46
Figura 13: Estrutura terciária do domínio HIRAN de <i>Saccharomyces cerevisiae</i>	47
Figura 14: Superfícies moleculares da estrutura terciária do domínio HIRAN da HLTF e da Rad5.....	48
Figura 15: Dinâmicas estruturais dos modelos hsHLTF _{APO} e scRad5.....	50
Figura 16: Flexibilidade residual a partir do RMSF dos modelos hsHLTF _{APO} e scRad5.	52
Figura 17: Desnaturalizações e diferenças estruturais nos modelos hsHLTF _{APO} e scRad5.	53
Figura 18: Estruturas secundárias formadas ao longo da etapa de produção do modelo hsHLTF _{APO}	55
Figura 19: Estruturas secundárias formadas ao longo da etapa de produção do modelo scRad5.	56
Figura 20: Número de resíduos com estruturas secundárias ao longo da etapa de produção dos modelos hsHLTF _{APO} e scRad5.....	57
Figura 21: Estruturas representativas das etapas de produção dos modelos hsHLTF _{APO} e scRad5.	59
Figura 22: Importância dos resíduos R71 e Y72 para a formação do sítio de interação a ssDNA.	61

Figura 23: Possível sítio de interação com a ssDNA do domínio HIRAN da Rad5.	62
Figura 24: Dinâmica estrutural da cadeia A do modelo hsHLTF _{HOLO}	63
Figura 25: Estruturas representativas da etapa de produção do modelo hsHLTF _{HOLO}	64
Figura 26: <i>Docking</i> molecular dos modelos hsHLTF _{HOLO/ssDNA} , hsHLTF _{APO/ssDNA} e scRad5 _{ssDNA} com poli-T.	65
Figura 27: Sistemas utilizados para a dinâmica molecular com ssDNA dos modelos hsHLTF _{HOLO/ssDNA} , hsHLTF _{APO/ssDNA} e scRad5 _{ssDNA}	66
Figura 28: Dinâmicas estruturais da cadeia A modelo hsHLTF _{HOLO/ssDNA} e da ssDNA.	67
Figura 29: Resíduos do modelo hsHLTF _{HOLO/ssDNA} que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.	69
Figura 30: Dinâmicas estruturais das porções proteica e nucleotídica do modelo hsHLTF _{APO/ssDNA}	70
Figura 31: Resíduos do modelo hsHLTF _{APO/ssDNA} que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.	72
Figura 32: Estruturas representativas da etapa de produção do modelo hsHLTF _{APO/ssDNA} sem a formação completa da cavidade de interação.	73
Figura 33: Dinâmicas estruturais das porções proteica e nucleotídica do modelo scRad5 _{ssDNA}	74
Figura 34: A diferença conformacional da ssDNA no modelo scRad5 _{ssDNA}	75
Figura 35: Resíduos do modelo scRad5 _{ssDNA} que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.	76
Figura 36: Estruturas representativas da etapa de produção do modelo scRad5 _{ssDNA}	78
Figura 37: Estruturas secundárias formadas ao longo da etapa de produção do modelo scRad5 _{ssDNA}	78
Figura 38: Número de resíduos com estruturas secundárias ao longo da etapa de produção do modelo scRad5 _{ssDNA}	79

ÍNDICE DE TABELAS

Tabela 1: Resumo das estruturas utilizadas nas dinâmicas moleculares.	37
Tabela 2: Valores de cortes utilizados pelo algoritmo de agrupamento	39
Tabela 3: RMSD médio e seu desvio-padrão das réplicas dos modelos hsHLTF _{APO} e scRad5.	51
Tabela 4: Número de resíduos médio com estruturas secundárias e seus desvios-padrões dos modelos hsHLTF _{APO} e scRad5.....	57

LISTA DE ABREVIACES

ATP - Adenosina tri-fosfato.

BER - Reparo por exciso de base, do ingls: *Base Excision Repair*.

CF - Campo de Fora.

CHFR - *Checkpoint protein with FHA and RING finger domain*.

Cryo-EM - *High resolution cryo-electron microscopy*.

DM - Dinmica Molecular.

HLTF - *Helicase-like transcription factor*.

HIRAN - HIP116 Rad5p N-terminal

MM - Mecnica Molecular.

MMS - Metanossulfonato de Metila.

MQ - Mecnica Quntica.

PBC - Condies peridicas de contorno, do ingls: *Periodic Boundary Condition*.

PCNA - Antgeno Nuclear de Clula em Proliferao, do ingls: *Proliferating Cell Nuclear Antigen*.

PDB - *Protein Data Bank*.

PHD - *Plant Homeodomain*.

PME - *Particle-Mesh Ewald*.

Poli-T – Sequncia de timinas.

PRR - Reparo ps replicao, do ingls: *Post Replication Repair*.

RFC - Fator de Replicao C, do ingls: *Replication Factor C*.

RMN - Ressonncia Magntica Nuclear.

RMSD - *Root Mean Square Deviation*.

RMSF - *Root Mean Square Fluctuation*.

RPA - Protena de replicao A, do ingls: *Replication Protein A*.

SAXS - *Small-angle X-ray scattering*.

SHPRH - *SNF2 Histone-linker PHD-finger RING-finger Helicase*.

ssDNA - DNA de fita simples, do ingls: *single stranded DNA*.

SWI/SNF - *Switching Defective/Sucrose Non-Fermentable*.

TLS - Sntese transleso, do ingls: *Translesion Synthesis*.

UV - Ultravioleta

SUMÁRIO

RESUMO.....	7
ABSTRACT	8
1) INTRODUÇÃO	15
1.1) REPLICAÇÃO DO DNA	15
1.2) FORQUILHA DE REPLICAÇÃO REVERSA	15
1.3) DOMÍNIO HIRAN	17
1.4) AS PROTEÍNAS DA FAMÍLIA SWI2/SNF2.....	19
1.4.1) RAD5	21
1.4.2) PROTEÍNAS HOMÓLOGAS A RAD5 EM HUMANOS.....	22
1.4.2.1) <i>HELICASE-LIKE TRANSCRIPTION FACTOR</i> (HLTF).....	23
1.4.2.2) <i>SNF2 HISTONE-LINKER PHD-FINGER RING-FINGER HELICASE</i> (SHPRH)	23
1.5) ESTUDOS <i>IN SILICO</i> DAS ESTRUTURAS DE PROTEÍNAS	24
1.5.1) MODELAGEM MOLECULAR COMPARATIVA.....	25
1.5.2) DINÂMICA MOLECULAR.....	27
1.5.3) ANCORAMENTO MOLECULAR.....	32
2) HIPÓTESE.....	35
3) OBJETIVOS.....	35
3.1) OBJETIVOS ESPECÍFICOS	35
4) METODOLOGIA	36
4.1) CONSTRUÇÃO DO MODELO scRad5	36
4.2) DINÂMICA MOLECULAR.....	36
4.2.1) TOPOLOGIA DOS ARQUIVOS	37
4.2.2) CAIXA DE SIMULAÇÃO	38
4.2.3) MINIMIZAÇÃO DE ENERGIA DO SISTEMA	38
4.2.4) EQUILIBRAÇÃO DO SISTEMA	38
4.2.5) ETAPA DE PRODUÇÃO.....	38

4.3) DOCKING MOLECULAR.....	40
4.3.1) GERAÇÃO DOS ARQUIVOS ESTRUTURAIS DO RECEPTOR E LIGANTE	40
4.3.2) SUPERFÍCIE MOLECULAR.....	40
4.3.3) CRIAÇÃO DA CAIXA DE INTERAÇÃO	41
4.3.4) ETAPA DE PRODUÇÃO.....	41
4.4) MODELO DE ssDNA UTILIZADO	41
5) RESULTADOS E DISCUSSÃO	43
5.1) O DOMÍNIO HIRAN DA RAD5 É MAIS EXTENSO NAS REGIÕES DE <i>LOOPS</i> .	43
5.2) O DOMÍNIO HIRAN DA PROTEÍNA HLTf APRESENTA MENOS ALTERAÇÕES ESTRUTURAIS	48
5.3) O SÍTIO DE INTERAÇÃO PODE APRESENTAR DUAS CONFORMAÇÕES	58
5.4) PREPARAÇÃO DAS ESTRUTURAS PARA SIMULAÇÃO COM A ssDNA	64
5.5) A ESTRUTURA TERCIÁRIA DO DOMÍNIO HIRAN É MAIS ESTÁVEL COM A ssDNA NO SÍTIO DE INTERAÇÃO	67
5.5.1) DOMÍNIO HIRAN DA PROTEÍNA HLTf EM SUA CONFORMAÇÃO DE INTERAÇÃO COM A ssDNA	67
5.5.2) DOMÍNIO HIRAN DA PROTEÍNA HLTf EM SUA CONFORMAÇÃO LIVRE COM A ssDNA	70
5.5.3) DOMÍNIO HIRAN DA PROTEÍNA RAD5	73
6) CONCLUSÕES E PERSPECTIVAS.....	80
ANEXO	88
APÊNDICE	101

1) INTRODUÇÃO

1.1) REPLICAÇÃO DO DNA

As células eucarióticas possuem mecanismos para garantir a fidelidade e eficiência da divisão celular e da multiplicação do seu material genético, a fim de evitar consequências que possam ser deletérias na sobrevivência e/ou viabilidade do organismo (O'DONNELL, 2013).

Ekundayo e Bleichert (2019) dividiram o processo de replicação do DNA em diferentes etapas: (a) iniciação: complexos proteicos que compõem a maquinaria de replicação, chamada de replissomo, reconhecem regiões do DNA e formam uma estrutura, bidirecional, denominada de bolha de replicação com 2 forquilhas de replicação, uma para cada sentido. (b) alongamento: o replissomo desenrola a fita dupla de DNA e sintetiza fitas complementares utilizando como molde as fitas já existentes. (c) terminação: após o reconhecimento de sítios específicos de terminação no DNA, outro complexo proteico de terminação irá realizar o desmonte do replissomo e terminar a replicação.

De forma geral, os replissomos de eucariotos envolvidos neste processo possuem as funções de desenovelar a fita dupla de DNA na forma de helicases, sintetizar as fitas líder e atrasada na forma de polimerases e sintetizar iniciadores em cada fita de DNA molde na forma de primases. Outras proteínas como as auxiliaadoras de ligação a DNA de fita simples (ssDNA¹) e o fator de processividade² PCNA³ (Antígeno Nuclear de Célula em Proliferação) também compõe este complexo (YANG, 2019).

O material genético pode sofrer lesões que impedem este processo e a progressão da forquilha de replicação durante a replicação do DNA. Este evento é um grande desafio para a maquinaria celular, uma vez que o organismo precisa contornar o problema para que o processo replicativo continue (PATEL, 2018).

1.2) FORQUILHA DE REPLICAÇÃO REVERSA

À medida que a forquilha de replicação avança, eventuais erros de DNA, como quebra de fita simples ou dupla e erros de pareamento de bases, podem ser encontrados. Desta forma, para a replicação não colapsar e terminar de forma abrupta, o replissomo tem o seu progresso interrompido (CORTEZ, 2019). Os complexos proteicos reconhecem a forquilha de replicação parada a partir de sinalizações específicas em proteínas presentes no replissomo e duas possíveis

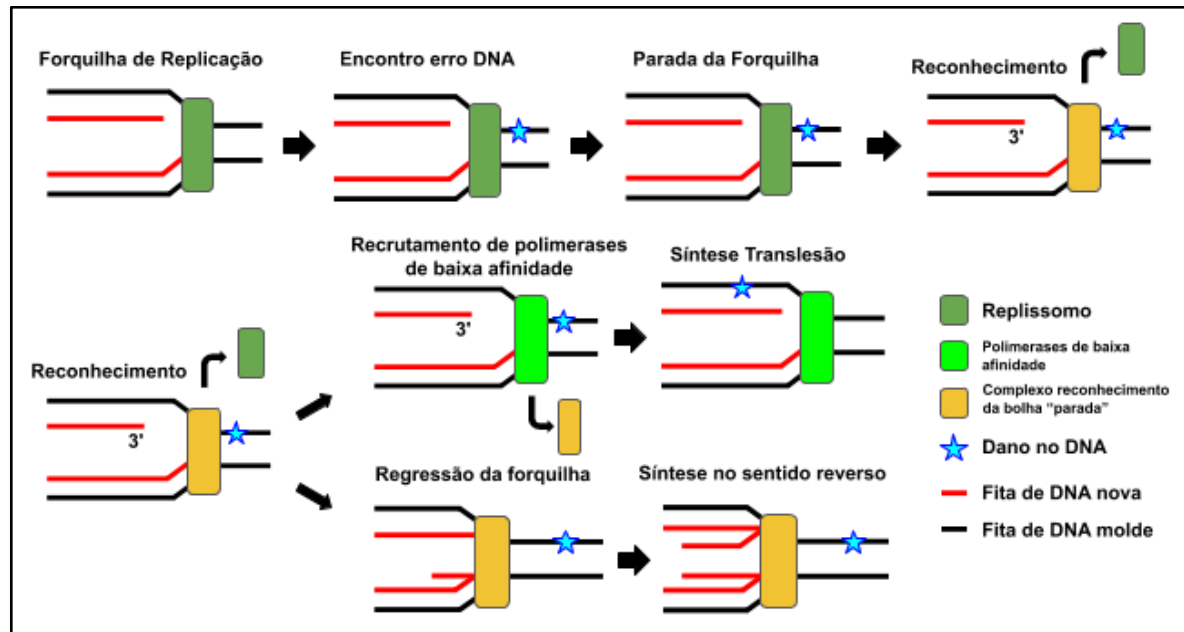
¹ Do inglês, *single stranded DNA*.

² Se referindo a uma característica enzimática de catalisar consecutivas reações sem liberar o substrato.

³ Do inglês, *Proliferating Cell Nuclear Antigen*.

soluções são observadas: (a) recrutamento de polimerases de baixa afinidade para que ocorra a síntese de DNA ignorando o erro e regressão da forquilha e (b) síntese de DNA no sentido reverso (**Figura 1**) (RICKMAN, 2019).

Figura 1: Processos envolvidos na formação da forquilha de replicação do DNA.



Representação esquemática da parada da replicação do DNA e reconhecimento por complexos proteicos para dar início a síntese translesão ou a síntese da fita de DNA no sentido inverso à progressão da forquilha.

A forquilha de replicação reversa consiste na síntese de DNA em sentido oposto ao do curso de movimento da forquilha de replicação, resultando em estruturas similares às junções de Holliday, as “*four-way junctions*”. Após a formação de uma junção de Holliday, a região onde o DNA está danificado torna-se acessível à maquinaria de reparo de DNA e esta estrutura também está associada ao mecanismo de recombinação homóloga (GIBBS, 2019). Independente do erro, o restabelecimento da forquilha de replicação é decisivo para a manutenção da integridade do genoma e para a sobrevivência das células. Este mecanismo de formação e o complexo proteico envolvido no processo de reparo necessitam de estudos complementares e de observações experimentais adicionais (QUINET, 2017).

A formação forquilha de replicação reversa pode ser considerada tanto uma via de reparo sem erros quanto similar a síntese translesão. Sugere-se que a formação da forquilha implique na parada da replicação do DNA para que ocorra acúmulo de fita simples de DNA (ssDNA), possibilitando a atuação de outras vias de reparo. Por outro lado, este mecanismo

pode ser comparado ao da via de tolerância ao erro da Síntese Translesão (TLS⁴), atuando como uma via que ignora o dano no material genético, o que permite o início e a continuação da replicação, para que, posteriormente, outras vias de reparo possam atuar e corrigir o erro remanescente (NEELSEN, 2015).

As modificações pós-traducionais na proteína PCNA, e a estabilização do material genético pelas proteínas estabilizadoras de ssDNA (RPA⁵), dão início a via de sinalização para a formação desta via de reparo. Taglialatela e colaboradores (2017) demonstraram que a atividade de ligação da proteína SMARCAL1, ao complexo ssDNA-RPA, é essencial para estabilização da replicação e formação da forquilha reversa. Vujanovic e colaboradores (2017) relataram que a função de translocase da ZRANB3, ao interagir com o PCNA poli-ubiquitinado, é essencial para retornar a forquilha ao seu estado inicial.

Na literatura são encontradas diversas informações a respeito da importância do domínio proteico HIP116 Rad5p N-terminal (HIRAN) nos processos envolvendo reconhecimento de ssDNA, estabilização do complexo atuante na forquilha de replicação parada, mono/poli-ubiquitinação do PCNA e início da via de reparo da forquilha reversa. (MASUDA, 2018; ELSERAFY, 2018; KILE, 2015; BLASTYAK, 2018; FAN, 2018).

1.3) DOMÍNIO HIRAN

O domínio HIRAN é descrito em diversas proteínas de ligação a ssDNA e possui conservação e distribuição em organismos eucariotos e procariotos. Em bactérias, já foi detectado em grupos gram-positivos com baixa quantidade de GC, bacteroides, actinobacteria e proteobacteria. O gene codificador do HIRAN foi encontrado em alguns bacteriófagos, sugerindo um modo de difusão entre bactérias (IYER, 2006).

O domínio HIRAN é comumente encontrado em eucariotos nas regiões N-terminais das proteínas que contém os motivos da família SWI2/SNF2 ATPases. É também encontrado fusionado com outros domínios que interagem com o DNA, como *TDPI-like DNA phosphoesterase* e VRR-Nuc, e nas proteínas de remodelagem e de modificação da cromatina, como o *DNA dependent SWI2/SNF2 ATPases* e o *RING finger E3 ubiquitin ligase* (IYER, 2006). Este domínio interage especificamente com a extremidade 3'-OH de DNA de fita simples, e facilita a formação inicial das junções de Holliday na forquilha de replicação reversa. Considerando o mecanismo de reconhecimento da forquilha de replicação parada proposto para

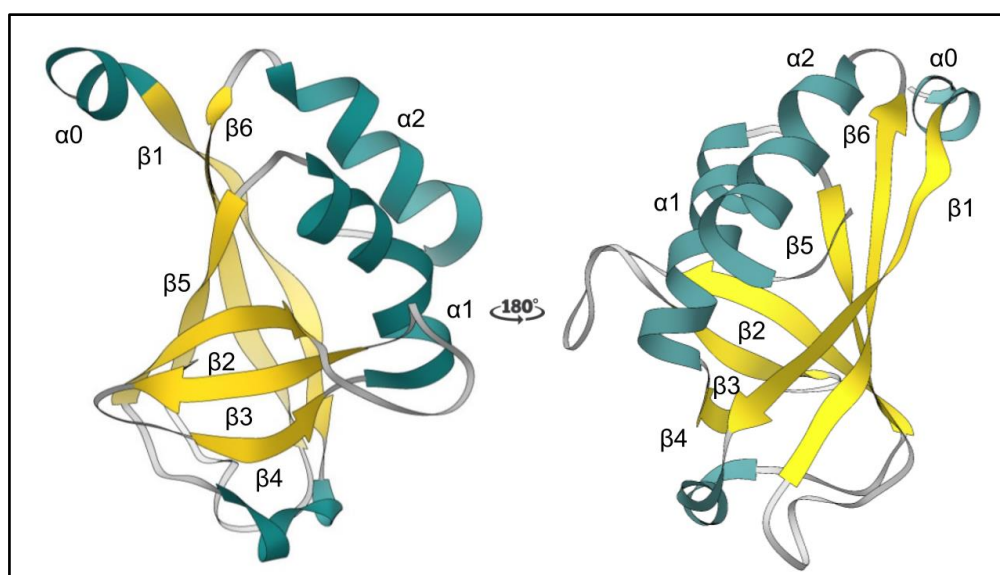
⁴ Do inglês, *Translesion Synthesis*.

⁵ Do inglês. *Replication Protein A*.

a proteína HLTF, quando a mesma interage com a forquilha de replicação parada, o domínio HIRAN confere estabilidade necessária para que o complexo proteico seja funcional (CHAVEZ, 2018).

Em 2016, Korzhnev e colaboradores calcularam a estrutura terciária do domínio HIRAN da proteína HLTF de humanos, em solução salina, a partir de dados obtidos por espectroscopia de ressonância magnética nuclear (RMN). A sua estrutura terciária foi descrita contendo 6 folhas-betas ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$) e 2 alfas-hélices (α_1, α_2) organizadas de modo $\beta_1 - \beta_2 - \beta_3 - \beta_4 - \alpha_1 - \beta_5 - \beta_6 - \alpha_2$ (**Figura 2**).

Figura 2: Estrutura terciária do domínio HIRAN da HLTF em solução.



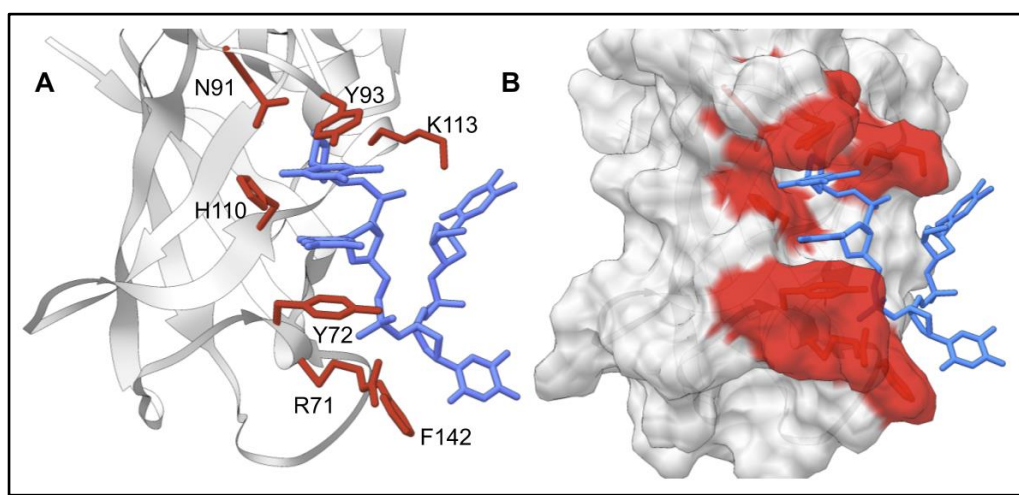
Estrutura do domínio HIRAN da proteína HLTF (PDB ID: 5K5F) resolvida pela técnica de espectroscopia Ressonância Magnética Nuclear. Nesta imagem é possível observar as 2 alfa-hélices e as 6 folhas-beta formando uma estrutura similar a um barril torcido. Fonte: adaptada de Korzhnev e colaboradores (2016).

A estrutura terciária funcional deste domínio é organizada em forma de um barril (**Figura 2**), diagonalmente torcido, com as folhas-betas $\beta_1 - \beta_6 - \beta_5 - \beta_2 - \beta_3 - \beta_4$ e com um dos lados coberto pelas 2 alfas-hélices que formam uma estrutura “*OB-like*” (*Oligonucleotide/Oligosaccharide - Binding*). Possui 2 faces de cargas opostas, sendo a positiva relacionada com a ligação ao DNA. Os aminoácidos presentes no reconhecimento do DNA estão localizados no *loop* $\beta_1 - \beta_2$ (V65, R71, Y72, Y73), no $\beta_2 - \beta_3$ (H110, K113) e no $\beta_5 - \beta_6$ (F142).

As análises da estrutura cristalográfica do domínio HIRAN complexado com ssDNA (**Figura 3**), sugerem que os resíduos Y72 e Y93 são responsáveis por interações essenciais dentro de uma cavidade formada, por meio de interações empilhamento de elétrons π com 2

nucleotídeos. Os resíduos Y93, N91 e H110 auxiliam na estabilização do ssDNA por meio de interações eletrostáticas. Os grupos fosfatos da cadeia de ssDNA interagem com os resíduos R71 e K113 e, fora da cavidade, o resíduo F142 interage com um terceiro nucleotídeo. As análises estruturais dos modelos de HIRAN ligado ao ssDNA e em solução, sugerem que o sítio de ligação a ssDNA pode ser dinâmico e apresentar conformações diferentes (KILE, 2015; KROZHNEV, 2016).

Figura 3: Resíduos responsáveis pela formação da cavidade de interação e estabilização da ssDNA do domínio HIRAN da proteína HLTF.



Estrutura cristalográfica do domínio HIRAN da proteína HLTF em sua conformação de ligação a ssDNA, que possui o código de acesso no PDB (PDB ID) 4S0N. Em A, os resíduos de aminoácidos responsáveis na estabilização da ssDNA e em B, a superfície molecular do HIRAN mostrando a formação do sítio de interação. Em cinza está representado a estrutura terciária do HIRAN, em vermelho os aminoácidos envolvidos na interação com ssDNA e em azul a ssDNA contendo 4 nucleotídeos de timina. Fonte: adaptada de Kile e colaboradores (2015).

Em 2020, Hashiki e colaboradores demonstraram, a partir da técnica de cristalografia de proteínas e difração de raios-x, que o domínio HIRAN da HLTF interage e complexa com a fita dupla de DNA de forma universal, ou seja, independe dos nucleotídeos presentes na região do material genético. O mecanismo molecular de interação ocorre de forma similar ao descrito por Kile e colaboradores (2015) e os autores ressaltam que este domínio possui maior afinidade por DNA de fita simples do que pela fita dupla.

1.4) AS PROTEÍNAS DA FAMÍLIA SWI2/SNF2

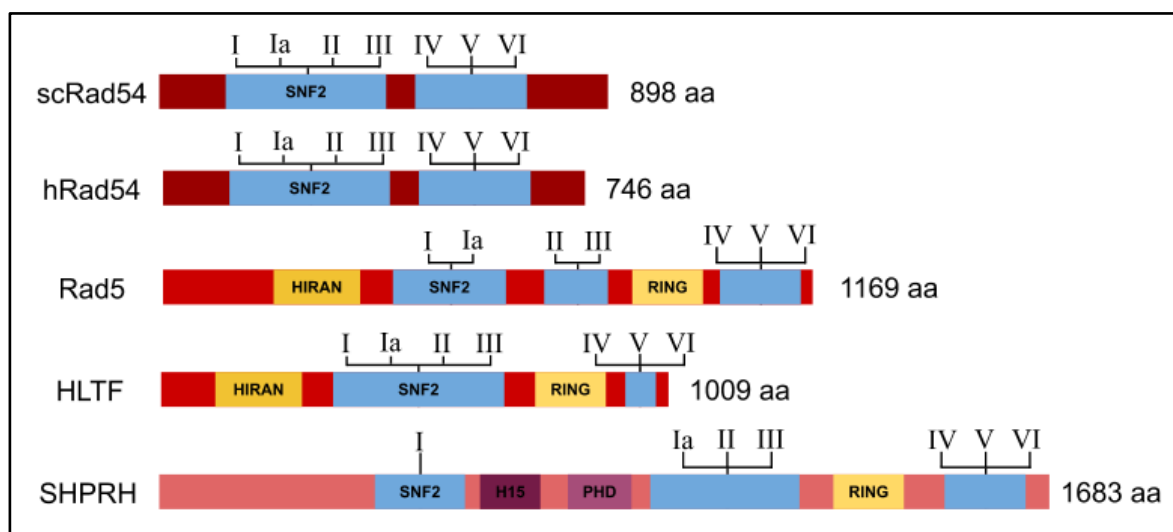
As proteínas da família SWI2/SNF2 (“*switching defective/sucrose non-fermentable*”) são formadas por um grande complexo de subunidades chamado SWI/SNF e são remodeladoras de cromatina ATP-dependentes com funções de DNA translocases. Os domínios da família

SWI/SNF são conservados ao longo da evolução e proteínas homólogas foram identificadas em plantas, leveduras e mamíferos. (HARGREAVES, 2011).

Segundo Hopfner e colaboradores (2012), alguns domínios/motivos podem estar associados de forma adjacente aos domínios SWI2/SNF2, como, por exemplo, o domínio HIRAN (que se liga a extremidade 3'-OH de fita simples de DNA), o RING *finger* (uma ubiquitina ligase) e o HARP (atividade de helicase aneladora ATP-dependente).

Análises de estruturas primárias das proteínas envolvidas no reconhecimento da forquilha de replicação parada e de proteínas reparo de DNA (**Figura 4**), sugerem que a proteína Rad5, encontrada em eucariotos mais simples, possui 2 ortólogos em humanos: a *Helicase-like transcription factor* (HLTF) e *SNF2 Histone-linker PHD-finger RING-finger Helicase* (SHPRH). Ambas possuem uma arquitetura de domínios bastante similares, sendo que a Rad5 possui 21,1% de identidade com a HLTF e 16,7% com a SHPRH (ELSERAFFY, 2018). A baixa identidade entre estas proteínas se dá devido a longas regiões não pertencentes aos domínios conservados.

Figura 4: Distribuição de domínios proteicos ao longo das sequências primárias das proteínas de levedura Rad5 e Rad54, e de humanos HLTF, SHPRH e Rad54.

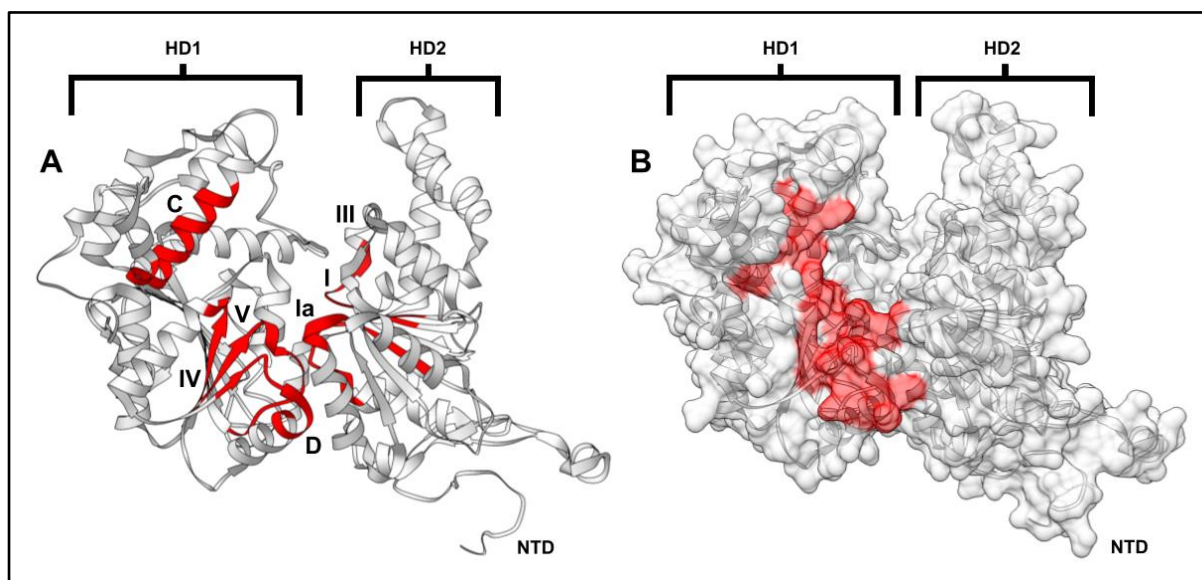


Arquitetura de domínios e motivos das proteínas da família SNF2/SWI2 envolvidas no reconhecimento da forquilha de replicação parada comparado com a proteína Rad54. scRad54 representa a proteína de levedura e hRad54 de humanos. A proteína Rad5 apresenta o domínio HIRAN, domínios da família SNF2 e um domínio RING *finger*, estrutura similar a HLTF, proteína em humanos com maior identidade. A SHPRH possui domínios da família SNF2, domínio H15, PHD e RING *finger*. Adaptado de Elserafy e colaboradores (2018) e Unk e colaboradores (2008).

A proteína Rad54 está presente na via de reparo de DNA por Recombinação Homóloga e possui atividade de translocase, dependente da hidrólise de ATP (CEBALLOS, 2011). A sua

estrutura primária apresenta uma arquitetura de domínios similar a Rad5, HLTF e SHPRH (**Figura 4**), apesar de não possuir o domínio HIRAN e não pertencer ao complexo proteico que interage com a forquilha de replicação parada (UNK, 2008). A estrutura terciária da Rad54 (**Figura 5**) revelou a conservação dos motivos pertencentes a super famílias de helicases (Motivo I, II, III, IV, V, VI) e novos motivos pertencentes a família SWI2/SNF2 (TxGx, Snf-C e Snf-D), tendo as funções já descritas: (a) Motivo I e II relacionados a interações com o ATP e íons Magnésio (Mg^{2+}), respectivamente; (b) Motivos Ia, TxGx, Snf-D, IV e V relacionados a interação translocação do DNA; (c) Motivos III e IV envolvidos na resposta estrutural pós hidrólise de ATP, atuando como um sensor (THOMA, 2005). Nogkhlaw e colaboradores (2012) identificaram um Motivo Q nas proteínas da família SWI2/SNF2 que, juntamente com o Motivo I, é responsável pela catálise do ATP.

Figura 5: Representação dos motivos de interação ao DNA da família SWI2/SNF2 na estrutura da proteína Rad54 de *zebrafish*.



Motivos conservados da família SWI2/SNF2, em eucariotos, que possuem interação com a fita dupla de DNA na estrutura da proteína Rad54 (PDB ID: 1Z3I). Em A, a representação da estrutura terciária da Rad54, resíduos 91-738, e em B a representação da superfície molecular. A proteína é dividida em 2 lóbulos, HD1 e HD2 e possui a região N-terminal destacada em NTD. Em vermelho estão representados os motivos que possuem interação com a fita dupla de DNA: Motivos I, Ia, III, IV, V, Snf-C (C) e Snf-D (D). Fonte: Adaptado de Thoma e colaboradores (2005).

1.4.1) RAD5

A proteína Rad5 é uma proteína de *Saccharomyces cerevisiae*, membro da família SWI/SNF ATPases que possui um motivo RING C_3HC_4 característico das ubiquitina ligases.

Na via de reparo denominada “reparo pós-replicação” (PRR⁶), a proteína Rad5 interage com o complexo enzimático Mm2-Ubc13 para promover a poli-ubiquitinação do PCNA, o que resulta na formação da forquilha de replicação reversa (a associação com Mm2-Ubc13 é dependente do motivo C₃HC₄). Em contraste, a mono-ubiquitinação do PCNA resulta na TLS, a partir do recrutamento de polimerases de baixa afinidade (BLASTYÁK, 2007).

O mecanismo de sinalização que propõe a participação da proteína Rad5 na via da TLS e na forquilha reversa baseia-se em sua interação com a proteína Rev1. Na via de Reparo por Excisão de Base (BER⁷), o PCNA mono-ubiquitinado pelo complexo Rad6-Rad18 é reconhecido pela Rev1 e pela Rad5, com participação da proteína Rad18. O complexo Rad5-Rev1 recruta as polimerases de menor afinidade e realiza a via TLS. Caso a TLS polimerase não consiga atuar no local da lesão do DNA, o complexo Mm2-Ubc13 liga-se à Rad5, poli-ubiquitina o PCNA e ativa a via de reparo da forquilha reversa (XU, 2016) (FAN, 2018).

A proteína Rad5 possui 7 domínios conservados da família SWI2/SNF2, além do motivo “RING *finger*”, relacionado com a interação com Ubc13, 1 zíper de leucina na região N-terminal e o domínio HIRAN, também no N-terminal, relacionados com a ligação a ssDNA (XU, 2016).

Hayashi e colaboradores (2018) relataram a regulação da proteína Rad5 por CDKs na fase G2/S do ciclo celular, por meio da fosforilação do resíduo S130 pelo complexo Cdc25/CDK1, o que confere a Rad5 um tempo de meia-vida menor, sugerindo uma relação entre a fosforilação e a degradação da mesma.

A proteína Rad5 é, então, capaz de reconhecer as forquilhas de replicação paradas, de interagir com a extremidade 3'-OH do DNA, devido a presença do domínio HIRAN, e pode ativar tanto a TLS quanto a formação da forquilha de replicação reversa (FAN, 2018).

1.4.2) PROTEÍNAS HOMÓLOGAS A RAD5 EM HUMANOS

Em humanos, 2 proteínas ortólogas à proteína Rad5 foram identificadas: a SHPRH e a HLTF. Ambas são capazes de reconhecer forquilhas de replicação paradas e promover a mono/poli-ubiquitinação do PCNA. Na literatura, um grupo de pesquisa sugeriu que estas proteínas podem ser dispensáveis nos processos de ubiquitinação e hipotetizaram a existência de outras proteínas envolvidas nesta via de sinalização (KRIJGER, 2011).

⁶ Do inglês, *Post Replication Repair*.

⁷ Do inglês, *Base Excision Repair*.

Elserafy e colaboradores (2018), em seu artigo de revisão, descreveram que a deleção/mutação das proteínas SHPRH e da HLTF resultam em desenvolvimento de cânceres. Lin e colaboradores (2011) relatam que a deleção, de forma alternada e simultânea, das proteínas SHPRH e HLTF foi capaz de aumentar a mutagênese causada por Ultravioleta (UV) e metanossulfonato de metila (MMS), descrevendo que a HLTF é mais responsiva a danos por UV, enquanto a SHPRH a danos causados pelo MMS, sugerindo uma resposta dano-específica. Lin e colaboradores (2011) também demonstraram que a HLTF é capaz de inibir a interação da SHPRH com Rad18 promovendo a mono-ubiquitinação da PCNA e, conseqüentemente, o recrutamento da Pol η na ativação da TLS.

1.4.2.1) HELICASE-LIKE TRANSCRIPTION FACTOR (HLTF)

A capacidade de reconhecimento da forquilha de replicação parada pela HLTF se dá pela interação de seu domínio HIRAN com a extremidade livre 3'-OH de ssDNA, sendo o único HIRAN com estrutura terciária elucidada depositada no banco de dados PDB (KORZHNEV, 2016; KILE, 2015). O mecanismo de ação é similar à Rad5 e, segundo Poole e Cortez (2017), a HLTF é capaz de catalisar a formação da forquilha reversa de forma similar às proteínas SMARCA1 e ZRANB3, porém não consegue catalisar o recomeço da forquilha.

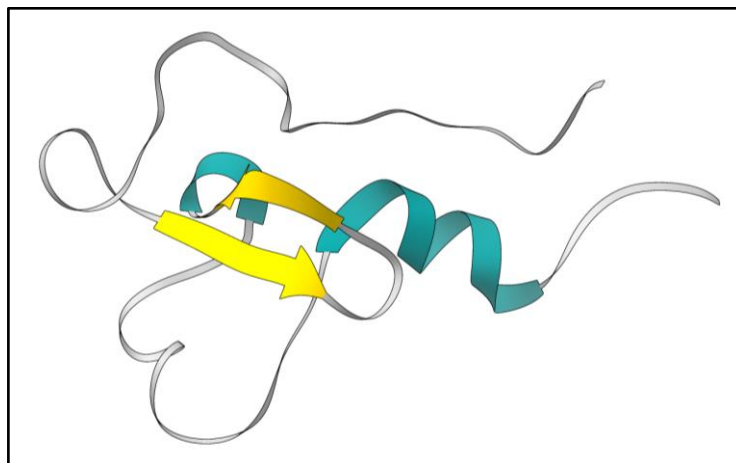
O mecanismo de sinalização proposto, similar a Rad5, é que a formação da forquilha reversa ocorre quando PCNA é poli-ubiquinado pelo complexo Rad6-Rad18, tornando-se o substrato para a HLTF, caso a forquilha de replicação esteja parada por um longo período de tempo. Este evento, dependente do Fator de Replicação C (RFC), e pode ocorrer caso a via da TLS não consiga atuar no local da lesão no DNA (MASUDA 2018).

A regulação da atividade da HLTF é realizada por 2 proteínas: a USP7 e a CHFR. A USP7 (*ubiquitin specific protease 7*) é uma enzima “desubiquitinante” que desubiquitina a HLTF, aumentando o tempo de meia-vida e, conseqüentemente, a poli-ubiquitinação do PCNA, enquanto a CHFR (*checkpoint protein with FHA and RING finger domain*) promove a ubiquitinação da HLTF e diminui o seu tempo de meia-vida (QING, 2011) (KIM, 2010).

1.4.2.2) SNF2 HISTONE-LINKER PHD-FINGER RING-FINGER HELICASE (SHPRH)

Diferente da HLTF, o domínio HIRAN não é descrito na estrutura primária e terciária da proteína SHPRH. Outra diferença é a presença do domínio PHD (*plant homeodomain*; **Figura 6**), entre os resíduos 625-716, que está relacionado com o remodelamento de cromatina e regulação da transcrição, sendo considerado sensor de metilação nas caudas das histonas, no resíduo lisina 4 da histona H3 (H3K4) (MACHADO, 2013).

Figura 6: Domínio PHD da proteína SHPRH.



Estrutura terciária do domínio PHD da proteína SHPRH (PDB ID: 2M85). Em ciano estão representadas as alfa-hélice e em amarelo as folhas-beta. Fonte: Adaptado de Machado e colaboradores (2013).

Tomi e colaboradores (2014) detectaram poli-ubiquitinação da PCNA em células que tinham HLTF silenciada e em diferentes estágios de inativação da SHPRH, sugerindo que deve haver uma outra ubiquitina E3 ligase *in vivo*.

Um outro grupo de pesquisa demonstrou que a proteína SHPRH está relacionada a regulação da transcrição de RNA ribossomal, ligando-se a região promotora e recrutando a RNA polimerase I (reconhecimento feito pelo domínio PHD), e que pode atuar como leitor de marcação epigenética em histonas (LEE, 2017) (LEE, 2018).

Por fim, a via de sinalização resultante do reconhecimento da forquilha de replicação parada, por parte da SHPRH, resulta somente na TLS e ainda não foi demonstrado experimentalmente a formação da forquilha de replicação reversa, apesar dos domínios SWI2/SNF2 conservados.

1.5) ESTUDOS *IN SILICO* DAS ESTRUTURAS DE PROTEÍNAS

A maioria dos estudos envolvendo a via de reparo de DNA da forquilha de replicação reversa descrevem as proteínas envolvidas no processo de formação e estabilização. Estudos adicionais envolvendo as informações a respeito dos mecanismos moleculares são necessários para contribuir na compreensão destes eventos.

Resultados anteriores do nosso grupo de pesquisa (**Anexo I**) demonstraram a dificuldade na obtenção da proteína recombinante referente ao domínio HIRAN da proteína Rad5 para realização dos estudos moleculares. Desta forma, uma alternativa para o estudo do mesmo é a partir de métodos computacionais validados e aplicados amplamente na literatura.

Os estudos das variações estruturais, *in silico*, podem ser realizados com as técnicas de modelagem molecular: Dinâmica Molecular, *Docking* Molecular e Modelagem Comparativa. Estes métodos permitem inferir o comportamento estrutural de uma molécula e correlacioná-lo com sua atividade/afinidade, uma vez que proteínas podem adquirir diversas conformações transitórias e ser estabilizadas a partir da interação com um ligante (NAQVI, 2018).

Nos estudos envolvendo sistemas biológicos, é importante conhecer os mecanismos moleculares, caracterizar e possivelmente entender funções de sistemas proteicos e não-proteicos. Cada sistema possui a sua individualidade, seja na composição, nas interações intra-residuais, nas estruturas primárias e terciárias e sua capacidade de interagir com o ambiente e outros compostos (SNEHA, 2016).

1.5.1) MODELAGEM MOLECULAR COMPARATIVA

As informações relativas as estruturas terciárias e secundárias de proteínas e peptídeos, obtidas de forma experimental, auxiliam as técnicas de bioinformática relacionadas às estruturas tridimensionais. Elucidar e caracterizar essas estruturas por técnicas de RMN, ou cristalografia de proteínas seguida por difração de raios-x, demandam mais tempo e mais recursos financeiros quando comparadas com as técnicas *in silico* de modelagem molecular. Estas técnicas não substituem as validações experimentais, entretanto nem sempre é possível obter uma proteína recombinante ou purificá-la do sistema onde é produzida naturalmente. Como uma alternativa, através da utilização de cálculos estatísticos e leis da física da mecânica clássica e quântica somos capazes de estudá-las a partir de métodos computacionais (VYAS, 2012; MUHAMMED, 2018; WATERHOUSE, 2018).

As técnicas envolvidas na predição teórica de proteínas/peptídeos podem ser divididas em: (a) modelagem *ab initio*, técnica que baseia-se em cálculos derivados de fórmulas físico-químicas para predição de interações entre os resíduos; (b) modelagem comparativa, ou por homologia, que se baseia na construção de um modelo a partir da substituição, inserção e deleção de aminoácidos, utilizando como molde uma estrutura já caracterizada por métodos experimentais (BITENCOURT-FERREIRA, 2019).

Os pesquisadores Vyas e colaboradores (2012) listaram diversas funções e aplicações da modelagem molecular, dentre elas, o estudo dos efeitos das mutações na estrutura terciária de uma proteína, construção de novos ligantes e receptores para serem empregados em técnicas como a dinâmica e *docking* molecular, refinamento de estruturas resolvidas a partir da cristalografia de proteínas seguida de difração de raios-x e planejamento e execução de novos experimentos computacionais que visam mudanças estruturais.

A etapa inicial para realizar a modelagem comparativa é a busca por estruturas em bancos de dados a partir da estrutura primária. Esta abordagem combina a ferramenta BLAST (MCGINNIS, 2004), utilizando os valores de identidade e similaridade, para realizar o alinhamento e selecionar o molde no PDB (BERMAN, 2003), é bastante difundida na literatura (PARULEKAR, 2013; NOCUA, 2014; BAVI, 2016; TARAZI, 2016; PETERSON, 2019). É recomendado o uso de estruturas com a maior identidade e ancestralidade próxima, além de ser possível utilizar mais de uma estrutura como referência (MUHAMMED, 2018). Posteriormente as estruturas selecionadas são alinhadas com a estrutura primária da proteína a ser modelada para a construção do modelo.

A construção de um modelo emprega diversas técnicas que podem ser agrupadas em (VYAS, 2012; WEBB, 2014; KHAN, 2016; MUHAMMED, 2018): (a) Montagem por Corpo Rígido⁸, realiza a divisão do modelo de referência em “núcleos” conservados de estruturas secundárias, posteriormente a substituição dos aminoácidos de acordo com o alinhamento e conecta as regiões com inserções dos resíduos não alinhados. Disponível nos programas SWISS MODEL (WATERHOUSE, 2018) e no ROSSETA (SONG, 2013); (b) Pareamento de Segmentos⁹, divisão do modelo de referência e a sequência primária da estrutura a ser modelada em pequenos segmentos e os compara com bancos de dados para definir a melhor orientação espacial. Disponível no programa SEGMOD/ENCAD (LEVITT, 1992); (c) Restrição Espacial¹⁰, utiliza cálculos de campos de força para criar restrições espaciais e adicionar os resíduos nas regiões indicadas pelo alinhamento da estrutura primária de referência com a que será calculada. Metodologia utilizada pelo programa MODELLER (SALI, 1993); (d) Evolução Artificial¹¹, no qual é feito substituições nos resíduos da estrutura de referência até que a sequência primária seja igual à da estrutura a ser modelada. Pode ser feito pelo programa NEST (PETREY, 2003).

A modelagem das cadeias laterais dos resíduos de aminoácidos baseia-se em estruturas análogas, fatores energéticos e estéricos. A biblioteca de rotâmeros, conformações dos resíduos cuja cadeia lateral possui baixa energia, auxilia a modelagem a inserir os resíduos de forma a evitar impedimentos estéricos na estrutura global da proteína/peptídeo (KHAN, 2016).

Os resíduos não alinhados a sequência primária de referência, quando a estrutura a ser modelada possui uma quantidade maior de aminoácidos, são adicionados no formato de *loops*.

⁸ do inglês, *rigid-body assembly*.

⁹ do inglês, *segment matching*.

¹⁰ do inglês, *spatial restraints*.

¹¹ do inglês, *artificial evolution*.

Estes *loops* podem ser descritos como uma inserção de aminoácidos que não possuem uma estrutura secundária definida, com a estrutura variável e de difícil predição. As metodologias adotadas pelos programas para modelagem destas estruturas foi dividida por Webb e colaboradores (2014) em: (a) busca em bancos de dados de todas as estruturas de proteínas depositadas para encontrar um padrão a ser utilizado e (b) variações conformacionais, baseadas em funções de ranqueamento, até estabelecer uma estrutura com energia mínima.

O refinamento do modelo gerado pela modelagem é posteriormente submetido a minimização de energia a partir de cálculos em modelos de campos de força de mecânica molecular, a fim de equilibrar o sistema em um mínimo energético global. Esta etapa pode ser realizada para evitar incoerências espaciais na estrutura terciária, como átomos em posições irregulares e sobreposições de nuvens eletrônicas, por exemplo (MUHAMMED, 2018).

A validação do modelo pode ser feita por plataformas disponíveis gratuitamente e que irão analisar as propriedades estereoquímicas da proteína (Verify3D; LUTHY, 1992), gerar o gráfico de Ramachandran (PROCHECK; LASKOWSKI, 1993) a partir dos ângulos *phi* e *psi*, realizar análises de simetria e geometria, como quiralidade, tamanho e ângulo das ligações (PROCHECK e WHATCHECK; HOOFT, 1996), analisar volume das nuvens eletrônicas para cristalografia de proteínas (PROVE; PONTIUS, 1996) e interações inter-residuais (ERRAT; COLOVOS, 1993)

1.5.2) DINÂMICA MOLECULAR

As simulações de dinâmica molecular (DM) são capazes de contribuir com o entendimento, *in silico*, do quão dinâmico é um determinado sistema, permitindo observações das biomoléculas em escalas atômicas. Portanto, o entendimento de processos biológicos, como por exemplo o enovelamento¹² de uma proteína, as variações estruturais tempo-dependentes, as mudanças conformacionais, a catálise enzimática e a interação de uma proteína com um ligante, tornam-se cada vez mais possíveis devido à melhora da capacidade de processamento e paralelismo computacional (LIU, 2017; CHENG, 2012; GANESAN, 2016).

Nos sistemas de DM, é possível destacar 3 grupos distintos de cálculos: (a) Mecânica Molecular (MM) (ALLEN, 1989), na qual os átomos são parametrizados, seguem funções de potenciais de energias e campos de força, e os dados são retirados de experimentos empíricos; (b) Mecânica Quântica (MQ) (KRESSE, 1993), a qual baseia-se na equação de Schrödinger para determinar o estado de cada núcleo e nuvem eletrônica de cada átomo; (c) MM-MQ

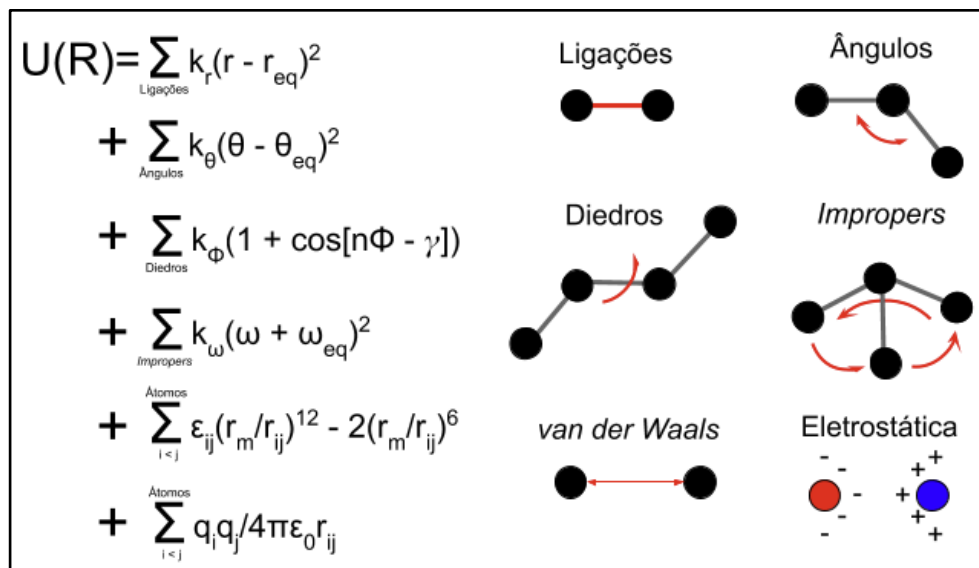
¹² Do inglês, *fold*ing.

(WARSHEL, 1976), uma abordagem híbrida dos métodos que visa aprimorar e complementar as abordagens durante a simulação (SELVARAJ, 2017).

Os parâmetros a respeito das interações atômicas são necessários para que, quando combinados com as posições atômicas e seus movimentos, seja possível calcular suas energias e as forças associadas. A energia total do sistema depende da temperatura, das acelerações e dos movimentos atômicos estimados a partir do campo de força. Os resultados informam a configuração e o estado de cada átomo do sistema em um determinado tempo de simulação (VLACHAKIS, 2014; DEVIVO, 2016).

O campo de força (CF) pode ser resumido em um conjunto de fórmulas matemáticas e parâmetros analíticos moleculares. Seus dados podem ser obtidos por meios de cálculos *ab initio* de MQ ou experimentalmente por dados de RMN, de espectroscopia de Raman e de Infravermelho e de Difração de Raios-X (DEVIVO, 2016). O CF descreve os parâmetros que serão utilizados nos cálculos da DM, dentre eles dados respectivos a geometria e as conformações da molécula, como o comprimento das ligações, ângulos, diedros e impróprios (**Figura 7**). Todas as nomenclaturas dos compostos e dos átomos, as energias de interações, barreiras de torção, frequência de vibrações e cargas são descritos. Ao longo do tempo, os campos de forças vão sendo atualizados e modificados, pelos desenvolvedores, para conter os parâmetros necessários para um determinado sistema, fazendo com que exista diferentes campos de força para moléculas específicas. A escolha do campo de força deve ser feita visando uma melhor adequação ao sistema (GANESAN, 2016).

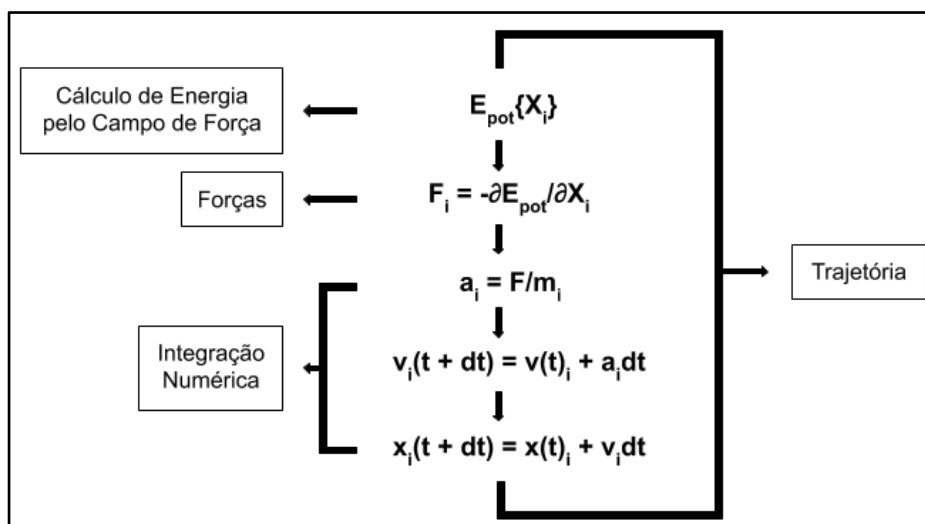
Figura 7: Exemplos de parâmetros e cálculos utilizados por um Campo de Força.



Campos de força clássicos utilizados em simulações de DM. $U(R)$ representa a função de potencial energético utilizada para derivar forças atômicas para movimentos moleculares; k_r , k_θ , k_ϕ e k_ω são constantes de forças; r_{eq} , θ_{eq} e ω_{eq} são posições de equilíbrio; r é o comprimento da ligação; θ é o ângulo atômico; ϕ é o ângulo diedro; ω é o ângulo dos impróprios; r_{ij} é a distância entre um átomo i e j ; n é a multiplicidade; γ é a troca de fase; ϵ_{ij} está relacionado ao mínimo da função de potencial Lennard-Jones entre o átomo i e j ; r_m é a distância que o potencial chega no mínimo; q_i e q_j são as cargas dos seus respectivos átomos; ϵ_0 é a constante dielétrica. Fonte: adaptado de Chang e colaboradores (2016).

Todos os parâmetros e classificações do CF são utilizados para o cálculo de energia no sistema (**Figura 7**) e os algoritmos de DM irão buscar, inicialmente, diminuir a energia do sistema a fim de encontrar conformações estruturais estáveis e em equilíbrio. Este cálculo de energia se torna importante para a etapa de produção da DM (**Figura 8**) (CHANG, 2016). A partir da energia do sistema, e conhecendo as coordenadas espaciais de cada átomo, é possível aplicar forças ao mesmo e obter aceleração resultante desta ação. A aceleração, então, implica em uma velocidade relativa e posteriormente uma mudança na posição espacial após percorrer uma distância por um tempo determinado. Por fim, há um novo cálculo de energia que irá gerar novas forças nos sistemas e novos movimentos, fechando um ciclo denominado de trajetória da DM. Dado estes fatos, a DM que utiliza a mecânica clássica pode também ser definida como uma integração das leis de Newton a partir da variável tempo (HOSPITAL, 2015).

Figura 8: Cálculos utilizados para o desenvolvimento da trajetória na etapa de produção da dinâmica molecular.



Exemplos de cálculos utilizados para a construção da trajetória da DM e criar a movimentação espacial dos átomos no sistema. E_{pot} representa a energia potencial oriunda do campo de força; t é o tempo de simulação; dt é o tempo de integração; X é a coordenada atômica; F é o componente força; a é a aceleração; m é a massa atômica; v é a velocidade; a letra i representada para indicar cada coordenada espacial de um número N de átomos simulados. Fonte: adaptado de Hospital e colaboradores (2015).

Outros tipos de algoritmos estão presentes na DM, como algoritmos para os cálculos de interações de curtas e longas distâncias: a lista de Verlet (VERLET, 1967) e o algoritmo de Ewald (DARDEN, 1993), respectivamente, e as interações de átomos não-ligados por potenciais de Lennard-Jones (LENNARD-JONES, 1931) e a Lei de Coulomb. O sistema de DM pode ser configurado para ocorrer em situações termodinâmicas bem definidas, de modo que se mantém constante o número de partículas, o volume, a pressão e a temperatura. A temperatura é um fator que impacta a cinética proteica, sendo que a sua elevação pode causar desestruturações. Os algoritmos que controlam a temperatura são denominados de termostatos, os quais controlam a energia cinética do sistema a partir da mudança de velocidades (CHENG, 2012; VLACHAKIS, 2014) e os algoritmos que controlam a pressão são denominados barostatos (PARRINELLO, 1981). Portanto, ao longo da trajetória, a integração das Leis de Newton, perante a variável tempo, gera as posições e as velocidades específicas de todos os átomos do sistema (VLACHAKIS, 2014).

Diversos pacotes¹³ para realizações de DM estão disponíveis, alguns de forma gratuita, e podem apresentar o seu campo de força próprio ou permitir a utilização daqueles

¹³ Do inglês, *packages*: se referindo ao conjunto de ferramentas e programas que constituem o pacote.

desenvolvidos por terceiros. Alguns destes pacotes tem utilizações específicas, sendo alguns bastantes difundidos na literatura, dentre eles: (a) CHARMM (JO, 2008), utilizado em sistemas que envolvem a presença de membranas lipídicas, glicoproteínas e peptídeos membrana ativos (SKJEVIK, 2015; PEZESHKIAN, 2017; MALLAJOSYULA, 2015; XIANG, 2018; GRASSO, 2018); (b) AMBER (CORNELL, 1995), amplamente difundido no estudo envolvendo proteínas, em todo os seus espectros, e moléculas que interagem com fita dupla e simples de DNA (SHAO, 2018; PRADHAN, 2018; HAMED, 2017; KAMARAJ, 2015; MARY, 2017; OPRZESKA-ZINGREBE, 2018); (c) OPLS (JORGENSEN, 1987), utilizado em parametrizações de moléculas de RNA e cálculos de cargas anisotrópicas de pequenas moléculas (ROBERTSON, 2019; YAN, 2017); (d) GROMOS (OOSTENBRINK, 2004), utilizado em sistemas envolvendo carboidratos (POL-FACHIN, 2012; NESTER, 2019).

A DM é realizada em um sistema que simula uma caixa de solvatação, contendo íons de sódio, cloro e potássio, para a neutralização das cargas do sistema, e das moléculas de água. É possível controlar a concentração destes sais para uma melhor mimetização do meio biológico. Nesta caixa é possível a criação de condições periódicas de contorno (PBC¹⁴) (ALLEN, 1987), que utilizam algoritmos como *particle-mesh Ewald* (PME) para cálculos de interações eletrostáticas de longa distância. As PBC permitem que um átomo não seja “deletado” ao sair da caixa, e sim se mova para o lado oposto. Deste modo, a simulação mantém o mesmo número de átomos no sistema, de modo que todas as coordenadas de velocidade e posição se mantenham (GANESAN, 2016).

A escolha do modelo da água a ser usado também interfere nas simulações de MD. Os modelos predizem as propriedades físico-químicas da água e sua estrutura tridimensional a ser usada. Quanto mais complexo o modelo, mais pesado computacionalmente. Cada modelo de água empregado depende da temperatura do sistema, número de átomos, fase atual do sistema (líquido ou sólido) e possui parâmetros como o tamanho das ligações hidrogênio-oxigênio, o ângulo da ligação, as medidas de polarização, a otimização de energia e a densidade eletrônica. Os modelos mais comuns utilizados são o SPC, SPC/E, TIP3P e TIP4P, sendo que cada campo de força possui um modelo recomendado (CHAPLIN, 2019).

A DM fornece informações teóricas, matemáticas e probabilísticas de um sistema e, portanto, possui limitações, tal como qualquer outro método científico. As limitações nos parâmetros do campo de força também interferem na veracidade dos cálculos de MD. Algumas validações experimentais podem ser feitas para comprovações: (a) as estruturas podem ser

¹⁴ Do inglês, *Periodic Boundary Conditions*.

validadas por experimentos de RMN e cristalografia de proteínas; (b) posições dos hidrogênios, em relação às interações de hidrogênio, comparadas com dados de difração de nêutrons; (c) interações inter-residuais confirmadas por experimentos de NOESY de RMN; (d) características de estruturas secundárias observadas em dicroísmo circular e/ou espectroscopia de Raman. Uma forma de validação *in silico* é a realização de repetições que partem de uma conformação inicial diferente (CHENG, 2012).

Por fim, a DM pode ser utilizada juntamente com a ancoragem molecular, uma vez que é capaz de abstrair informações conformacionais de um receptor, analisar as interações ligante-receptor, observar o papel do solvente no sítio ativo, refinar a estrutura final do *docking* e observar se as interações são estáveis ao longo de toda a trajetória (DE VIVO, 2016; SALMASSO, 2018). No mais, MD pode ser capaz de: (a) prever induções conformacionais do receptor com o ligante, conformações essas que o receptor só adquire caso o ligante esteja presente (DEVIVO 2016); (b) buscar conformações ótimas, com a energia global mínima, para a realização da ancoragem molecular (PINIZ, 2018; CAPORUSCIO, 2016); (c) verificar se o ligante se manterá interagindo com o receptor (HUANG, 2011); (d) refinar as estruturas e compreender quais as interações que mais contribuem para a estabilização do ligante com o receptor (PINZI, 2019).

1.5.3) ANCORAMENTO MOLECULAR

O objetivo da ancoragem molecular (*docking* molecular) é o entendimento e a predição das interações entre diferentes compostos. Esta técnica permite sugerir sítios ativos de enzimas, sítios causadores de efeitos alostéricos, energias de interação e afinidades ligante-receptor. Além de prever potenciais conformações de um ligante em um sítio ativo, o *docking* é capaz de ranquear os ligantes de acordo com sua interação com o alvo. Portanto, é possível criar hipóteses, do ponto de vista estrutural, como um receptor interage com uma molécula, a conformação e orientação de um ligante dentro de uma região do receptor e estipular as interações que possivelmente estão estabilizando o sistema. Caso a região de interação seja desconhecida, é possível utilizar metodologias e programas para prever o sítio ativo de interação mais plausível (MENG, 2011; GUEDES, 2014; TRIPATHI, 2017).

No desenvolvimento de novos fármacos, o *docking* auxilia em diversas aplicações, como descoberta de novas moléculas, otimização de compostos, hipóteses para mecanismos de ação, triagem virtual¹⁵ e estudos de mutagênese. Com isso, ocorre a diminuição de custos, de

¹⁵ Do inglês, *virtual screening*.

mão de obra e tempo na pesquisa com novos fármacos, uma vez que indica e prioriza os possíveis melhores candidatos. Esta técnica, porém, não quantifica a interação de forma exata e não substitui as técnicas experimentais (PINZI, 2019).

As metodologias e os protocolos de *docking* podem considerar o receptor e o ligante rígidos; o receptor rígido e o ligante flexível; e um sistema do qual ambas as partes são flexíveis, ou parcialmente flexíveis (TORRES, 2019). Em sistemas que possuem mudanças de conformações induzidas pelo ligante, utilizar métodos flexíveis permitirá maior probabilidade de encontrar a conformação de menor energia entre ligante-receptor (MENG, 2011). Os sistemas que utilizam maiores graus de liberdade possuem a vantagem de conseguir incluir melhor o efeito do solvente nas interações e permitir todos os movimentos dos átomos possíveis durante a simulação (GUEDES, 2014). Os cálculos com conformações rígidas, com menores graus de liberdade, demandam um tempo de processamento menor. Consequentemente, caso o usuário insira o ligante em uma posição inadequada, não será possível extrair informações a respeito das interações que melhores se adequam ao sistema estudado e cabe ao mesmo avaliar a validade das inferências. Os algoritmos de *docking*, então, exercem funções quantitativas de predição de energias de ligação de maneira a convergir à energia mínima do sistema. (FERREIRA, 2016; PAGADALA, 2017).

A escolha da metodologia de ancoragem molecular a ser aplicada é aquela que, após devidamente testada e parametrizada, seja adequada ao sistema estudado, além de recorrer a estratégias aplicadas em outros trabalhos que quando implementadas obtiveram resultados satisfatórios (GUEDES, 2014). Algumas técnicas, como a cristalografia de proteínas, RMN, *small-angles X-ray scattering* (SAXS) e *high-resolution cryo-electron microscopy* (cryo-EM) auxiliam na validação do *docking* fornecendo informações de mudanças no deslocamento químico dos átomos do receptor que estão participando da interação com ligante e formas das nuvens eletrônicas oriundas do complexo ligante-receptor (CIEMNY, 2018).

A re-ancoragem, do inglês *re-docking*, pode ser utilizada para validar a metodologia e se baseia em realizar, *in silico*, o *docking* de uma estrutura terciária tridimensional já resolvida do receptor interagindo com o seu ligante. Desta forma, é possível observar se o modelo simulado corresponde ao observado experimentalmente (SHNEIDER, 2012; UCHIKOGA, 2013). O valor de RMSD entre as poses, que são as conformações adquiridas pela molécula, de menor energia pode ser utilizado como método de validação ao comparar com a pose modelo de cristalografia já existente (MORRIS, 2008).

Na ancoragem molecular é necessário a estrutura tridimensional, tanto do receptor quanto do ligante, e, caso não esteja disponível, as técnicas de modelagem comparativa podem ser aplicadas e as metodologias pós-*docking* podem ser utilizadas para validar as simulações. Os sistemas envolvendo a modelagem e *docking* molecular de proteínas que interagem com DNA, tal como o domínio HIRAN, já foram implementadas na literatura (TARAZI, 2016; PARULEKAR, 2013).

2) HIPÓTESE

Neste trabalho hipotetizamos que os domínios HIRAN das proteínas Rad5 e HLTF possuem diferenças na afinidade com a ssDNA, fazendo com que o complexo de estabilização da forquilha parada ocorra de forma diferenciada, interferindo no tempo necessário para sinalizar a via de reparo da forquilha de replicação reversa e a via de TLS.

3) OBJETIVOS

O objetivo do trabalho foi estudar, *in silico*, as variações moleculares do domínio HIRAN das proteínas HLTF e Rad5 em seus estados de conformação livre e complexado com DNA. Além disso, tentar correlacionar, evolutivamente, se a redundância de função da HLTF, no mecanismo de reconhecimento da forquilha de replicação parada pela SHPRH, está relacionada com a estrutura terciária do HIRAN.

3.1) OBJETIVOS ESPECÍFICOS

- Construir o modelo tridimensional do domínio HIRAN da proteína Rad5;
- Comparar os comportamentos estruturais dos domínios HIRAN das proteínas HLTF e Rad5, nas suas conformações livre e de interação à ssDNA;
- Extrair informações estruturais dos domínios HIRAN da Rad5 e HLTF quando há a presença da ssDNA no sítio de interação;
- Analisar as interações, entre resíduos de aminoácidos e nucleotídeos, que estabilizam os sistemas envolvendo o domínio proteico e o material genético.

4) METODOLOGIA

4.1) CONSTRUÇÃO DO MODELO scRad5

O programa MODELLER v9.22 foi utilizado para realizar a modelagem comparativa para obtenção do modelo tridimensional do domínio HIRAN da proteína Rad5 de *Saccharomyces Cerevisiae* (scRad5), a partir da busca por estruturas homólogas utilizando a plataforma BLAST para o alinhamento local com o banco de dados PDB.

O alinhamento da estrutura primária da scRad5 (GenBank ID: CAA97556.1) com o modelo 5K5F, domínio HIRAN da HLTF, foi feito com o *script alignment.salign*¹⁶() e a geração dos modelos com o *script automodel*¹⁷() descritos no manual do programa. A sequência proteica do domínio scRad5 foi obtida pela plataforma Pfam (FINN, 2014), por meio do código de acesso PF08797.

Foram calculadas e geradas 25.000 estruturas, utilizando novamente como referência o modelo com PDB ID de 5K5F e selecionada a estrutura de menor energia e menor DOPE *score* (SHEN, 2006). Esta estrutura teve sua energia global minimizada utilizando o programa CHIMERA v1.13.1 a partir do campo de força AMBER ff99SB com parâmetros de 100 *steepest descent steps*, 0.02 *steepest descent step* (Å), 10 *conjugate gradient steps*, 0.02 *conjugate gradient step size* (Å) e 10 *update interval*.

Posteriormente somente a estrutura após a minimização de energia foi submetida a plataforma SAVES v5.0 para análises de validação estrutural a partir dos programas ERRAT e PROCHECK e selecionada para ser utilizada nas etapas de dinâmica e *docking* molecular.

4.2) DINÂMICA MOLECULAR

O programa GROMACS v2018.3 foi utilizado para realização das dinâmicas. As estruturas dos domínio HIRAN utilizadas foram: (a) domínio HIRAN da HLTF em solução salina calculada a partir de dados de RMN, na sua conformação livre (PDB ID: 5K5F), com e sem uma fita simples de DNA poli-T; (b) Modelo obtido pela modelagem comparativa scRad5 com e sem uma fita simples de DNA poli-T; (c) estrutura resolvida a partir da cristalografia e difração de raios-x do HIRAN da HLTF, em sua conformação complexada com ssDNA, com e sem uma fita simples de DNA poli-T. A deleção dos átomos que representam os nucleotídeos foi feita utilizando o programa CHIMERA para a realização da etapa de produção dos sistemas sem material genético. Para todos os sistemas sem ssDNA foi utilizado o campo de força

¹⁶ Disponível em: <https://salilab.org/modeller/9v1/manual/node263.html#CMD:alignment.align>

¹⁷ Disponível em: <https://salilab.org/modeller/9v8/manual/node42.html>

AMBER99SB-ILDN e para os sistemas com ssDNA o campo de força AMBER14SB, que contêm os parâmetros do campo de força ff99bsc0 necessários para simulações com nucleotídeos de DNA. Todas as estruturas e sistemas foram preparados utilizando os mesmos critérios e algoritmos.

Tabela 1: Resumo das estruturas utilizadas nas dinâmicas moleculares.

Estruturas	Replicatas	Tempo (ns)	Campos de Força
HIRAN da HLTF conformação livre sem ssDNA	3	500	AMBER99SB-ILDN
HIRAN da Rad5 conformação livre sem ssDNA	3	500	AMBER99SB-ILDN
HIRAN da HLTF complexada sem ssDNA	1	500	AMBER99SB-ILDN
HIRAN da HLTF conformação livre com ssDNA	1	500	AMBER14SB
HIRAN da Rad5 conformação livre com ssDNA	1	500	AMBER14SB
HIRAN da HLTF complexada com ssDNA	1	500	AMBER14SB

A estrutura do domínio HIRAN da HLTF na sua conformação livre é representada pelo modelo de PDB ID 5K5F e calculada a partir de dados de RMN. A estrutura do domínio HIRAN da Rad5 foi obtida pela modelagem comparativa em sua conformação livre. A estrutura do domínio HIRAN da HLTF complexada é representada pelo modelo de PDB ID 4S0N e resolvida a partir da cristalografia e difração de raios-x. NS é a contração para nanosegundos.

4.2.1) TOPOLOGIA DOS ARQUIVOS

Os arquivos *.pdb* obtidos do banco de dados PDB tiveram a sua topologia e coordenadas atômicas geradas a partir do comando *pdbgmx* do próprio GROMACS, sendo que para a estrutura derivada de RMN e da modelagem comparativa (cadeia A #0.1 do *NMR ensemble* PDB ID 5K5F e scRad5) foi preciso utilizar a *flag -igh* pois já continham os átomos de hidrogênios. A protonação dos resíduos do modelo scRad5 foi realizada pelo programa Modeller e verificada utilizando o programa H++ (GORDON, 2005) no pH 6.8, de acordo com o pH de obtenção da estrutura molde de RMN. O modelo de água utilizado foi o TIP3P (*Transferable Intermolecular Potential, three-point*), modelo recomendado para o campo de força empregado.

4.2.2) CAIXA DE SIMULAÇÃO

O comando *editconf* foi utilizado para criar a caixa de simulação cúbica e com as estruturas centralizadas no centro da caixa. Em seguida, a caixa foi solvatada com modelos de água TIP3P com o comando *solvate*. Os programas *genion* e *grompp* (contidos no pacote GROMACS) foram utilizados para neutralizar o sistema com íons de Sódio e Cloro. A quantidade e íons adicionada depende da carga total do sistema e foi utilizado a *flag -neutral* para o sistema ser neutralizado (somatório de cargas igual a 0). A caixa contém parâmetros de PBC nas três dimensões e as interações coulômbicas tratadas pelo algoritmo PME com valores de corte de 1 nm para as interações de curta distância do tipo eletrostática e de van der Waals. As estruturas do domínio HIRAN, com e sem a presença de ssDNA, foram centralizadas na caixa à 1 nm de distância do centro de cada arestas.

4.2.3) MINIMIZAÇÃO DE ENERGIA DO SISTEMA

Foi utilizado o programa *grompp* e o comando *mdrun* para realizar a minimização de energia do sistema para evitar maus contatos das interações das moléculas de água e dos íons adicionados com as estruturas do sistema. A minimização se deu a partir do algoritmo *steepest descent* e parou quando a força máxima do sistema foi menor que 1000.0 kJ/mol/nm. Como parâmetros, foi utilizado PBC, interações coulômbicas tratadas por PME, e 1 nm de valor de corte para as interações de curta distância.

4.2.4) EQUILIBRAÇÃO DO SISTEMA

A equilibração do sistema foi feita utilizando os *ensembles* NVT e NPT, dos quais se mantém o número de partículas, volume, temperatura (somente em NVT) e pressão (somente em NPT) do sistema constantes. Ambas equilibrações possuem tempo de duração de 300 ps, integração coulômbica do tipo PME, 1 nm de valor de corte para interações de curta distância, PBC em três dimensões e forças harmônicas restritivas de $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. O algoritmo de termostato utilizado foi o *V-rescale* (Berendsen modificado) com uma temperatura de referência de 300 K. O algoritmo de acoplamento de pressão foi o barostato de Parrinello-Rahman, do tipo isotrópico e com valor de compressibilidade isotérmica da água de 1 bar^{-1} .

4.2.5) ETAPA DE PRODUÇÃO

A dinâmica molecular utilizou todos os arquivos e parâmetros gerados nas etapas anteriores, com um integrador de tempo de 2 fs (femtosegundos) e gravando a cada 50 ps (picosegundos) as trajetórias, velocidades e energias dos átomos. As análises das trajetórias foram extraídas utilizando os comandos do pacote GROMACS (*rms*, *trjconv*, *gyrate*, *rmsf*,

make_ndx e *pairdist*) e os gráficos criados pelo programa GRACE v1:5.1.25-1build2. Os valores do eixo Y dos gráficos de RMSF foram convertidos em *b-factor* pelo programa GROMACS a partir da utilização da *flag -oq* no comando *rmsf*. Estes valores foram renderizados na forma de cores na estrutura do modelo tridimensional inicial de cada modelo pela ferramenta *Render by Attribute* do CHIMERA.

As coordenadas espaciais atômicas durante todo o curso da dinâmica, assim como suas interações, são escritas em arquivos gráficos de trajetória para serem analisadas posteriormente. Algumas análises são importantes para o entendimento da dinâmica, dentre elas: (a) *Root Mean Square Deviation* (RMSD), que analisa a variação de distância média dos átomos em relação a sua posição inicial a cada intervalo de tempo, possibilitando inferir o momento de estabilidade do sistema; (b) *Root Mean Square Fluctuation* (RMSF), que diferente do RMSD, analisando as flutuações de movimento dos átomos individualmente, indicando aquele que mais contribuiu para as diferenças no RMSD. O RMSF pode ser utilizado para comparar estruturas selvagens com mutantes, observando se a mobilidade/flexibilidade do resíduo (ou da região) é afetada; (c) Interações de contato e eletrostáticas, as quais são de extrema importância para a estabilidade e as formações de estruturas secundárias e terciárias de uma proteína, por exemplo. Normalmente, as estruturas secundárias proteicas são calculadas a partir de padrões nas interações de hidrogênio (SNEHA, 2016).

As ferramentas de agrupamento permitem análises conformacionais que o sistema apresenta em determinados tempos da simulação (CHENG, 2012). O algoritmo de agrupamento de estruturas representativas utilizado foi o GROMOS (DAURA, 1998) a partir do comando *gmx cluster* com a *flag method*. Este algoritmo utiliza os valores de RMSD para agrupar as estruturas semelhantes a partir de valores de corte (**Tabela 2**), que foram escolhidos de forma a obter um número representativo de estruturas.

Tabela 2: Valores de cortes utilizados pelo algoritmo de agrupamento

Modelo	Valor de Corte (nm)	Número de Estruturas
5K5F Réplica 1	0,25	8
5K5F Réplica 2	0,25	7
5K5F Réplica 3	0,25	7
scRad5 Réplica 1	0,55	7
scRad5 Réplica 2	0,45	6
scRad5 Réplica 3	0,45	6

4S0N Livre	0,20	6
4S0N ssDNA **	1,20	6
5K5F ssDNA *	0,30	6
5K5F ssDNA **	0,30	5
scRad5 ssDNA *	0.35	7
scRad5 ssDNA **	0,275	7

*Utilizando os átomos principais como referência e ** utilizando a ssDNA

As análises de contatos foram realizadas utilizando a ferramenta *cpptraj*¹⁸, do pacote de dinâmica molecular do AMBER. Foram utilizados os comandos *nativecontacts*, para obter os contatos nativos e não-nativos, e *maxdist distance*, para estabelecer uma distância máxima de 4 Å. Os contatos nativos são aqueles presentes inicialmente no sistema, enquanto os não-nativos são aqueles adquiridos ao longo da etapa de produção (CASE, 2018).

4.3) DOCKING MOLECULAR

Para realização do *docking* molecular, foi utilizado o programa UCSF DOCK v6.9 e o programa CHIMERA, com o campo de força AMBER FF14SB, para o preparo dos arquivos e das estruturas a serem utilizados. As etapas subsequentes são equivalentes ao protocolo para o programa UCSF DOCK6, outros programas de *docking* podem não realizar os mesmos passos.

4.3.1) GERAÇÃO DOS ARQUIVOS ESTRUTURAIS DO RECEPTOR E LIGANTE

Os receptores, estrutura terciária do domínio HIRAN, foram preparados a partir da ferramenta *Dock Prep* do CHIMERA. Foram gerados 2 arquivos em extensão *.mol2* para cada estrutura: (a) um arquivo contendo os receptores com o todos os átomos de hidrogênios, sendo que a estrutura de cristalografia 4S0N teve a deleção de todas as cadeias duplicadas e do ssDNA, restando somente a cadeia “A”. A adição de hidrogênios foi feita pela ferramenta *Add H* do CHIMERA e verificada pelo programa H++; (b) um arquivo contendo os receptores sem hidrogênio, que foram removidos por meio do programa CHIMERA; O ligante teve adição de cargas em sua estrutura partir da ferramenta *Add Charge* do CHIMERA, computadas pelo ANTECHAMBER, e utilizando como molde de ssDNA a cadeia “G” poli-T do modelo 4S0N.

4.3.2) SUPERFÍCIE MOLECULAR

A geração da superfície molecular dos receptores foi feita utilizando a ferramenta *Write DMS* do CHIMERA, em extensão *.dms*, a partir da estrutura que não possui hidrogênios. Em

¹⁸ Disponível em: <https://amber-md.github.io/cpptraj/PPTRAJ.xhtml>

seguida, as esferas foram geradas pelo programa *sphgen* do pacote DOCK utilizando o arquivo *.dms*, e como parâmetros a superfície externa do receptor, todos os pontos da superfície molecular, raio máximo das esferas de 4 Å e mínimo de 1,4 Å. O *cluster* de maior número de esferas e que melhor representava o sítio de ligação foi selecionado e convertido em *.pdb* pelo programa *showsphere* do pacote DOCK.

4.3.3) CRIAÇÃO DA CAIXA DE INTERAÇÃO

Foi criada uma caixa em volta do sítio de interação para ocorrer o cálculo do *grid*, região do receptor da qual o programa irá considerar os resíduos para fazer os cálculos necessários. A caixa, em formato cúbico, foi criada com arestas de 4 nm e seu centro com o mesmo vetor de posição do centro do grupo de esferas selecionado. Em seguida, o cálculo do *grid* foi realizado com o programa *grid* do pacote do DOCK, com os parâmetros padrões, exemplificados de acordo com um tutorial disponível no site dos desenvolvedores.

4.3.4) ETAPA DE PRODUÇÃO

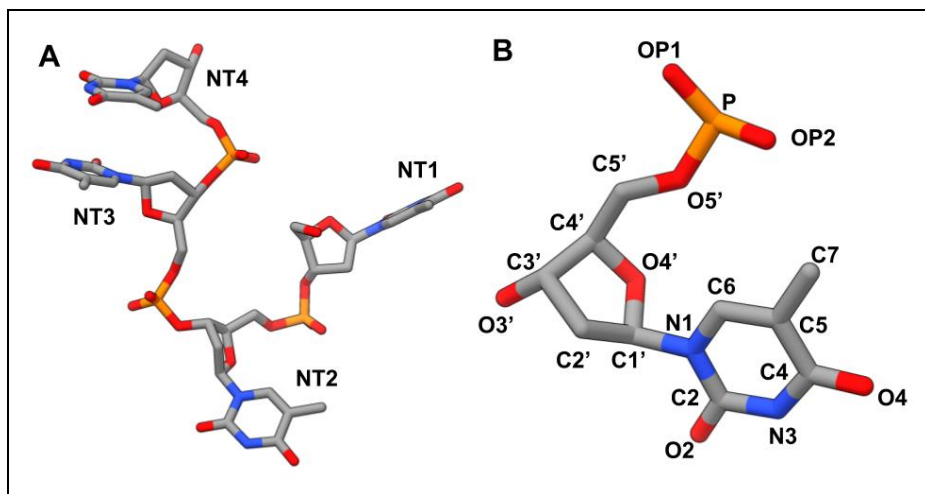
O *docking* foi realizado utilizando o programa *dock6* do pacote DOCK, considerando o ligante flexível, o receptor rígido, calculando 100 conformações, e as 10 poses que apresentaram menor energia global foram selecionadas para geração do *footprint rescore*, e parâmetros padrões explicados no tutorial disponível no site dos desenvolvedores¹⁹.

4.4) MODELO DE ssDNA UTILIZADO

Para as simulações de dinâmica molecular e nas aplicações de *docking* molecular do domínio HIRAN com a fita simples de DNA, foi utilizado a cadeia G da estrutura com PDB ID de 4S0N (**Figura 9**). Essa cadeia representa uma sequência de timinas (Poli-T) com 4 nucleotídeos na conformação complexada com o sítio de interação da HLTF.

¹⁹ Disponível em:
http://dock.compbio.ucsf.edu/DOCK_6/tutorials/ligand_sampling_dock/ligand_sampling_dock.html

Figura 9: Estrutura e nomenclatura atômica da ssDNA.



Representação tridimensional do modelo de ssDNA utilizado das etapas de produção da dinâmica e *docking* molecular. Estrutura retirada do PDB sob o código de acesso 4S0N, cadeia G. Em cinza, estão representados os átomos de carbono, em laranja, os átomos de fósforo, em vermelho, os átomos de oxigênio e em azul, os átomos de nitrogênio. Os oxigênios ligados ao fósforo recebem nomenclatura de OP, ao açúcar de O' e a base nitrogenada de O. Os átomos de carbono do açúcar recebem a nomenclatura C'. NT é a abreviação de nucleotídeo.

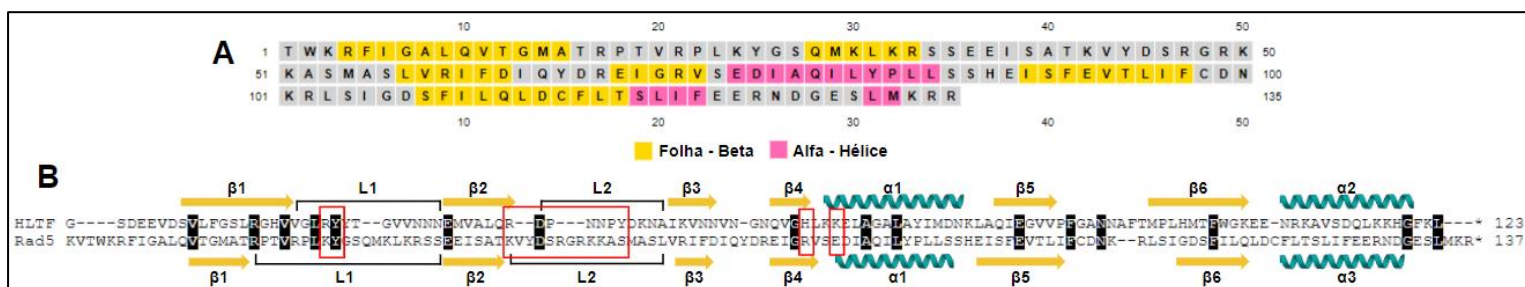
5) RESULTADOS E DISCUSSÃO

5.1) O DOMÍNIO HIRAN DA RAD5 É MAIS EXTENSO NAS REGIÕES DE *LOOPS*

A plataforma BLASTp foi utilizada para buscar proteínas com estruturas terciárias de maior identidade com a sequência do domínio HIRAN de *Saccharomyces cerevisiae* com filtro na plataforma do Protein Data Bank (PDB). Alguns trabalhos como de Rejik e colaboradores (2015), para a proteína MRE11, e de Turchick e colaboradores (2015), para a proteína Rad51, utilizaram esta mesma abordagem para identificar estruturas de proteínas envolvidas nas interações com o DNA. As abordagens que combinam a ferramenta BLAST, utilizando os valores de identidade e similaridade, para realizar o alinhamento e selecionar o modelo no PDB, são bastantes difundidas na literatura (PARULEKAR, 2013; NOCUA, 2014; BAVI, 2016; TARAZI, 2016; PETERSON, 2019).

O banco de dados PDB tem 7 estruturas do domínio HIRAN depositadas, as quais são referentes à proteína HLTf: (a) 5K5F, estrutura calculada a partir de dados de RMN em solução salina na forma apo; (b) 4S0N, estrutura resolvida a partir da cristalografia e difração de raios-x na forma holo, complexada com ssDNA com resolução de 1,5 Å; (c) 4XZG estrutura resolvida a partir da cristalografia e difração de raios-x na forma apo, sem ssDNA; (d) 5BNH, estrutura resolvida a partir da cristalografia e difração de raios-x na forma holo, com resolução de 1,7 Å; (e) 2MZN, estrutura calculada a partir de dados de RMN na conformação de ligação ao ssDNA; (f) 4XZF, estrutura resolvida a partir da cristalografia e difração de raios-x na forma holo, complexada com DNA fita dupla e resolução de 1,38 Å; (g) 6KCS estrutura resolvida a partir da cristalografia e difração de raios-x na forma holo, complexada com fita dupla de DNA e resolução de 2,1 Å. O modelo 5K5F (hsHLTF_{APO}) foi selecionado para ser utilizado como referência na modelagem comparativa, do domínio HIRAN da proteína Rad5 (scRad5), por ser uma estrutura de conformação livre e em solução salina. Esta etapa garante que possamos estudar e analisar, de forma comparativa, as estruturas dos domínios sem a interferência do ligante. O alinhamento demonstrado na **Figura 10** foi realizado pelo programa Modeller e demonstra uma porcentagem de identidade de 12,50% entre a sequência do domínio HIRAN da proteína HLTf com o da proteína Rad5. Apesar da baixa identidade, que se quantifica pelo número de resíduos idênticos, os domínios apresentam uma similaridade de 51,47%, quantificada a partir de resíduos que possuem cadeia lateral de propriedades físico-químicas similares. Além disso, compartilham alguns resíduos chaves para a interação com a ssDNA e resíduos com similaridade nas regiões do possível sítio de interação (**Figura 10 B**).

Figura 10: Predição de estrutura secundária do domínio HIRAN da Rad5 e alinhamento com a sequência do domínio da proteína HLTf.



Em A, predição da estrutura secundária obtida a partir do programa PSIPRED sendo em rosa a predição de alfa-hélice e em amarelo a de folha-beta. Em B, alinhamento utilizado para a construção do modelo scRad5 utilizando o domínio HIRAN da HLTf como referência. Na região superior e inferior do alinhamento estão representadas as estruturas secundárias do modelo tridimensional dos modelos (em ciano, representação de alfa-hélice, em amarelo, de folha beta). Resíduos idênticos estão destacados em preto e os resíduos que interagem com a ssDNA, em humanos, estão destacados por retângulos vermelhos.

A predição de estruturas secundárias (**Figura 10 A**) foi realizada utilizando o programa PSIPRED (JONES, 1999) e demonstrou um padrão de distribuição de alfa-hélice e folha-beta similar com a estrutura obtida a partir da modelagem comparativa. Porém, a região e o número de resíduos nestas posições apresentaram diferenças: (a) o número de resíduos presentes na primeira folha-beta ($\beta 1$) foi superior ao do modelo tridimensional; (b) a região da segunda folha-beta ($\beta 2$) predita pelo PSIPRED se encontra na região do loop 1 (L1); (c) a região da segunda alfa-hélice ($\alpha 2$) apresentou um número de resíduos menor e fragmentada em 2 regiões quando comparada com a estrutura terciária calculada pelo Modeller. Estes dados demonstram a diferença de como as definições de interações que ocorrem no enovelamento de uma estrutura terciária são calculadas, o que pode resultar nas diferenças das predições de estruturas secundárias a partir da sequência de aminoácidos. Iyer e colaboradores (2006), utilizando o programa JPRED, predisseram que o domínio HIRAN da HLTf era composto por 8 folhas-beta sendo que sua estrutura terciária se apresenta em outra conformação (**Figura 13 A**).

Os cálculos de estruturas pelo Modeller baseiam-se tanto na identidade das sequências quanto na similaridade para realizar a construção do modelo. A utilização do parâmetro de DOPE score nos permite selecionar uma estrutura de confiança, uma vez que átomos em posições irregulares, sobreposições de nuvens eletrônicas e ângulos diedros em discrepância com o campo de força aumentam este valor. Portanto, após a seleção do modelo de menor energia, a estrutura foi minimizada pelo programa CHIMERA e submetida a avaliação dos ângulos *phi* e *psi* pelo PROCHECK (**Figura 11**) e das interações inter-residuais pelo ERRAT (**Figura 12**). As análises dos ângulos diedros demonstraram a presença de 99 resíduos (79,8%) em regiões mais favorecidas, 23 resíduos (18,5%) em regiões adicionais permitidas, 2 resíduos

(1,6%) em regiões permitidas generosamente e nenhum resíduo em regiões desfavoráveis. Os 2 resíduos encontrados em regiões permitidas generosamente estão posicionados em regiões de *loop* na estrutura terciária. Os resíduos nestas regiões possuem um maior grau de liberdade, pois possuem um menor número de interações atômicas com resíduos vizinhos e suas cadeias laterais externalizadas para o solvente. As pontuações observadas no gráfico ERRAT demonstraram que somente resíduos em posições de *loop* apresentam um valor de pontuação elevado e ultrapassam a região de alerta. Estes dados suportam, a partir dos cálculos estatísticos e comparativos com os bancos de dados de estruturas de proteínas, uma estrutura que possa representar a conformação do domínio HIRAN da proteína Rad5 em solução.

Figura 11: Análises angulares *phi* e *psi* do modelo scRad5.

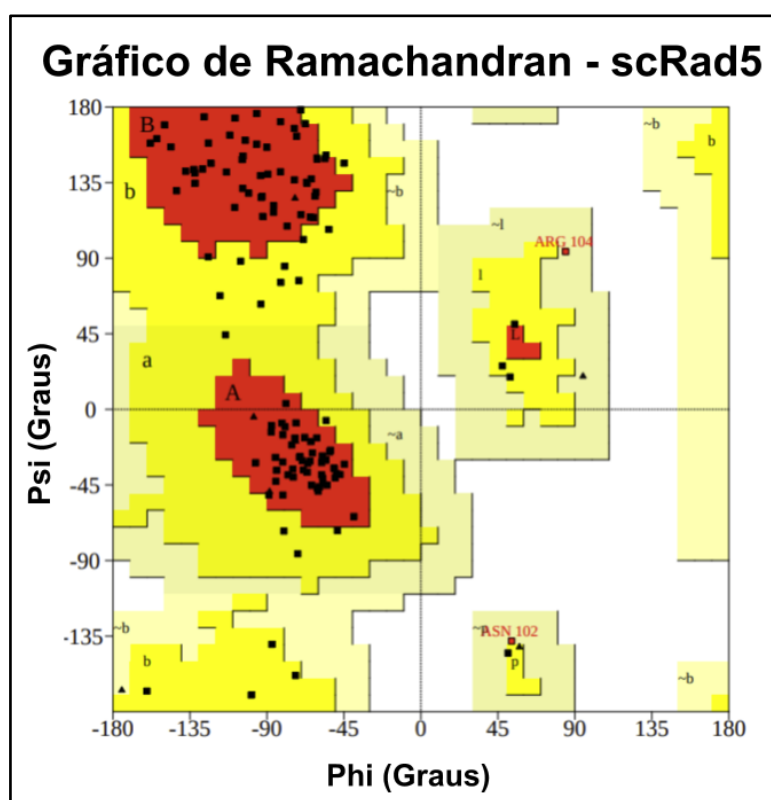


Gráfico de Ramachandran gerado pelo programa PROCHECK. Em vermelho está representado as regiões mais favorecidas, em amarelo escuro as em regiões adicionais permitidas, em amarelo claro as regiões permitidas generosamente e em branco as regiões desfavoráveis. A numeração dos resíduos está normalizada, sendo Arg104 a Arg275 e a Asn102 a Asn273 na estrutura primária da Rad5.

Figura 12: Análises das interações inter-residuais do modelo scRad5.

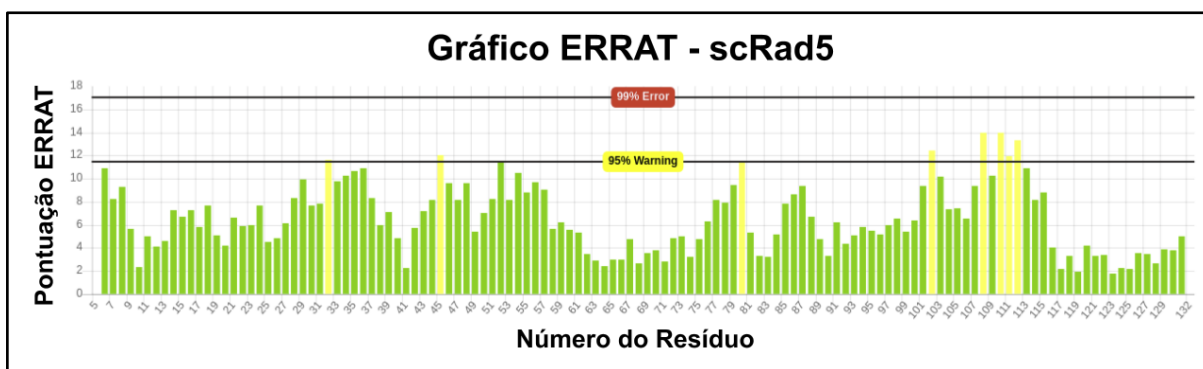
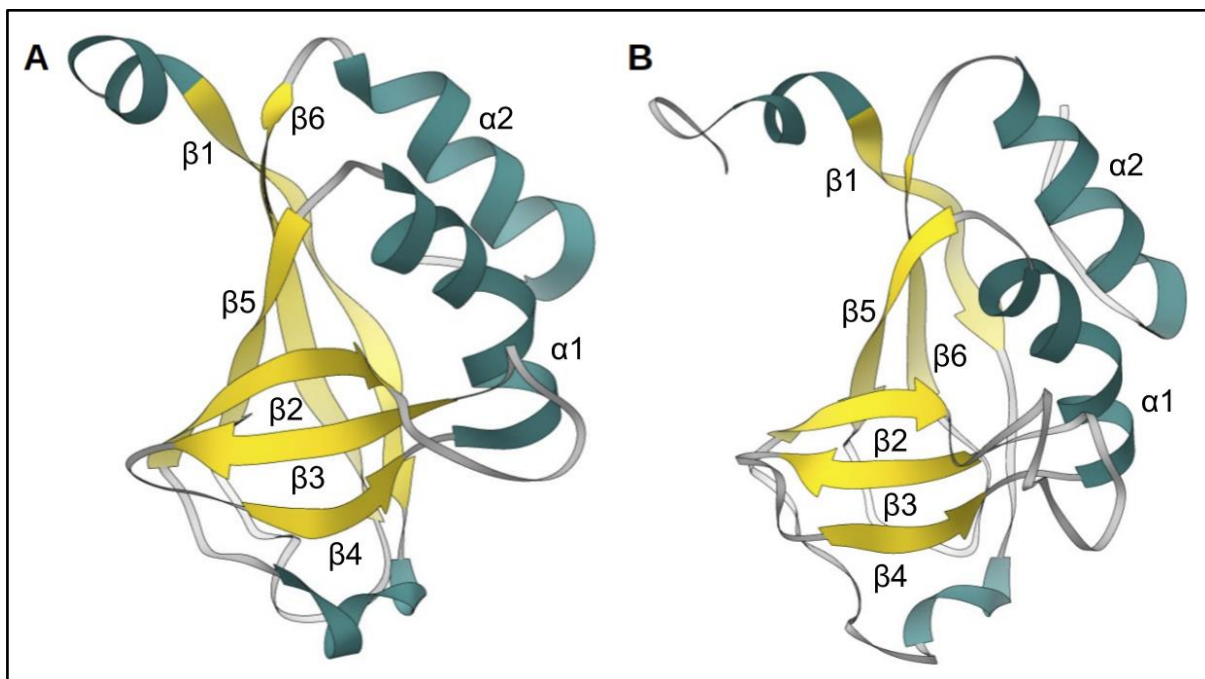


Gráfico de análises das interações inter-residuais gerado pelo programa ERRAT. Em verde está representado os resíduos que obtiveram uma pontuação abaixo do valor de corte de aviso e em amarelo os resíduos que obtiveram pontuação abaixo do valor de erro e acima de aviso. A numeração dos resíduos está normalizada, sendo o resíduo 1 o resíduo 169 da estrutura primária da Rad5.

A escolha da metodologia adotada para a modelagem comparativa gerou um modelo de confiança para ser utilizado em etapas posteriores, apesar da baixa percentagem de identidade, uma vez que foi validada estatisticamente. As proteínas possuem variações genéticas adquiridas ao longo da evolução a partir da seleção natural, porém, aquelas que compartilham um semelhante microambiente de ação molecular tendem a possuir estruturas e funções similares a partir de pressões e restrições ambientais (WORTH, 2009). Outros trabalhos, utilizaram o programa MODELLER para modelar proteínas que interagem com o DNA, como a proteína UvrB, envolvida no processo de reparo de excisão de nucleotídeos (BAVI, 2016), a proteína UvrC (PARULEKAR, 2013) e as proteínas DNA ligases NAD^+ dependentes (SHRIVASTAVA, 2015), mostrando a sua importância na aquisição de informações a respeito dessa classe de proteínas..

A estrutura terciária gerada para o modelo de scRad5 (**Figura 13 B**) apresenta 6 folhas betas e 2 alfas hélices, similar à estrutura do HIRAN da HLTF (hsHLTF_{APO}, **Figura 13 A**), além de 2 alfas hélices localizada na região N-terminal, uma região de *loop* e um valor de RMSD de 0,559 Å quando comparado ao molde. A estrutura apresenta, também, regiões de *loops* mais extensas, sugerindo que estes resíduos possam apresentar uma alta motilidade, uma vez que suas cadeias laterais não possuem restrições espaciais com resíduos adjacentes e suas superfícies moleculares estão expostas para o solvente.

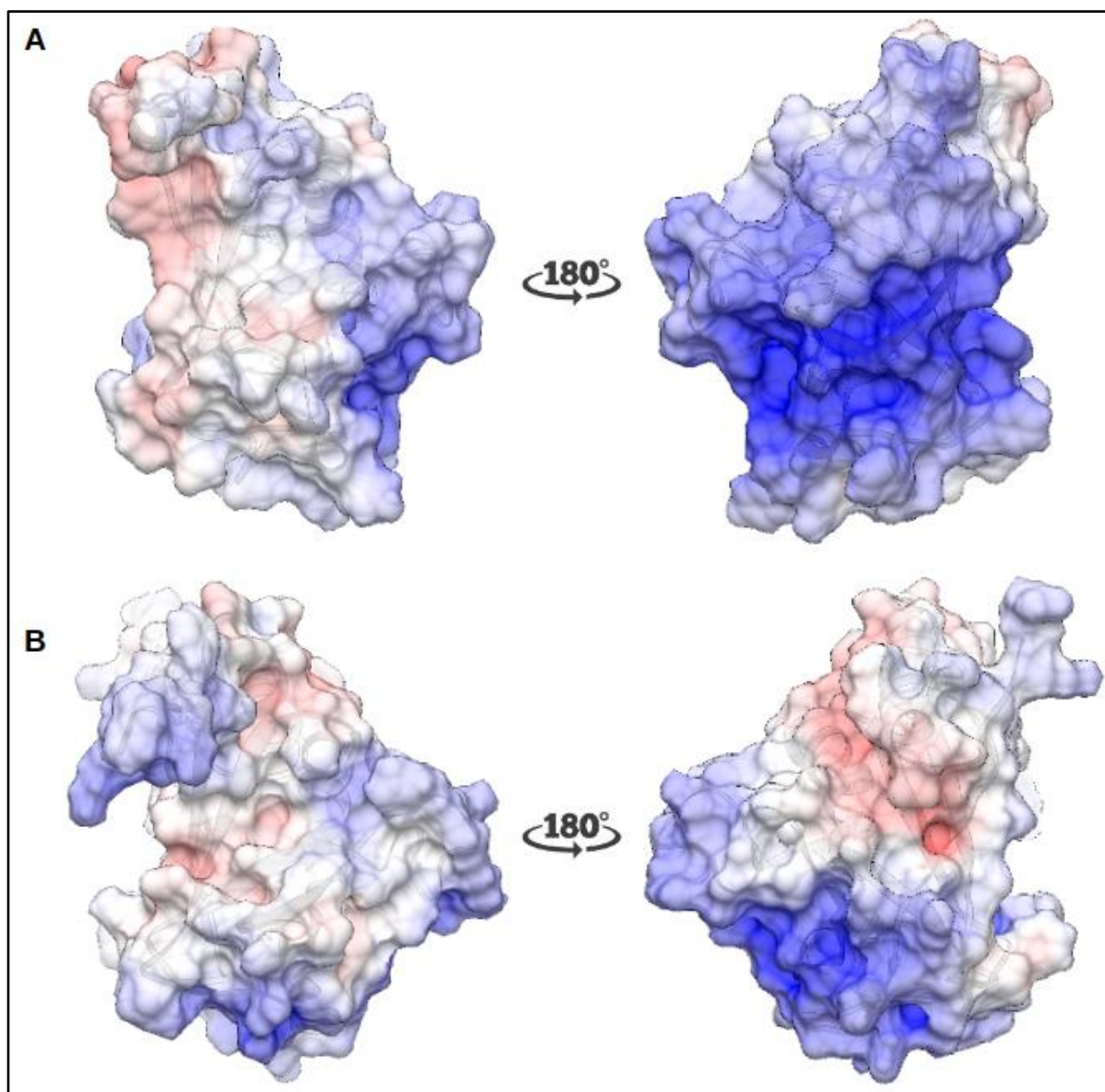
Figura 13: Estrutura terciária do domínio HIRAN de *Saccharomyces cerevisiae*.



Estruturas terciárias dos modelos hsHLTF_{APO} e scRad5. As estruturas possuem um valor de RMSD de 0,559 Å entre elas calculado pelo programa CHIMERA. Em A, modelo hsHLTF_{APO}, em B, estrutura terciária do modelo scRad5 obtida por modelagem molecular por homologia. Em amarelo estão representadas as estruturas de folha beta e em ciano as de alfa-hélice.

A estrutura terciária do domínio HIRAN da HLTf possui duas faces com somatória de cargas eletrostáticas opostas (**Figura 14 A**), sendo uma mais positiva e outra mais negativa (KILE, 2015). Para conferir se o modelo da proteína Rad5 possuía as mesmas características, foi feita uma análise de superfície eletrostática por meio da ferramenta APBS dentro do programa CHIMERA no valor de pH 7.0. Os resultados demonstraram que somente a região do suposto sítio de interação com o DNA possui uma concentração de resíduos de aminoácidos com cadeias laterais positivas, podendo hipotetizar que esta região possa realizar interações eletrostáticas com a cadeia de fosfatos do ssDNA (**Figura 14 B**). A estrutura terciária do hsHLTF_{APO} possui uma maior extensão de aminoácidos positivos em uma das faces do domínio quando comparado com o scRad5.

Figura 14: Superfícies moleculares da estrutura terciária do domínio HIRAN da HLTF e da Rad5.



Potencial eletrostático calculado pela ferramenta APBS do programa CHIMERA, do qual os valores variam de -10 a +10 kTe. Em vermelho os valores mais negativos, em branco os valores neutros e em azul os valores mais positivos. Em A, modelo hsHLTF_{APO} e, em B, modelo scRad5. Observar que em azul, em ambos os modelos, há uma concentração de resíduos positivos nas regiões de interação com ssDNA.

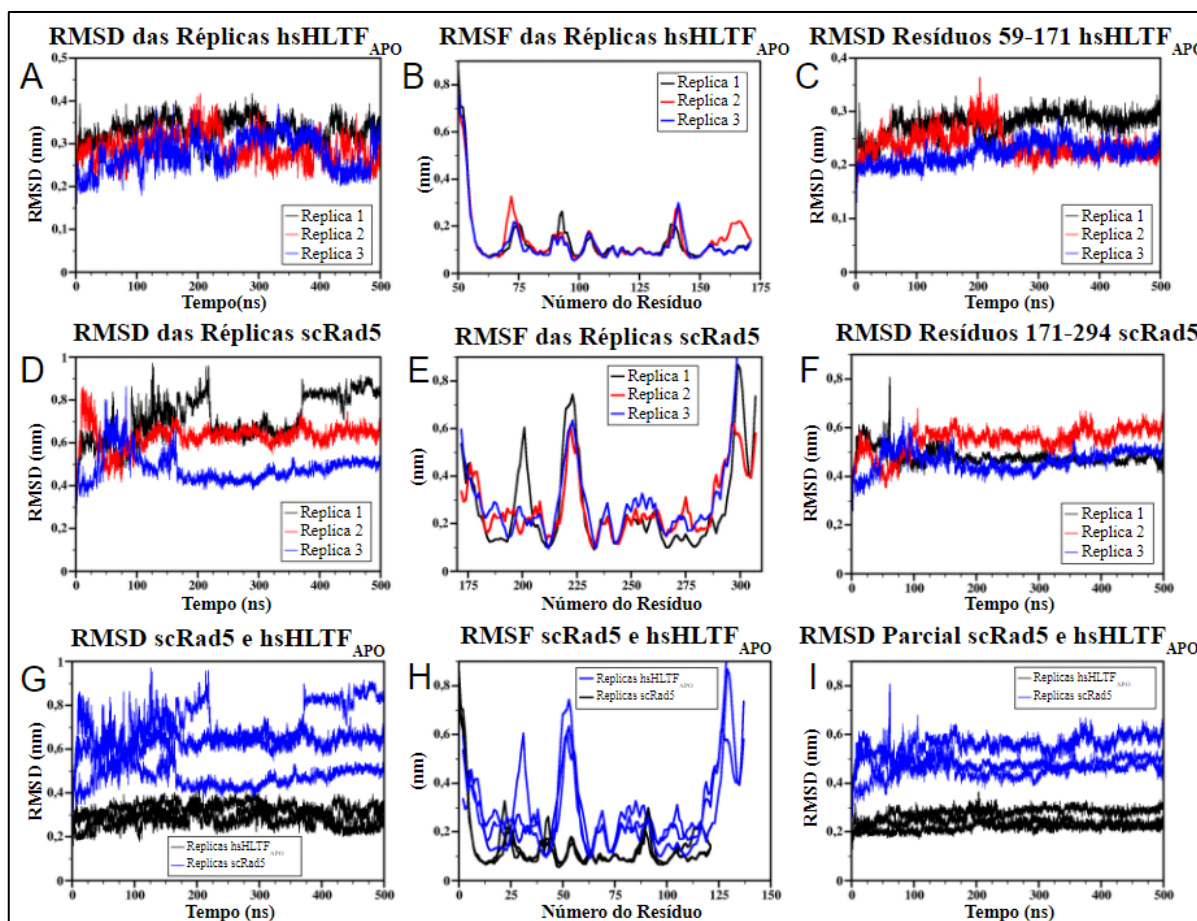
5.2 O DOMÍNIO HIRAN DA PROTEÍNA HLTF APRESENTA MENOS ALTERAÇÕES ESTRUTURAIS

A dinâmica molecular foi utilizada para avaliar e comparar a dinâmica estrutural dos domínios em solução aquosa. Todos os sistemas foram criados e executados nas mesmas condições para permitir inferir as características estruturais comparativas, *in silico*, dos modelos. Para ambos os modelos foram realizadas 3 replicatas de 500 nanosegundos cada, e

com diferentes estruturas iniciais geradas a partir do comando *gen_seed* -1 como argumento nos arquivos de preparação da dinâmica.

As análises das dinâmicas estruturais foram feitas a partir dos gráficos de RMSD e RMSF. O gráfico de RMSD do modelo hsHLTF_{APO}, considerando todos os resíduos (**Figura 15 A**) mostra uma estrutura com certa instabilidade, uma vez que a variação das distâncias dos átomos não se estabilizou. A fim de observar as variações de posições dos resíduos de forma individual, o gráfico de RMSF (**Figura 15 B**) foi extraído do arquivo de trajetória e observado que a região dos resíduos 50-59 (região N terminal) não apresentou uma estrutura secundária e possuiu uma alta mobilidade. Em seguida, foi traçado o gráfico de RMSD somente dos resíduos 59 ao 171 (**Figura 15 C**) e observado que o sistema se encontrava em equilíbrio e que o tempo da dinâmica foi suficiente para explorar as diferentes conformações do modelo. O gráfico de RMSD de todos os resíduos do modelo scRad5 (**Figura 15 D**) demonstrou estabilidade a partir de 200 ns de simulação, porém com diferentes variações entre as réplicas. Similar ao motivo do modelo hsHLTF_{APO}, o gráfico de RMSF foi traçado (**Figura 15 E**) e observado uma desnaturação da região C-terminal das réplicas. Em seguida o gráfico de RMSD dos resíduos 171 ao 294 (região sem o C terminal; **Figura 15 F**) apresentou estruturas com uma dinâmica estrutural mais próximas e um maior equilíbrio dos sistemas. Estes gráficos demonstram que as regiões N e C-terminal possuem uma alta motilidade, porém, estes dados podem ser um artefato gerado pela metodologia. É possível que este fenômeno não seja aplicável no contexto biológico, uma vez que o domínio HIRAN possui domínios adjacentes.

Figura 15: Dinâmicas estruturais dos modelos hsHLTF_{APO} e scRad5.



Gráficos de RMSD e RMSF dos modelos hsHLTF_{APO} e scRad5 ao longo de 500 ns de simulação. Em A e D estão representados os gráficos de RMSD de todos os resíduos, em B e E os gráficos de RMSF para identificação de movimentos de resíduos individuais e em C e F, o gráfico de RMSD com exclusão dos resíduos com desnaturação e alta mobilidade. De A a C está representado o modelo hsHLTF_{APO} e de D a F o modelo scRad5. De G a I estão representadas as sobreposições dos gráficos de RMSD total, RMSF e RMSD parcial de todas as réplicas de cada modelo, respectivamente. De A a F, as cores pretas representam as Réplicas 1, as vermelhas as Réplicas 2 e as azuis as Réplicas 3. De G a I, as cores pretas representam o modelo hsHLTF_{APO} e as azuis o modelo scRad5.

A sobreposição dos gráficos de RMSD total e parcial de ambas as réplicas dos modelos (**Figura 15 G e I; Tabela 2**) demonstraram que a estrutura terciária do modelo scRad5 possui maior variação nas posições dos resíduos do que o hsHLTF_{APO}. Isso pode ser devido ao número superior de resíduos em regiões de *loops*, que possuem maior grau de liberdade de movimento, e à desestruturação de estruturas secundárias.

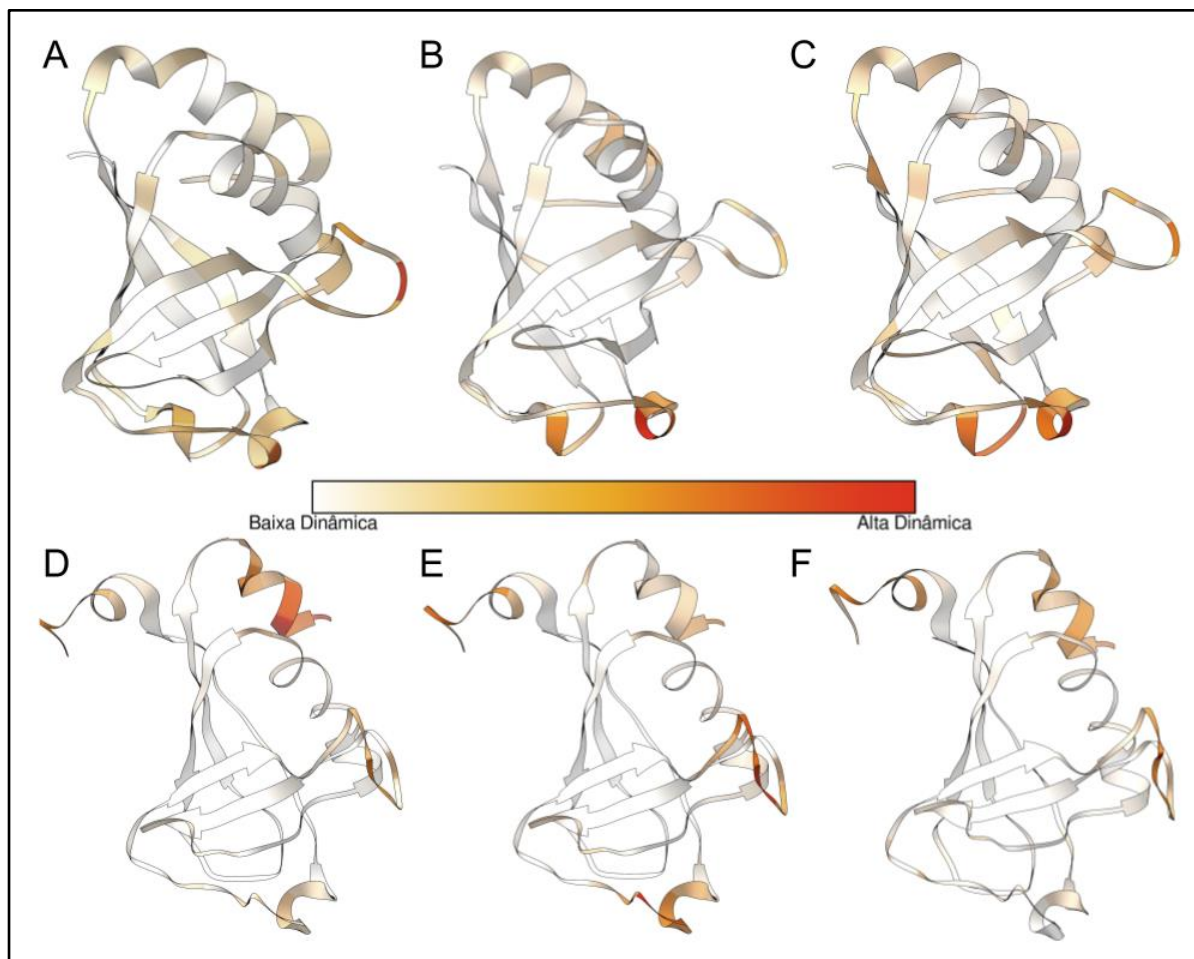
Tabela 3: RMSD médio e seu desvio-padrão das réplicas dos modelos hsHLTF_{APO} e scRad5.

RMSD (nm)						
Modelo	Réplica 1	Réplica 1*	Réplica 2	Réplica 2*	Réplica 3	Réplica 3*
hsHLTF _{APO}	0,33±0,03	0,27±0,02	0,28±0,03	0,23±0,03	0,27±0,04	0,22±0,02
scRad5	0,71±0,11	0,48±0,04	0,62±0,06	0,55±0,05	0,48±0,07	0,46±0,05

* Domínio HIRAN sem as regiões com valores de RMSF elevados. Para a hsHLTF_{APO}, resíduos 59-171. Para a scRad5, resíduos 171-294.

Os valores de RMSF convertidos em b-factor nos permite traçar uma correlação positiva de flutuação a partir destes valores: quanto maior o valor do resíduo no RMSF, maior será o seu *b-factor* (**Figura 16**). Como a ferramenta *Render by Attribute* utiliza os valores médios de todos os resíduos, foi retirado do cálculo aqueles resíduos que apresentaram um valor de RMSF elevado, restando os resíduos 59-171 do hsHLTF_{APO} e 171-294 do scRad5. Foi observado que todos os *loops* possuem uma maior movimentação espacial quando comparado com as estruturas secundárias, em ambos os modelos. No hsHLTF_{APO}, os resíduos que possuíram maior flutuação foram as tirosinas Y72 e Y93, que são responsáveis pela interação com a ssDNA (KILE, 2015). No modelo scRad5, a região da alfa-hélice $\alpha 2$ possui uma dinâmica de movimento superior às de regiões de *loops*, indicando uma desnaturação de sua estrutura secundária e os resíduos que apresentaram maior flutuação foram K197, Y198, G199 e Y217, todos em regiões de *loops*. Essas variações estruturais, tanto dos resíduos indicados, quanto dos *loops*, podem ser essenciais para o reconhecimento e estabilização da ssDNA.

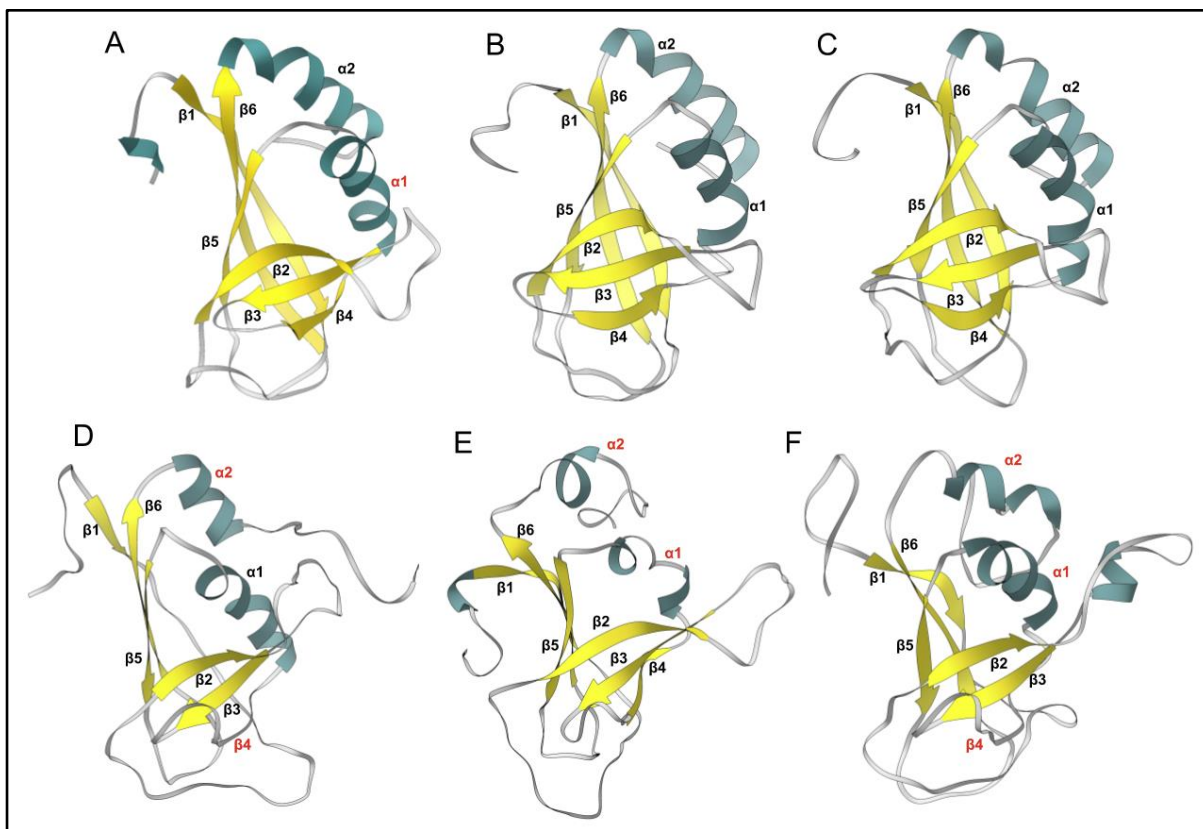
Figura 16: Flexibilidade residual a partir do RMSF dos modelos hsHLTF_{APO} e scRad5.



Os dados extraídos do RMSF convertidos em *b-factor* para análises de movimento dos resíduos. Os dados foram normalizados a partir da retirada das regiões com valores elevados de RMSF. De A a C, está representado o modelo hsHLTF_{APO} e de D a F o modelo scRad5, com a numeração das Réplicas alinhadas de forma crescente. A cor branca representa um valor de *b-factor* menor e vermelho de *b-factor* maior. A correlação entre o valor de *b-factor* com a dinâmica de movimento do resíduo é positiva e proporcional.

Ao atingir o equilíbrio energético na DM, espera-se que todas as estruturas das réplicas apresentem o mesmo valor na variação do RMSD. A diferença no RMSD final das réplicas do modelo hsHLTF_{APO} se deve ao comportamento dinâmico da $\alpha 1$ na Réplica 1 (**Figuras 15 e 16 A**). No modelo scRad5, a diferença foi devido a desestruturação de diversas estruturas secundárias, sendo a estrutura mais afetada a Réplica 2 (**Figura 17 E**) e as estruturas secundárias de folhas beta $\beta 4$ e as alfa-hélices $\alpha 1$ e $\alpha 2$ (**Figura 17 D, E, F**). Para o domínio HIRAN da HLTF, as análises das estruturas finais e das estruturas secundárias formadas ao longo da etapa de produção (**Figura 18**) também demonstraram que a $\alpha 0$ e a alfa-hélice formada entre as folhas beta $\beta 2$ e $\beta 3$ são de fato transientes, tal como especulado por Kile e colaboradores (2015) e Korzhnev e colaboradores (2016).

Figura 17: Desnaturações e diferenças estruturais nos modelos hsHLTF_{APO} e scRad5.



Representação da estrutura terciária da última coordenada de posição de cada átomo ao final de 500 ns de simulação. De A a C, está disposto as réplicas do modelo hsHLTF_{APO} e de D a F as réplicas do modelo scRad5, todas as Réplicas representadas de forma crescente: Réplica 1 (A e D), 2 (B e E) e 3 (C e F). Em vermelho está em destaque as estruturas secundárias mais dinâmicas ao longo da etapa de produção.

Após análises das estruturas finais dos modelos, foi utilizado o programa DSSP (KABSCH, 1983) para análises de formações e quantificação de estruturas secundárias ao longo de toda a trajetória da DM. O gráfico gerado pelo DSSP, com a numeração dos resíduos normalizada (o resíduo 1 corresponde ao primeiro resíduo do modelo), nos permite analisar quais resíduos estão se estruturando em *coil*, folha-beta, *beta-bridge*, *bend*, *turn*, alfa-hélice e em alfa-hélices parciais com 3 e 5 resíduos. No modelo hsHLTF_{APO} (**Figura 18**), foi possível perceber que as estruturas secundárias se mantiveram estáveis, exceto pela alfa-hélice $\alpha 2$ da réplica 2 (resíduos 105-125) durante o espaço de tempo de 25 ns a 200 ns, apresentando uma desestruturação parcial. Foi possível observar, também, que a alfa-hélice inicial (resíduos 0-10) e a localizada entre folhas beta $\beta 2$ e $\beta 3$ (resíduos 40-50) são transitórias, estruturando-se esporadicamente ao longo da etapa de produção, confirmando as afirmações de outros grupos de pesquisa (KILE, 2015; KORZHNEV, 2016). O modelo scRad5 (**Figura 19**) apresentou diversas desnaturações de estruturas secundárias: na Réplica 1, a alfa-hélice $\alpha 2$ (resíduos 115-125), e as folhas beta $\beta 1$ e $\beta 2$ (resíduos 5-15 e 35-45, respectivamente); na Réplica 2, as duas

alfa-hélices $\alpha 1$ e $\alpha 2$ (resíduos 75-85 e 115-125, respectivamente); na Réplica 3, a alfa-hélice $\alpha 2$ (resíduos 115-125) as folhas betas $\beta 1$, $\beta 2$ e $\beta 4$ (resíduos 5-15, 35-45 e 65-75, respectivamente), além de uma estruturação em alfa-hélice entre as folhas beta $\beta 2$ e $\beta 3$ (resíduos 50-60). Estas análises permitem inferir que o modelo hsHLTF_{APO} possui uma estabilidade estrutural maior que o modelo scRad5.

A estabilidade do domínio HIRAN da scRad5 pode estar dependente da estruturação de domínios adjacentes, uma vez que análises da dinâmica de uma modelagem molecular, da proteína Rad5 completa, auxiliada por SAXS (espalhamento de Raio-X de baixo ângulo) (GILDENBERG, 2019), não apresentou tal desestruturação. Gildenberg e colaboradores (2019) utilizam o PDB ID 4XZG para a modelagem do domínio HIRAN, o 4R8P para o domínio RING e os domínios da helicase os modelos 1Z3I, 6GEJ, 3MWY, 5O9G, 6FML e 6G7E. A técnica de DM utilizada foi a *coarse-grained* num período de tempo de 5 μ s e relataram que o HIRAN da proteína Rad5 possui uma interação com os domínios helicases e que quando próximos se mostraram mais favoráveis energeticamente. A conclusão dos autores é que estes domínios possuem uma interação e é preciso aguardar a resolução experimental da estrutura da proteína Rad5 para melhor compreendê-la. Os dados observados por este grupo de pesquisa estão de acordo com os dados apresentados neste trabalho e com dados do **Anexo I**, sugerindo que o domínio HIRAN da proteína Rad5 necessite dessas interações com outros domínios para manter sua estrutura terciária na forma livre.

Figura 18: Estruturas secundárias formadas ao longo da etapa de produção do modelo hsHLTF_{APO}.

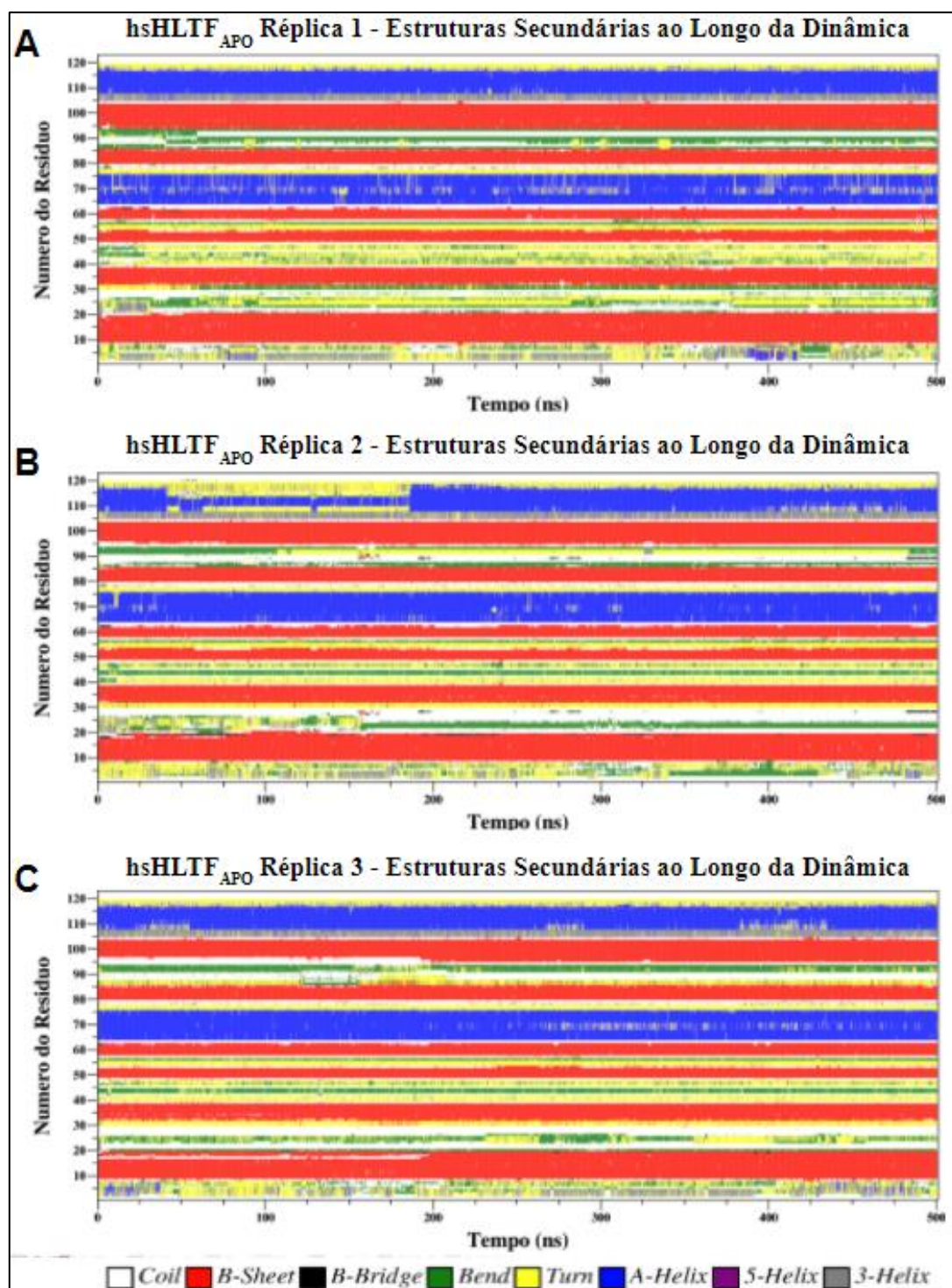


Gráfico de análise de estruturas secundárias formadas, gerado pelo programa DSSP, das réplicas do modelo hsHLTF_{APO} ao longo da trajetória. Em (A) réplica 1, em (B) réplica 2 e em (C) réplica 3. Em branco, estruturas do tipo *Coil*, em vermelho do tipo Folhas Beta, em preto do tipo *Beta-Bridge*, em verde do tipo *Bend*, em amarelo do tipo *Turn*, em azul do tipo Alfa-Hélice, em roxo e em cinza do tipo Alfa-Hélice de 5 e 3 resíduos, respectivamente. As numerações dos resíduos estão normalizadas, contando como resíduo 1 o primeiro resíduo da estrutura.

Figura 19: Estruturas secundárias formadas ao longo da etapa de produção do modelo scRad5.

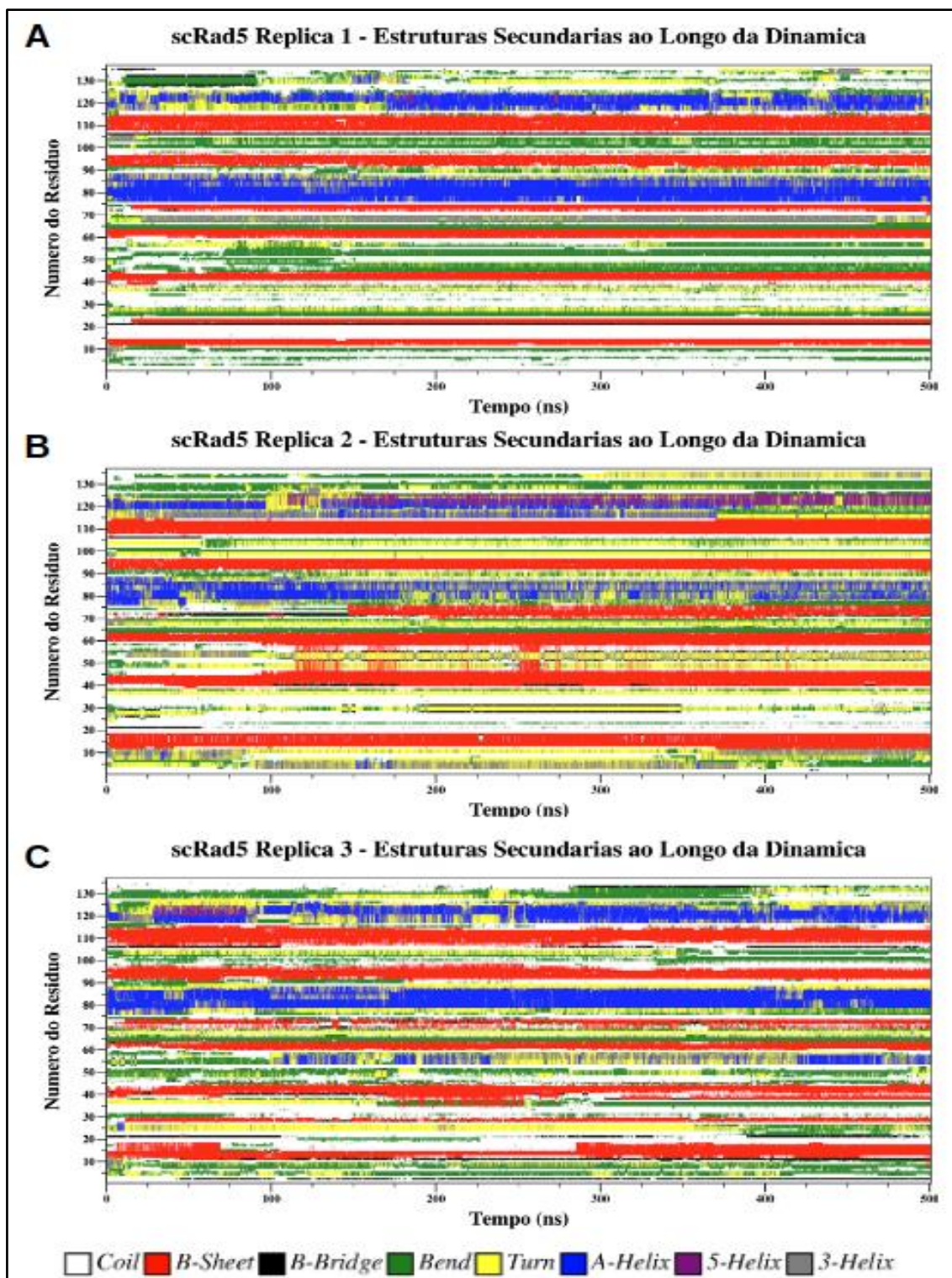


Gráfico de análise de estruturas secundárias formadas, gerado pelo programa DSSP, das réplicas do modelo scRad5 ao longo da trajetória. Em (A) réplica 1, em (B) réplica 2 e em (C) réplica 3. Em branco, estruturas do tipo *Coil*, em vermelho do tipo Folhas Beta, em preto do tipo *Beta-Bridge*, em verde do tipo *Bend*, em amarelo do tipo *Turn*, em azul do tipo Alfa-Hélice, em roxo e em cinza do tipo Alfa-Hélice de 5 e 3 resíduos, respectivamente. As numerações dos resíduos estão normalizadas, contando como resíduo 1 o primeiro resíduo da estrutura.

O programa DSSP quantifica a somatória dos resíduos que estão adquirindo estruturas secundárias (alfa-hélice e folha-beta), em forma de um gráfico, e foi usado para melhor elucidar o panorama estrutural dos modelos (**Figura 20**). Para a elaboração da **Tabela 3**, foi calculado a média e o desvio padrão do eixo y.

Figura 20: Número de resíduos com estruturas secundárias ao longo da etapa de produção dos modelos hsHLTF_{APO} e scRad5.

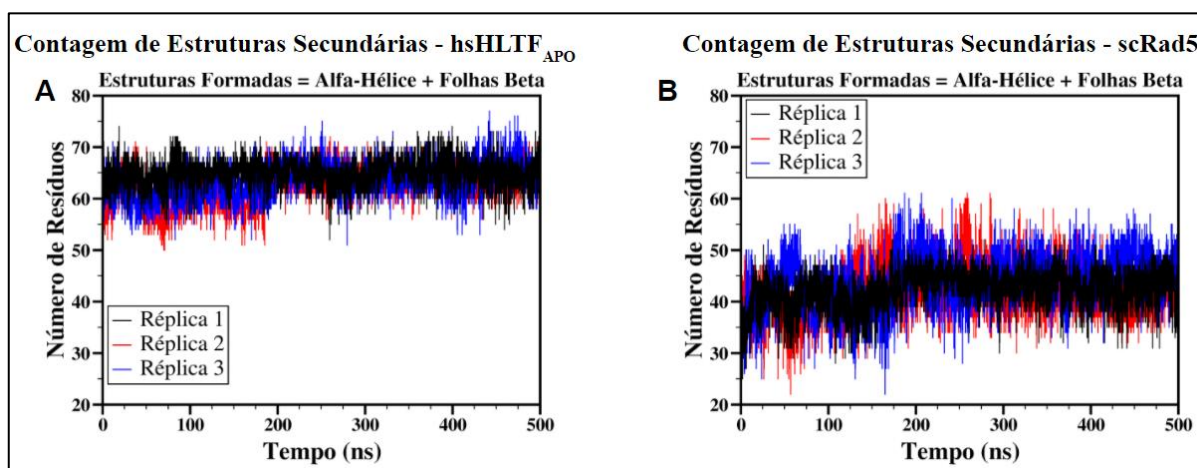


Gráfico de contagem dos resíduos que se estruturam em estruturas secundárias de alfa-hélice, folha beta e realizada pelo programa DSSP. Em A, contagem de estruturas secundárias do modelo scRad5 e em B do modelo hsHLTF_{APO}. Em ambas, as cores pretas representam as Réplicas 1, as vermelhas as Réplicas 2 e em azuis as Réplicas 3.

Tabela 4: Número de resíduos médio com estruturas secundárias e seus desvios-padrões dos modelos hsHLTF_{APO} e scRad5.

Número de Resíduos com Estruturas Secundárias*			
Modelo	Réplica 1	Réplica 2	Réplica 3
hsHLTF _{APO}	64,87 ± 2,70	62,38 ± 3,41	63,15 ± 3,23
scRad5	42,30 ± 3,89	42,84 ± 5,21	44,06 ± 5,35

*Estruturas secundárias: somatório das alfa-hélices e folhas betas.

As análises da **Figura 20** e da **Tabela 3** mostram que o modelo hsHLTF_{APO} apresentou um número de resíduos com estrutura secundárias mais homogêneo e maior que o modelo scRad5, estando de acordo com os dados das **Figuras 15** e **17**. A Réplica 2 do scRad5 apresentou a maior variação no número de resíduos quando comparado com as outras réplicas, sendo a Réplica 1 a que menor contém estruturas secundárias. Este número de resíduos pode ser correlacionado à estabilidade do modelo, uma vez que a heterogeneidade da variação deste

número indica que a estrutura está sofrendo mudanças conformacionais mais bruscas, alterando os seus padrões de interações de hidrogênio e inter-residuais.

Todas as réplicas das simulações dos modelos foram feitas a fim de encontrar conformações ideais do sítio de ligação a ssDNA e estudar a dinâmica dos aminoácidos envolvidos na interação e estabilização da fita. Estas análises irão permitir inferir a afinidade, os mecanismos de interações e, se possível, traçar um aspecto evolutivo do domínio HIRAN em humanos comparado com o mesmo em levedura.

O uso do programa GROMACS, combinado com o campo de força AMBER, o termostato *V-rescale*, o barostato de Parrinello-Rahman e o integrador *leapfrog*, se mostraram eficientes para conduzir a etapa de produção da dinâmica molecular. O GROMACS (BERENDSEN, 1994) foi utilizado em sistemas de DNA fita simples, provendo informações sobre os efeitos da ureia na sua estrutura tridimensional (OPRZESKA-ZINGREBE, 2018), na elucidação do papel das interações hidrofóbicas na inibição da atividade da proteína MDM2 por polifenóis (VERMA, 2016) e no sistema do complexo Rad4-DNA de Pitta e colaboradores (2018), que também utilizou-se do integrador *leapfrog* e do *V-rescale*. O barostato Parrinello-Rahman (PARRINELLO, 1981) é encontrado em diversos trabalhos envolvendo sistemas biológicos (ANDREWS, 2017; OPRZESKA-ZINGREBE, 2018).

Os campos de força AMBER (CORNELL, 1995) são comumente empregados em dinâmicas envolvendo proteínas em geral, como no estudo de enovelamento de peptídeos (SHAO, 2018) e na simulação de cadeias laterais de aminoácidos (PETROVIC, 2018). Liu e colaboradores (2019) demonstraram como a mutação pontual R249S na proteína p53 interfere negativamente na interação com o sulco maior do DNA, Zhang e colaboradores (2019) conseguiram inferir a participação de íons de Potássio na interação da proteína EndoMS/NucS com a fita dupla de DNA e Lakhani e colaboradores (2017) identificaram possíveis aminoácidos responsáveis pela interação da MutS em DNA fita dupla com erros de pareamento de bases. Diversos outros trabalhos são encontrados na literatura e auxiliam no entendimento de proteínas e moléculas que interagem com fita dupla de DNA (PRADHAN, 2018; HAMED, 2017; KAMARAJ, 2015; MARY, 2017) e fita simples de DNA (OPRZESKA-ZINGREBE, 2018).

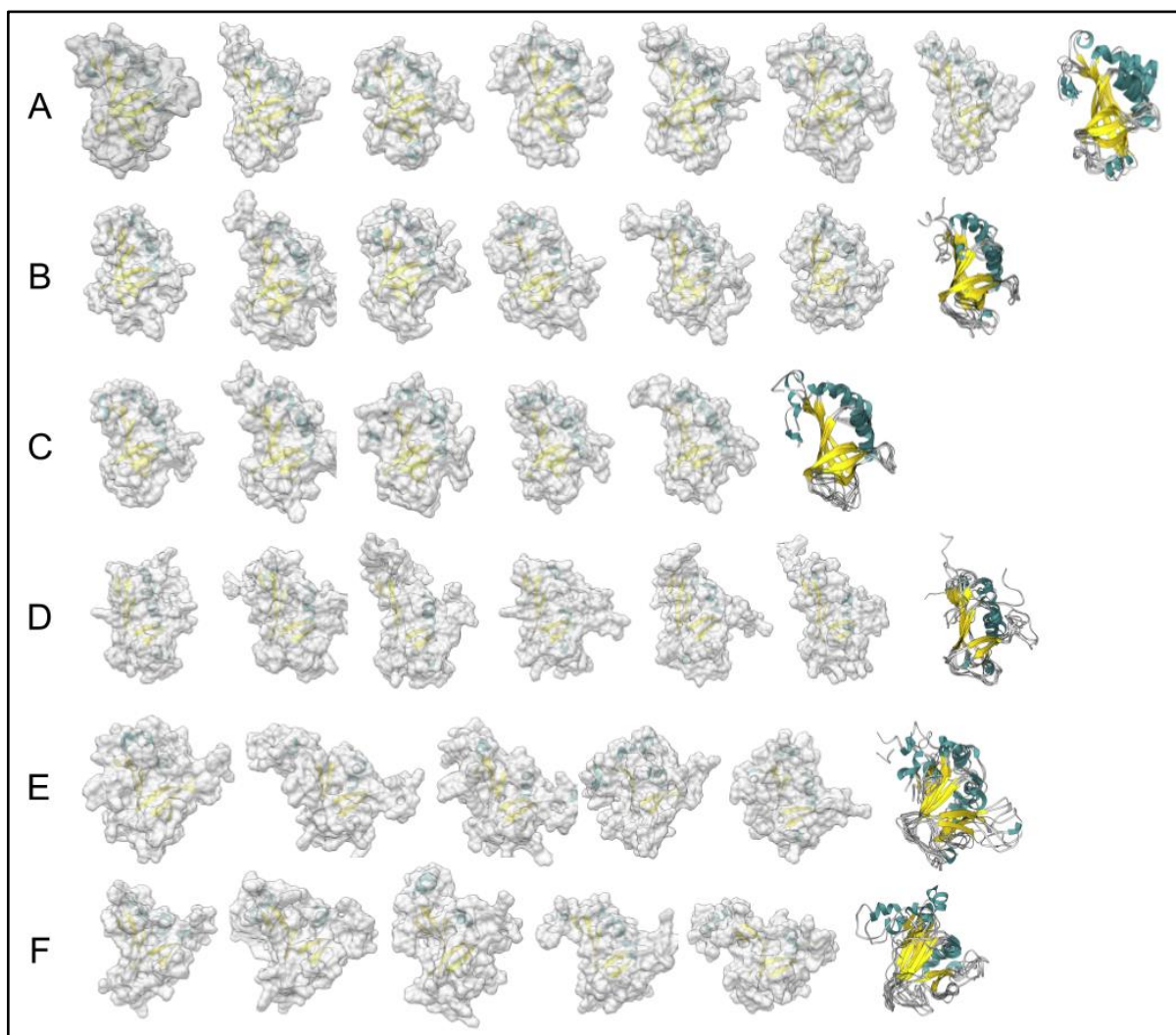
5.3) O SÍTIO DE INTERAÇÃO PODE APRESENTAR DUAS CONFORMAÇÕES

O estudo envolvendo a cristalização do domínio HIRAN da HLTF com uma ssDNA aponta para a importância dos aminoácidos Y72, Y93, N91, H110, R71 e F142 na interação e estabilização da fita dentro da cavidade formada (KILE, 2015). De forma majoritária, é possível observar que a fita de DNA é estabilizada por interações de empilhamento de elétrons π , da

cadeia lateral das tirosinas com as bases nitrogenadas, e forças eletrostáticas de momentos positivo com a cadeia fosfato, que possui momento negativo.

O algoritmo GROMOS de agrupamento, que utiliza a variação de RMSD de todos os átomos, foi utilizado para encontrar estruturas representativas da trajetória da dinâmica dos modelos hsHLTF_{APO} e scRad5 (**Figura 21**). Para cada modelo foi utilizado um valor de corte específico a fim de encontrar um número mínimo de estruturas que representam as conformações estruturais adotadas ao longo de toda a etapa de produção. Estas estruturas, juntamente com a análises completa de todos os quadros da etapa de produção, ajudam a entender o comportamento do resíduo de aminoácidos presentes na cavidade de interação com o ssDNA.

Figura 21: Estruturas representativas das etapas de produção dos modelos hsHLTF_{APO} e scRad5.



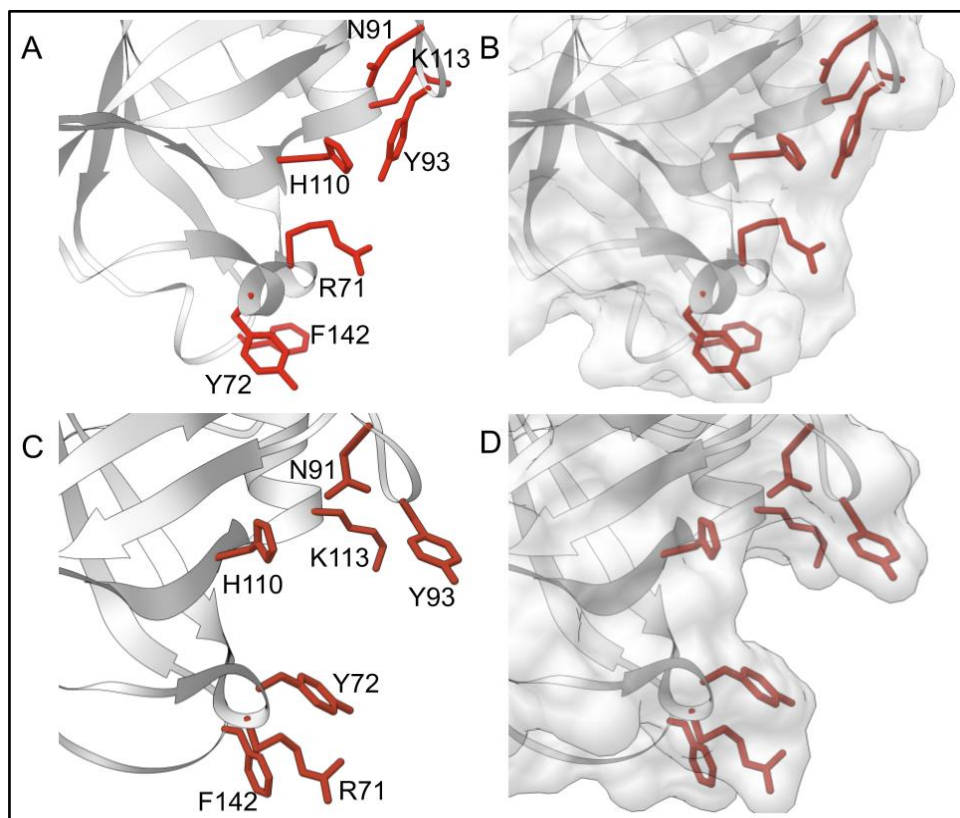
Estruturas representativas extraídas através de um algoritmo de agrupamento e valores de corte baseado no RMSD global dos modelos hsHLTF_{APO} e scRad5. De A a C estão representadas as estruturas do hsHLTF_{APO} e de D a F do

scRad5, seguindo a ordem crescente das numerações das réplicas. Em amarelo, estão representadas as folhas betas, em ciano as alfa-hélices e em cinza claro a superfícies molecular das estruturas. A última estrutura de cada letra é a sobreposição de todas as estruturas obtidas sem a superfície molecular.

As análises das estruturas representativas permitiram observar que o domínio HIRAN da HLTF se comporta de maneira mais uniforme enquanto a de levedura de forma mais heterogênea, estando de acordo com os resultados anteriores. A cavidade de interação da ssDNA não se forma de maneira completa em nenhum momento da simulação, em ambos os modelos. É possível inferir que este bolsão não se forma de maneira espontânea em solução aquosa e levanta a hipótese de que este bolsão se estrutura apenas na presença da ssDNA, uma vez que as regiões de *loops* e os aminoácidos responsáveis pela estabilização possuem um valor de RMSF mais elevado e portanto possuem uma elevada dinâmica espacial. De acordo com a revisão de Du e colaboradores (2016), para que ocorra a indução de conformação, a região necessita de uma região complementar entre os pares e diferentes tentativas de colisões e interações entre a proteína e o ligante, justificando a necessidade do sítio de interação conter uma alta motilidade na sua forma livre. Nesta revisão os pesquisadores afirmam que a indução de conformação, por parte do ligante, permite a proteína interagir com diferentes conformações de substratos, permitindo mutações e variações no ligante, e ser uma vantagem evolutiva.

O modelo de PDB ID 4S0N (hsHLTF_{HOLo}) é a estrutura do domínio HIRAN de HLTF no momento de interação com a ssDNA resolvida a partir da cristalografia e difração de raios-x. Este foi usado como comparativo para analisar as posições espaciais dos aminoácidos situados na cavidade formada (**Figura 22**).

Figura 22: Importância dos resíduos R71 e Y72 para a formação do sítio de interação a ssDNA.



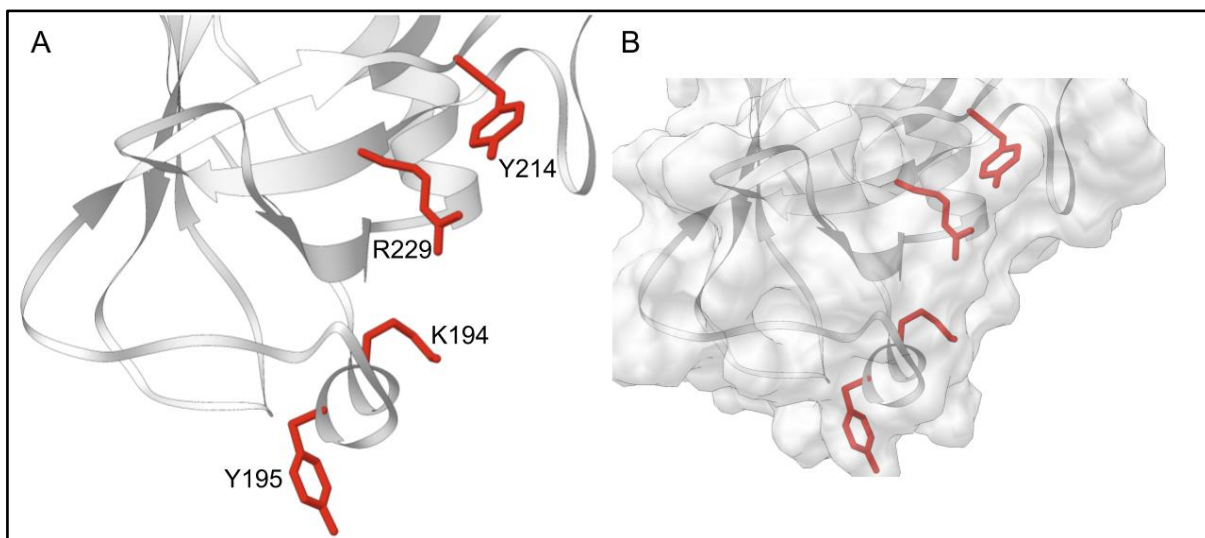
A formação da cavidade de interação está diretamente ligada às posições dos resíduos R71 e Y72 no *loop* inferior do sítio de interação a ssDNA. Em A e C, destaque dos aminoácidos responsáveis pela estabilização da ssDNA dos modelos hsHLTF_{APO} e hsHLTF_{HOLO}, respectivamente. Em B e D as superfícies moleculares dos modelos hsHLTF_{APO} e hsHLTF_{HOLO}, respectivamente. Em cinza, a estrutura terciária do domínio HIRAN e em vermelho as cadeias laterais dos aminoácidos destacados.

A comparação das posições espaciais dos aminoácidos dos modelos hsHLTF_{APO} e hsHLTF_{HOLO} na região do bolsão de interação a ssDNA revelou a importância dos resíduos R71 e Y72 para a formação do mesmo. O hsHLTF_{APO} (**Figura 22 A**) apresenta a R71 internalizada na cavidade de interação, de forma que sua cadeia lateral restringe o espaço para a entrada da ssDNA, enquanto a R71 do hsHLTF_{HOLO} (**Figura 22 C**) se encontra na parte mais externa, liberando espaço para a movimentação da Y72 e se projetar para parear com os nucleotídeos juntamente com a Y93. Importante ressaltar que essa transição de conformação espacial, do modelo hsHLTF_{APO} para o hsHLTF_{HOLO}, dos resíduos R71 e Y72, não acontece durante a etapa de produção.

O provável sítio de interação do domínio HIRAN da Rad5 se comporta de forma similar ao do modelo hsHLTF_{APO}. De fato, os aminoácidos da região de interação com ssDNA são diferentes, porém resíduos com cadeias laterais similares são encontrados em posições semelhantes (**Figura 23**). As tirosinas Y195 e Y214 são encontradas nos *loops* inferior e

superior, e os resíduos de arginina, R71, e histidina, H110, são substituídos por resíduos de lisina, K194, e arginina, R229, respectivamente. Este sítio também apresenta uma flexibilidade alta e possivelmente se comporta de maneira similar à estrutura do domínio HIRAN da HLTF. A flexibilidade das cadeias laterais da K194 e Y195 se assemelha a R71 e Y72, de modo que a K194, quando internalizada na cavidade de interação, restringe estericamente a entrada da ssDNA (**Figura 23 B**).

Figura 23: Possível sítio de interação com a ssDNA do domínio HIRAN da Rad5.

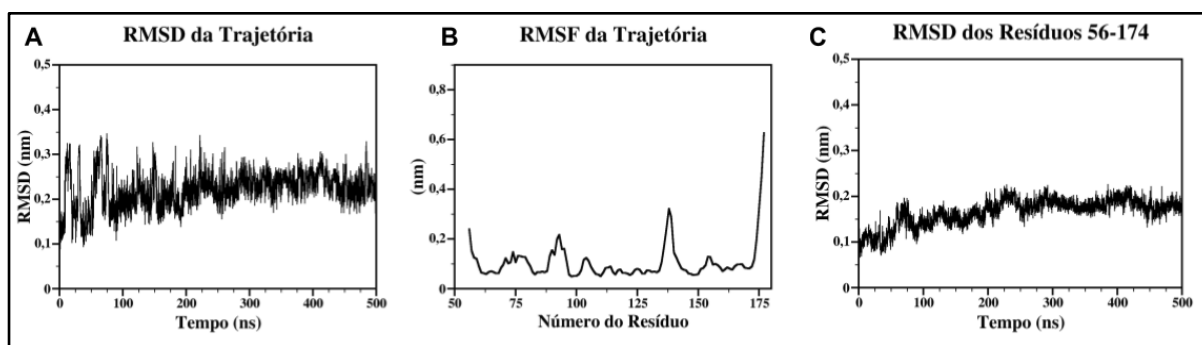


Os aminoácidos encontrados na região do provável sítio de interação com a ssDNA do domínio HIRAN da Rad5. Os aminoácidos representados em vermelho se assemelham em características de cadeia lateral dos resíduos do modelo hsHLTF_{APO}. Em A, estrutura terciária do modelo scRad5, na cor cinza, e em B com sua superfície molecular aparente.

A cadeia A do modelo 4S0N (hsHLTF_{HOLO}), sem o ligante, foi utilizada na simulação de dinâmica molecular para analisar se a cavidade de interação, após formada e estabilizada com a ssDNA, adquire a conformação encontrada em sua forma livre. O modelo foi submetido às mesmas condições e tempo de simulação que o modelo hsHLTF_{APO}.

O gráfico de RMSD de todos os resíduos do modelo hsHLTF_{HOLO} (**Figura 24 A**) mostrou uma estrutura com considerável dinamismo estrutural e ainda instável dado o tempo de simulação. O gráfico de RMSF (**Figura 24 B**) revelou que a região C-terminal do modelo não possui estruturação e uma alta mobilidade, portanto, o gráfico de RMSD dos resíduos 56-174 (**Figura 24 C**), com exclusão desta região, foi traçado.

Figura 24: Dinâmica estrutural da cadeia A do modelo hsHLTF_{HOLO}.

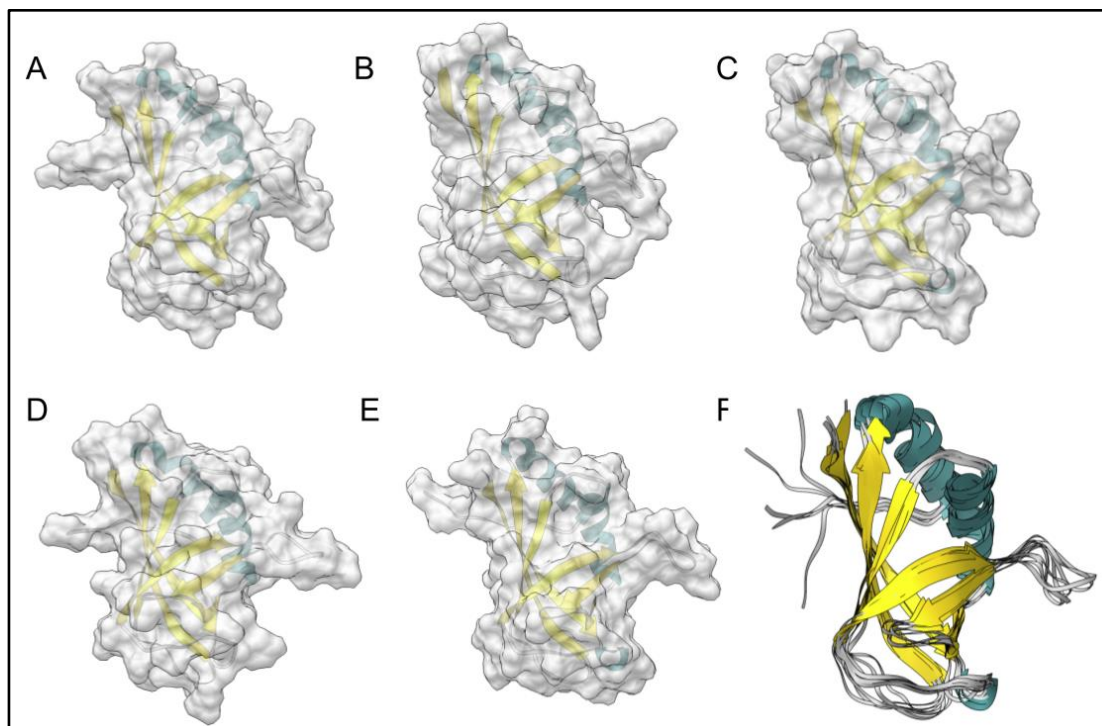


Gráficos de RMSD e RMSF do modelo hsHLTF_{HOLO} em solução aquosa ao longo de 500 ns de simulação. Em A, RMSD global, com todos os resíduos envolvidos, em B, RMSF de todos os resíduos e em C, o RMSD parcial, contando somente dos resíduos 56 a 174.

O RMSD médio de todos os resíduos do modelo hsHLTF_{HOLO} foi de 0,22 nm com um desvio-padrão de 0,40 nm, enquanto o RMSD médio dos resíduos 56 a 174 foi de 0,16 nm \pm 0,03 nm. Estes valores, juntamente com os gráficos da **Figura 20** e da **Tabela 3**, mostram que esta estrutura possui uma menor mobilidade, quando comparada com o modelo hsHLTF_{APO}. Este fato pode se dar por ser uma estrutura originada de uma cristalização, portanto sendo uma conformação situacional, mais rígida, ou devido a maior estabilidade do HIRAN em sua conformação de interação com a ssDNA, após a estabilização e interação com a mesma. A partir do gráfico de RMSD é possível concluir que o tempo de simulação adotado foi suficiente para explorar as diferentes conformações *in silico* do modelo, dado a pequena variação do mesmo.

Utilizando o algoritmo de agrupamento, foi possível adquirir as estruturas representativas da etapa de produção do modelo hsHLTF_{HOLO} (**Figura 25**). Os critérios de seleção foram os mesmos para os modelos hsHLTF_{APO} e scRad5, mudando somente o valor de corte.

Figura 25: Estruturas representativas da etapa de produção do modelo hsHLTF_{HOLO}.



Estruturas representativas extraídas através de um algoritmo de agrupamento e valores de corte baseado no RMSD global do modelo hsHLTF_{HOLO}. De A a E estão representadas as estruturas obtidas com sua superfície molecular a mostra. Em F, a sobreposição de todos os modelos sem a superfícies molecular. Em amarelo, estão representadas as folhas betas, em ciano as alfa-hélices e em cinza claro a superfícies molecular das estruturas.

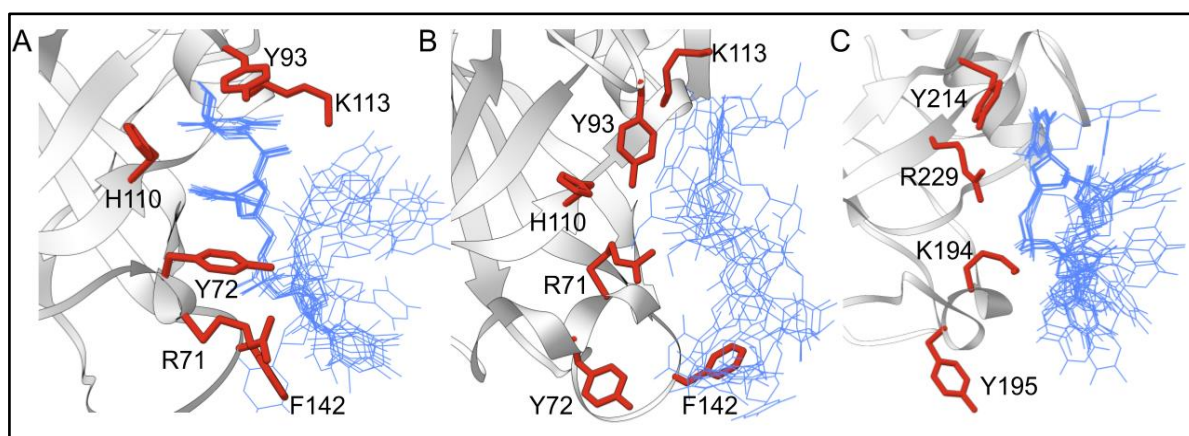
As estruturas representativas do modelo hsHLTF_{HOLO} apresenta um perfil homogêneo, a partir de uma inspeção visual, sem grandes variações estruturais e de acordo com os dados da **Figura 24**. Juntamente com a análises dos quadros da etapa de produção, é possível perceber que o sítio de interação a ssDNA não se desfaz e a conformação espacial das cadeias laterais dos resíduos R71 e Y72 não adquire a conformação encontrada no modelo hsHLTF_{APO}. Esta análise nos permite inferir que o domínio HIRAN da HLTf, ao adquirir uma conformação de interação com a fita de DNA, se mantém estável e, possivelmente, seria necessário um tempo de simulação maior, ou outra abordagem de DM, para poder adquirir e observar o estado de transição entre estas estruturas.

5.4) PREPARAÇÃO DAS ESTRUTURAS PARA SIMULAÇÃO COM A ssDNA

O programa UCSF DOCK6 foi utilizado para gerar poses estruturais a partir do *docking* molecular da ssDNA dentro do sítio de interação nos modelos hsHLTF_{HOLO} (hsHLTF_{HOLO/ssDNA}), hsHLTF_{APO} (hsHLTF_{APO/ssDNA}) e scRad5 (scRad5_{ssDNA}). As estruturas utilizadas foram as mesmas estruturas iniciais dos tópicos 5.2 e 5.3: estruturas retiradas do bando de dados e calculada a partir da modelagem comparativa. O *re-docking* do modelo

hsHLTF_{HOLo/ssDNA} se mostrou eficiente e em concordância com os dados apresentados por Kile e colaboradores (2015) após a cristalização do domínio HIRAN da HLTF com uma ssDNA poli-T. As mesmas interações e conformações descritas pelo grupo foram observadas e as 10 poses de menor energia possuíam um RMSD de 0,51 Å (pose de menor energia como referência), mostrando que o programa UCSF DOCK6 foi eficiente na condução da etapa de busca por estruturas (**Figura 26 A**). A cadeia G do PDB ID 4S0N foi utilizada como o molde do ligante para todos os modelos e posicionada ao próximo ao sítio, evitando sobreposições de superfícies moleculares.

Figura 26: *Docking* molecular dos modelos hsHLTF_{HOLo/ssDNA}, hsHLTF_{APo/ssDNA} e scRad5_{ssDNA} com poli-T.



As 10 poses de menor energia após o *docking* molecular dos modelos hsHLTF_{HOLo}, hsHLTF_{APo} e scRad5 com ssDNA poli-T. Em A, o modelo hsHLTF_{HOLo/ssDNA}, em B, o modelo hsHLTF_{APo/ssDNA} e em C, o modelo scRad5_{ssDNA}. Em azul estão representadas as estruturas adquiridas pela ssDNA e em vermelho as cadeias laterais dos resíduos indicados na estabilização da fita.

O modelo hsHLTF_{APo/ssDNA} apresentou poses estruturais heterogêneas e obteve um RMSD entre as estruturas de 1,4 Å (**Figura 26 B**). É possível observar que, como o sítio de interação não está formado, as estruturas do ligante oscilam e não se observa um mínimo comum. Este fato pode se dar pela internalização da tirosina Y93 e a externalização da Y72 na região do bolsão, o que impede a fita de parear e realizar o empilhamento de elétrons π . Como esta região apresenta um dinamismo espacial muito alto, a estrutura do HIRAN da HLTF pode apresentar outras conformações para que haja a interação com a ssDNA.

O *docking* envolvendo o modelo scRad5 apresentou estruturas mais homogêneas, quando comparadas com as do modelo hsHLTF_{APo/ssDNA}, com um RMSD entre as poses de 1,362 Å (**Figura 26 C**). Apesar do valor de RMSD próximo ao do domínio HIRAN da HLTF na sua conformação livre, a superfície molecular do resíduo Y214 estava sem impedimento

estérico e a ssDNA conseguiu parrear com a mesma, apresentando conformações similares e mais agrupadas nesta região.

A simulação de dinâmica molecular da pose de menor energia de cada modelo foi realizada para analisar as interações e a estabilidade da ssDNA dentro do sítio de interação (**Figura 27**).

Figura 27: Sistemas utilizados para a dinâmica molecular com ssDNA dos modelos $hsHLTF_{HOLO/ssDNA}$, $hsHLTF_{APO/ssDNA}$ e $scRad5_{ssDNA}$.

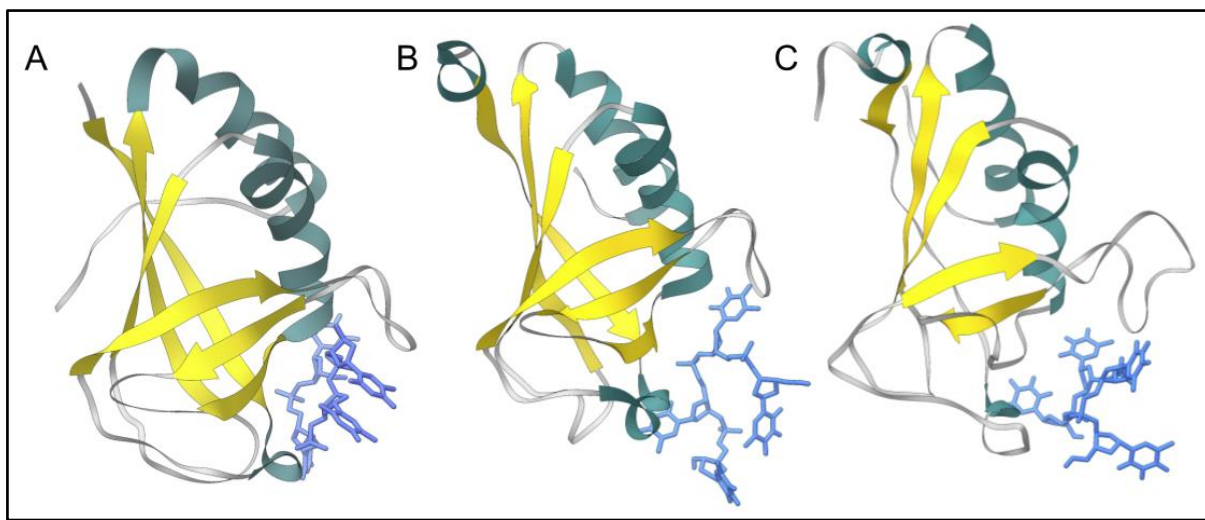


Figura com as estruturas de menor energia que foram utilizadas como sistemas inicial nas dinâmicas dos domínios HIRAN com ssDNA. Em A, o modelo $hsHLTF_{HOLO/ssDNA}$, em B, o modelo $hsHLTF_{APO/ssDNA}$ e em C, o modelo $scRad5_{ssDNA}$. Em azul está representado a pose de ssDNA poli-T.

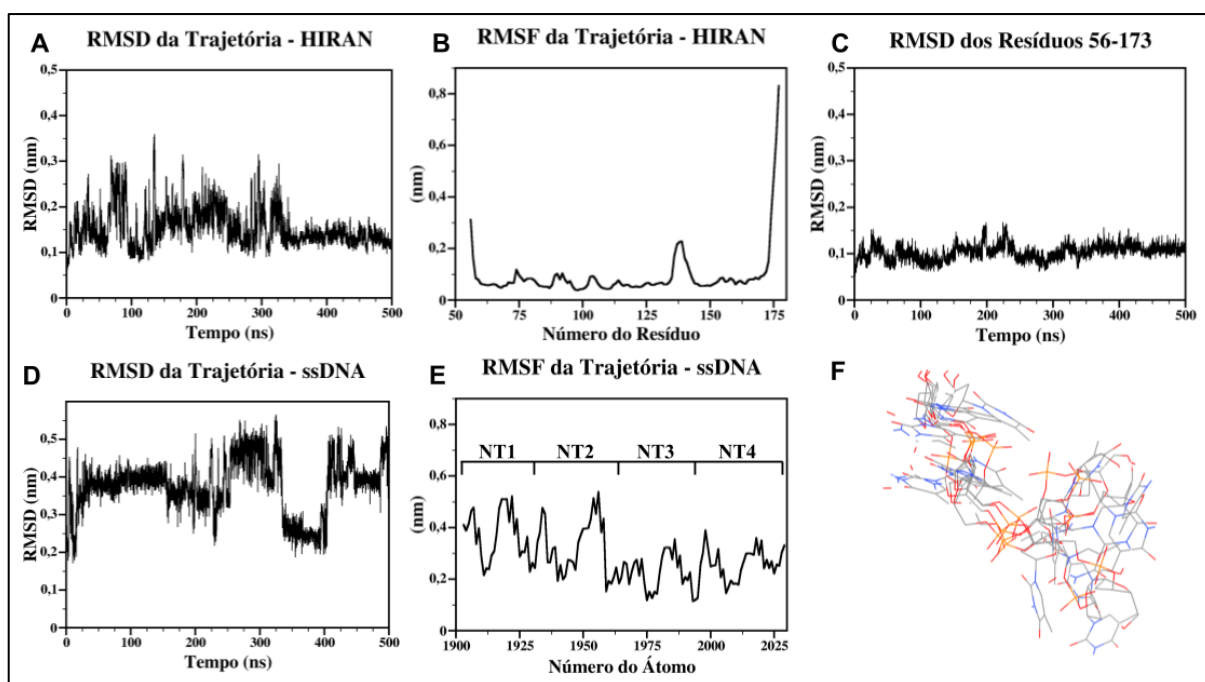
O programa UCSF DOCK6 se mostrou eficiente em trabalhos envolvendo a proteína de reparo de DNA Rad51, do qual Fornander e colaboradores (2012) demonstraram a uma melhor interação com o ATP na presença de íons cálcio, comparado com íons magnésio, e a proteína Rad52, do qual Li e colaboradores (2018) identificaram compostos inibitórios a partir de um *screening* virtual. Nestes trabalhos, foram utilizadas pequenas moléculas, com um menor número de rotâmeros que de uma fita de DNA, porém, o propósito do *docking* nesta etapa foi a seleção de estruturas com menor energia, evitando átomos em posições irregulares e garantir um início em comum para todos os modelos.

5.5) A ESTRUTURA TERCIÁRIA DO DOMÍNIO HIRAN É MAIS ESTÁVEL COM A ssDNA NO SÍTIO DE INTERAÇÃO

5.5.1) DOMÍNIO HIRAN DA PROTÉINA HLTF EM SUA CONFORMAÇÃO DE INTERAÇÃO COM A ssDNA

A simulação de dinâmica molecular do modelo hsHLTF_{HOLO/ssDNA} foi utilizada como referência para as simulações de dinâmicas envolvendo o domínio HIRAN com a ssDNA. Todos os modelos utilizaram as mesmas condições e parâmetros para ser possível estudar, de forma comparativa, as variações moleculares envolvidas no processo de estabilização e interação com a fita. As análises das dinâmicas espaciais foram feitas a partir de gráficos de RMSD, RMSF e agrupamento de estruturas representativas (**Figura 28**).

Figura 28: Dinâmicas estruturais da cadeia A modelo hsHLTF_{HOLO/ssDNA} e da ssDNA.



Análises das dinâmicas estruturais do domínio HIRAN da HLTF interagindo com a ssDNA, no modelo hsHLTF_{HOLO/ssDNA}. Em A, O gráfico RMSD e, em B, o de RMSF de todos os resíduos do modelo. Em C, gráfico de RMSD parcial, envolvendo os resíduos 56-173. Em D e E, os gráficos de RMSD e RMSF da ssDNA, respectivamente. Em F, as 6 estruturas representativas da etapa de produção da ssDNA extraída a partir do algoritmo de agrupamento.

O gráfico de RMSD do modelo hsHLTF_{HOLO/ssDNA} (**Figura 28 A**) demonstrou que o tempo de dinâmica foi suficiente para o sistema entrar em equilíbrio e explorar as diferentes conformações estruturais do domínio, porém, apresentou diversas instabilidades. O gráfico de RMSF (**Figura 28 B**) apresentou a mesma região com desestruturação e alta mobilidade do

modelo hsHLTF_{HOLO} em solução e, conseqüentemente, foi traçado o gráfico de RMSD dos resíduos 56 ao 173 (Região do N-terminal; **Figura 28 C**) e mostrou um sistema em equilíbrio. O RMSD médio global apresentado foi de $0,15 \pm 0,04$ nm e parcial de $0,10 \pm 0,02$ nm, se mostrando uma estrutura bem mais estável e com menos mobilidade quando comparada com os modelos sem a ssDNA: (a) hsHLTF_{HOLO}: médio total de $0,22 \pm 0,40$ nm, médio parcial de $0,16 \pm 0,03$ nm; (b) hsHLTF_{APO} Réplica 1: médio total de $0,33 \pm 0,03$ e médio parcial de $0,27 \pm 0,02$ nm; (c) hsHLTF_{APO} Réplica 2: médio total de $0,28 \pm 0,03$ e médio parcial de $0,23 \pm 0,03$ nm; (d) hsHLTF_{APO} Réplica 3: médio total de $0,27 \pm 0,04$ e médio parcial de $0,22 \pm 0,02$ nm.

O gráfico de RMSF (**Figura 28 B**) revelou que as regiões de *loops* envolvidas na interação com a ssDNA possuíram uma menor mobilidade espacial. Juntamente com os dados de RMSD, podemos sugerir que a presença da fita interfere na estabilidade e dinâmica estrutural do domínio HIRAN da HLTF.

Os gráficos de RMSD e RMSF da ssDNA (**Figura 28 D e E**) demonstraram uma estrutura com alta dinâmica conformacional. Este fato se deve pela exposição de 2 nucleotídeos com o solvente e que não estão internalizados na cavidade de interação do domínio. É possível confirmar esta ocorrência através das análises do agrupamento das estruturas representativas (**Figura 28 F**), nas quais os nucleotídeos internos apresentaram uma maior homogeneidade espacial quando comparado com os externalizados.

As análises dos contatos nativos e não-nativos realizados pelo domínio HIRAN com a ssDNA (**Figura 29**) estão de acordo com os resultados obtidos por Kile e colaboradores (2015), exceto pela F142 que apresentou interação com o NT1 por 25% do tempo de simulação. Esta análise quantifica o número de quadros que a distância entre os átomos esteve abaixo do valor de corte em relação à quantidade de quadros totais. Portanto, esse valor indica que os quadros que contêm esta interação equivale a 25% do tempo de simulação total. Este fato se deve pela exposição dos nucleotídeos iniciais do modelo com solvente e pode não ocorrer biologicamente, uma vez que, no momento da parada da forquilha de replicação, a fita de DNA não se apresenta com 4 nucleotídeos e sim de forma numerosa e estabilizada por regiões de fita dupla. De fato, tal como demonstrado por Hishiki e colaboradores (2020), o resíduo F142 é responsável pela separação da fita filha da fita molde no evento de transcrição do DNA, se ligando ao sulco menor da estrutura tridimensional do material genético. Este resíduo também está presente na sequência e estrutura terciária do domínio HIRAN de leveduras (**Figura 10**), de modo que a F267 possa ter função similar quando comparada com a F142 no domínio de humanos.

Figura 29: Resíduos do modelo hsHLTF_{HOLo/ssDNA} que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.

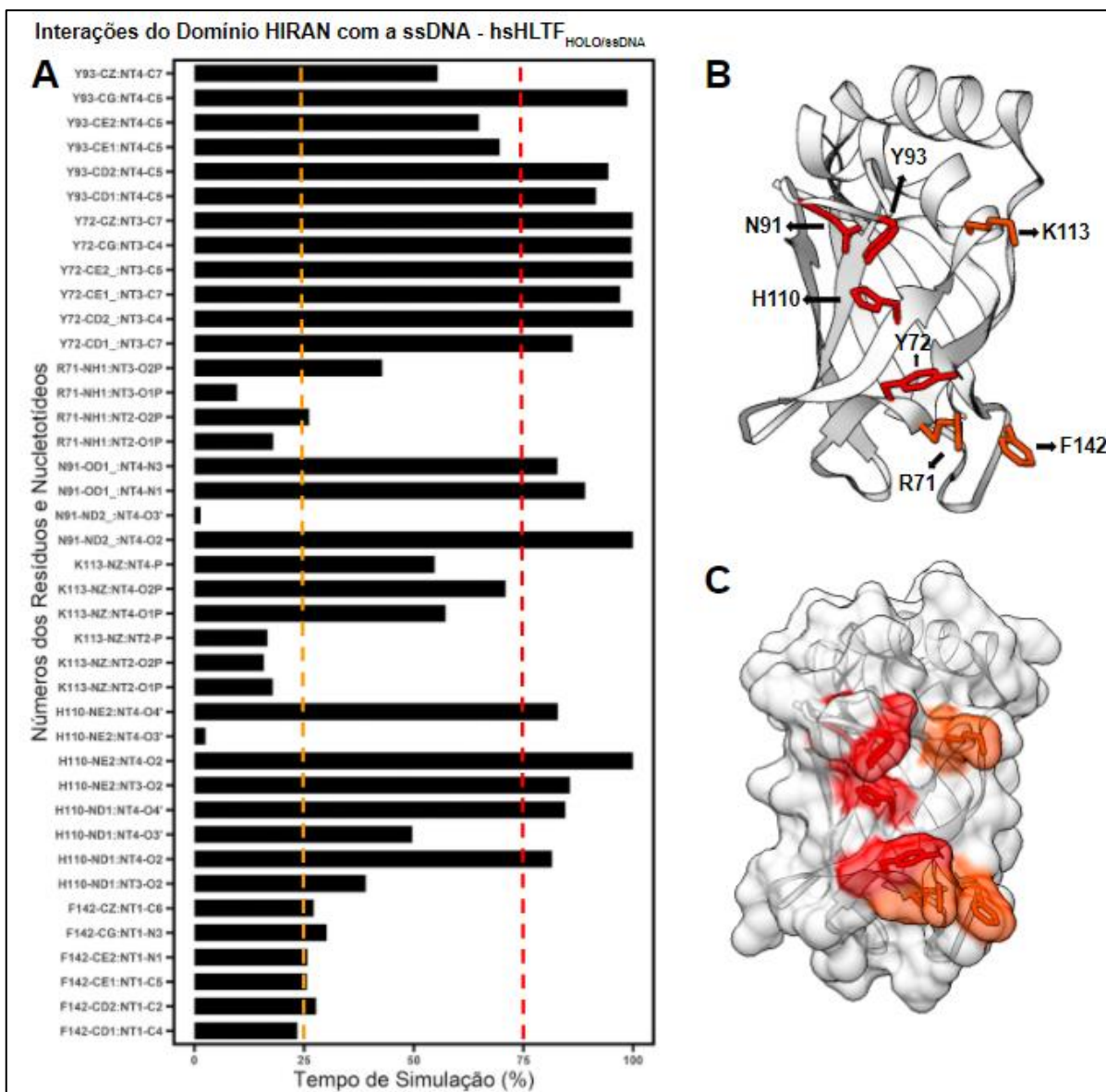


Gráfico de contatos gerado a partir das informações extraídas da ferramenta *cpptraj* e do comando *nativecontacts* do modelo hsHLTF_{HOLo/ssDNA}. Em A, os átomos dos resíduos que interagem com seus respectivos átomos de nucleotídeos (NT), em B, a estrutura terciária do modelo e em C, a representação destes resíduos com sua superfície molecular. Resíduos que apresentaram contatos durante um período maior que 75% do tempo de simulação estão representados em vermelho e aqueles que apresentaram entre 25% e 75% estão em laranja.

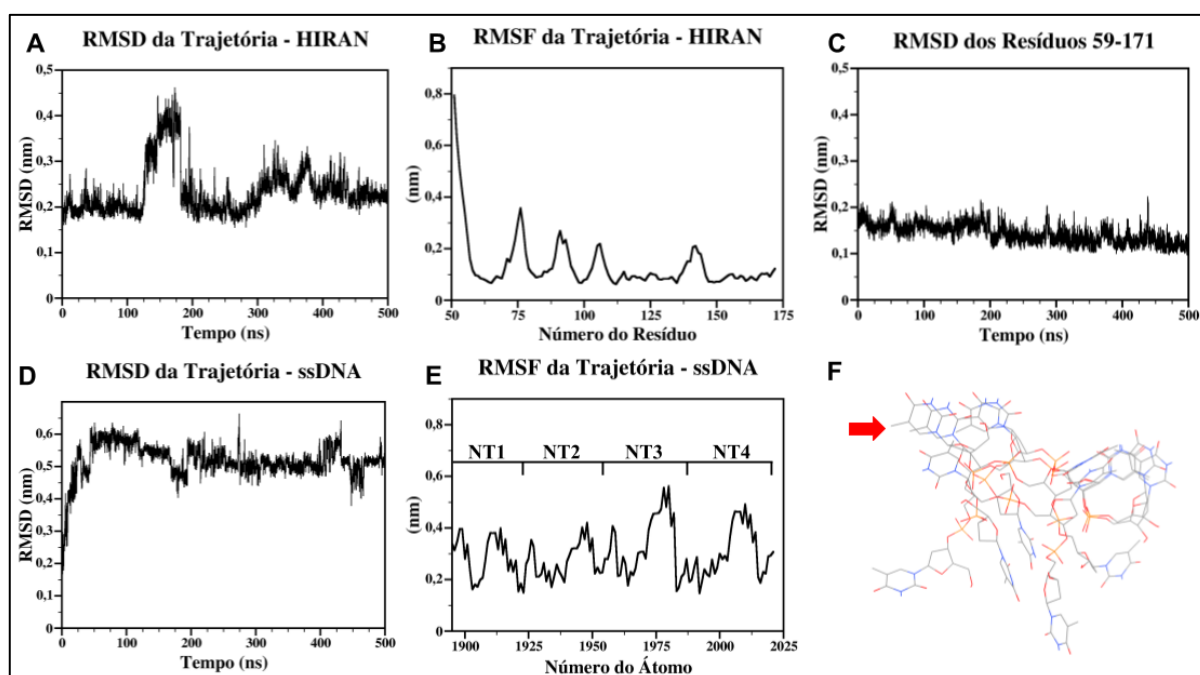
As interações de empilhamento de elétrons π das tirosinas Y72 e Y93 com a ssDNA, tal como as interações eletrostáticas da N91 e H110, se mantiveram ao longo de todo o tempo de simulação e demonstraram a importância destes resíduos na estabilização da fita dentro da estrutura terciária do sítio de interação. Os contatos dos resíduos R71 e K113 com ssDNA se

mantiveram por 25% a 75% do tempo de simulação, demonstrando que também ajudam a estabilizar os nucleotídeos.

5.5.2) DOMÍNIO HIRAN DA PROTEÍNA HLTF EM SUA CONFORMAÇÃO LIVRE COM A ssDNA

A simulação dinâmica molecular do modelo $hsHLTF_{APO/ssDNA}$ foi utilizada para verificar se a mudança conformacional na região do sítio dos resíduos R71 e Y72 ocorre por meio da indução com o ligante. Os gráficos de RMSD e RMSF e o agrupamento das estruturas representativas foram feitos para analisar a dinâmica estrutural do modelo (**Figura 30**).

Figura 30: Dinâmicas estruturais das porções proteica e nucleotídica do modelo $hsHLTF_{APO/ssDNA}$.



Análises das dinâmicas estruturais do domínio HIRAN da HLTF interagindo com a ssDNA, no modelo $hsHLTF_{APO/ssDNA}$. Em A, O gráfico RMSD e, em B, o de RMSF de todos os resíduos do modelo. Em C, gráfico de RMSD parcial, envolvendo os resíduos 56-173. Em D e E, os gráficos de RMSD e RMSF da ssDNA, respectivamente. Em F, as 5 estruturas representativas da etapa de produção da ssDNA extraída a partir do algoritmo de agrupamento. A seta vermelha evidencia o nucleotídeo que parece com a Y72, exposta para a cavidade de interação.

O gráfico de RMSD de todos os resíduos (**Figura 30 A**) demonstrou uma estrutura que ainda não atingiu o seu equilíbrio e o gráfico de RMSF (**Figura 30 B**), similar a etapas anteriores e ao sistema sem a ssDNA, demonstrou que a região N-terminal possui uma grande mobilidade e desestruturação de estrutura secundária. Portanto, o gráfico de RMSD dos resíduos 59-171 foi traçado e demonstrou que o domínio está em equilíbrio, não apresentando

variações espaciais tal como a **Figura 30 A**. O RMSD médio de todos os resíduos foi de $0,23 \pm 0,05$ nm e médio parcial de $0,14 \pm 0,02$ nm, valores próximos aos de RMSD observado nas réplicas do modelo hsHLTF_{APO} livre, porém menores, devido a restrição espacial imposta pela presença da ssDNA.

As análises do gráfico de RMSF permitiram inferir que as regiões de *loops* ainda continuam com grande mobilidade, diferente do modelo hsHLTF_{HOLO/ssDNA}, e que como a cavidade de interação não foi formada, essas estruturas não encontraram um mínimo energético local nem sua conformação mais estável. Com esses dados, podemos hipotetizar que o tempo necessário para observar o mecanismo completo de formação deste sítio, a partir de sua estrutura livre, seja superior ao utilizado neste trabalho, necessitando de simulações na escala de microssegundos.

Os gráficos de dinâmica estrutural e o agrupamento das estruturas representativas da ssDNA (**Figura 30 D, E e F**) demonstrou uma mobilidade de todos os nucleotídeos e uma estrutura fora do equilíbrio dado o tempo de simulação. A análise dos quadros gerados pela etapa de produção e do gráfico de contatos nativos e não-nativos (**Figura 31**) evidenciou que o resíduo Y93, quando exposto para o interior do sítio e sem restrição espacial, consegue parear e interagir com um dos nucleotídeos da ssDNA (**Figura 30 F**). A Y93 pode ser o resíduo inicial de interação com a ssDNA e induzir a formação da cavidade de interação ao permitir que a Y72 interaja com outro nucleotídeo, por meio da mudança conformacional com a R71, e consequentemente a estabilização da fita a partir das interações eletrostáticas dos resíduos N91, K113 e H110. Kile e colaboradores (2015) relataram que a mutação dos resíduos Y72 e Y93 por um resíduo de alanina impede o domínio HIRAN da HLTF de interagir e complexar com a ssDNA.

O gráfico de interações do domínio HIRAN com a ssDNA (**Figura 31**) demonstrou que a Y93 de fato interage com o NT4 por volta de 40% do tempo de simulação. Além disso, a estabilização da fita nas proximidades do modelo hsHLTF_{APO/ssDNA} foi realizada majoritariamente (maior que 50% do tempo de simulação) por interações eletrostáticas dos resíduos R71, N91 e H110. Estes fatos permitem hipotetizar uma etapa inicial para formação da cavidade de interação e que consequentemente permitirá o pareamento das tirosinas com os nucleotídeos NT4 e NT3. Ao contrário do modelo do domínio HIRAN na sua forma complexada, a R71 obteve um tempo de interação com a ssDNA maior pois o mesmo está exposto para o interior do sítio, enquanto a Y72 está externalizada e possui apenas interações transientes com o NT3 (menores que 20% do tempo de simulação). A F142 não apresentou

interação com o NT1, sugerindo que seu pareamento com o nucleotídeo deve ocorrer após a formação completa do bolsão, sendo então uma das etapas finais para a estabilização do domínio HIRAN da HLTF com uma fita simples de material genético.

Figura 31: Resíduos do modelo $hsHLTF_{APO/ssDNA}$ que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.

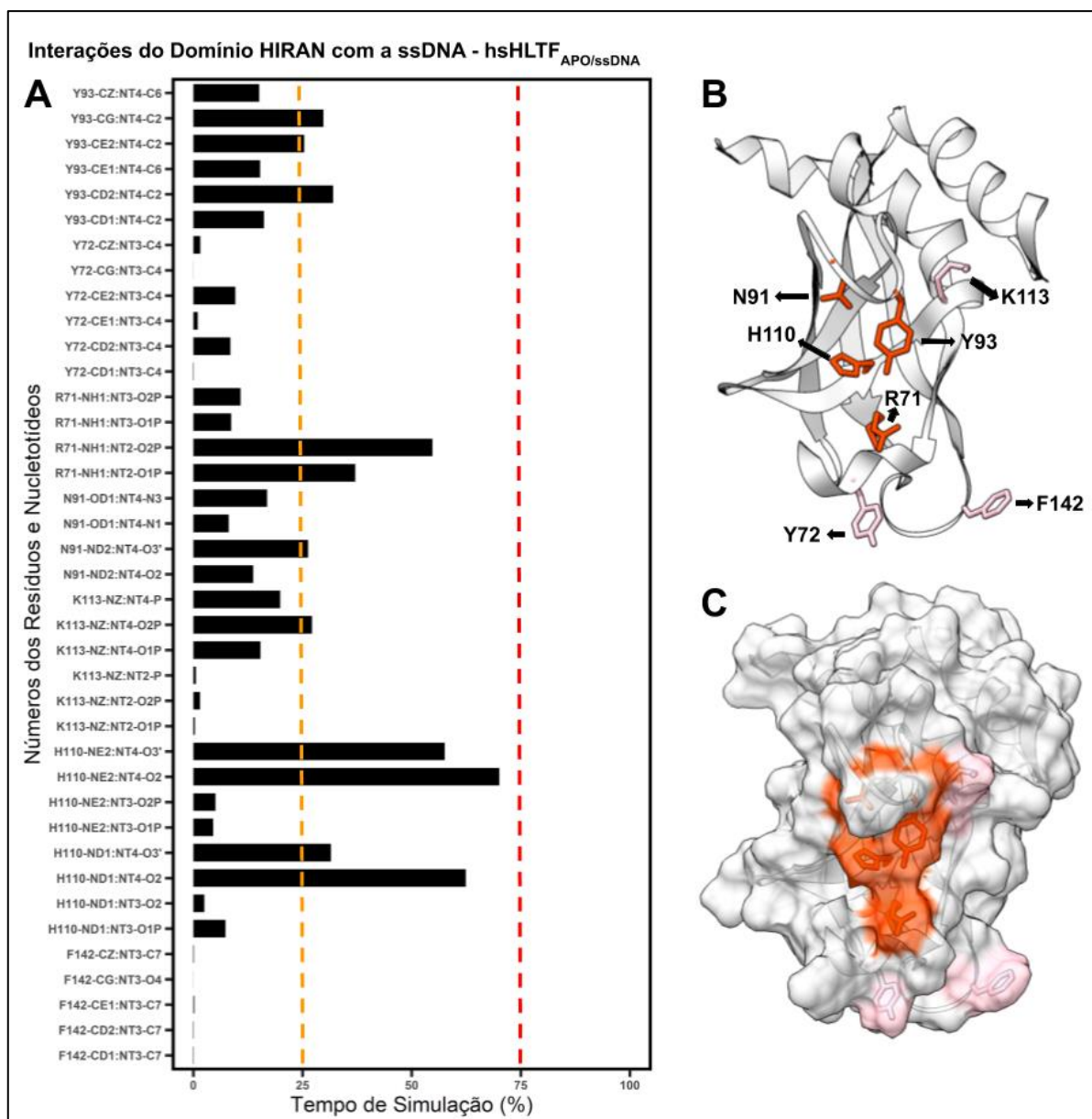
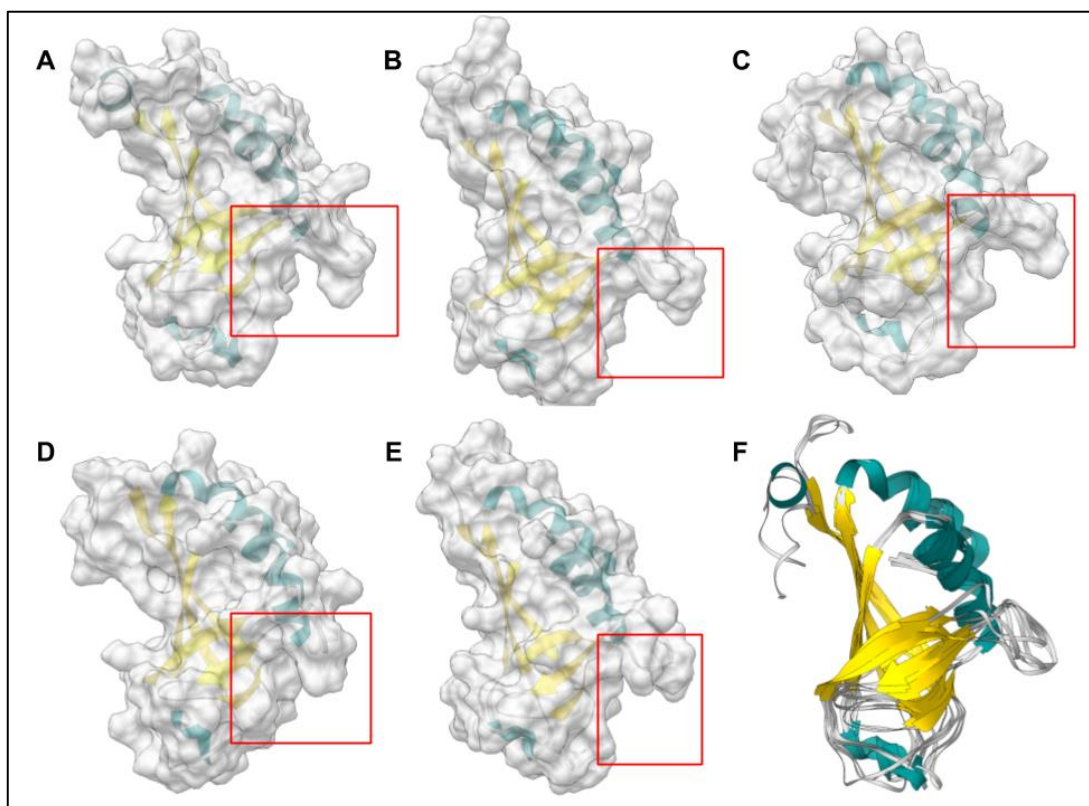


Gráfico de contatos gerado a partir das informações extraídas da ferramenta *cpptraj* e do comando *nativecontacts* do modelo $hsHLTF_{APO/ssDNA}$. Em A, os átomos dos resíduos que interagem com seus respectivos átomos de nucleotídeos (NT), em B, a estrutura terciária do modelo e em C, a representação destes resíduos com sua superfície molecular. Resíduos que apresentaram contatos durante um período entre 25% e 75% do tempo de simulação estão representados em microssegundos laranja e os menores que 25% estão em rosa.

O agrupamento das estruturas representativas da porção proteica do sistema hsHLTF_{APO/ssDNA} (**Figura 32**) demonstrou que não ocorre a formação da cavidade de interação com 500 ns de simulação e que somente a Y93 se encontra externalizada para a interação com os nucleotídeos.

Figura 32: Estruturas representativas da etapa de produção do modelo hsHLTF_{APO/ssDNA} sem a formação completa da cavidade de interação.



Estruturas representativas extraídas através do algoritmo de agrupamento e valores de corte baseado no RMSD global do modelo hsHLTF_{APO/ssDNA} com ssDNA. De A a E estão representadas as estruturas obtidas com sua superfície molecular a mostra. Em F, a sobreposição de todos os modelos sem a superfícies molecular. Os retângulos vermelhos indicam a posição do sítio de interação na estrutura terciária. Em amarelo estão representadas as folhas betas, em ciano as alfa-hélices e em cinza claro a superfícies molecular das estruturas

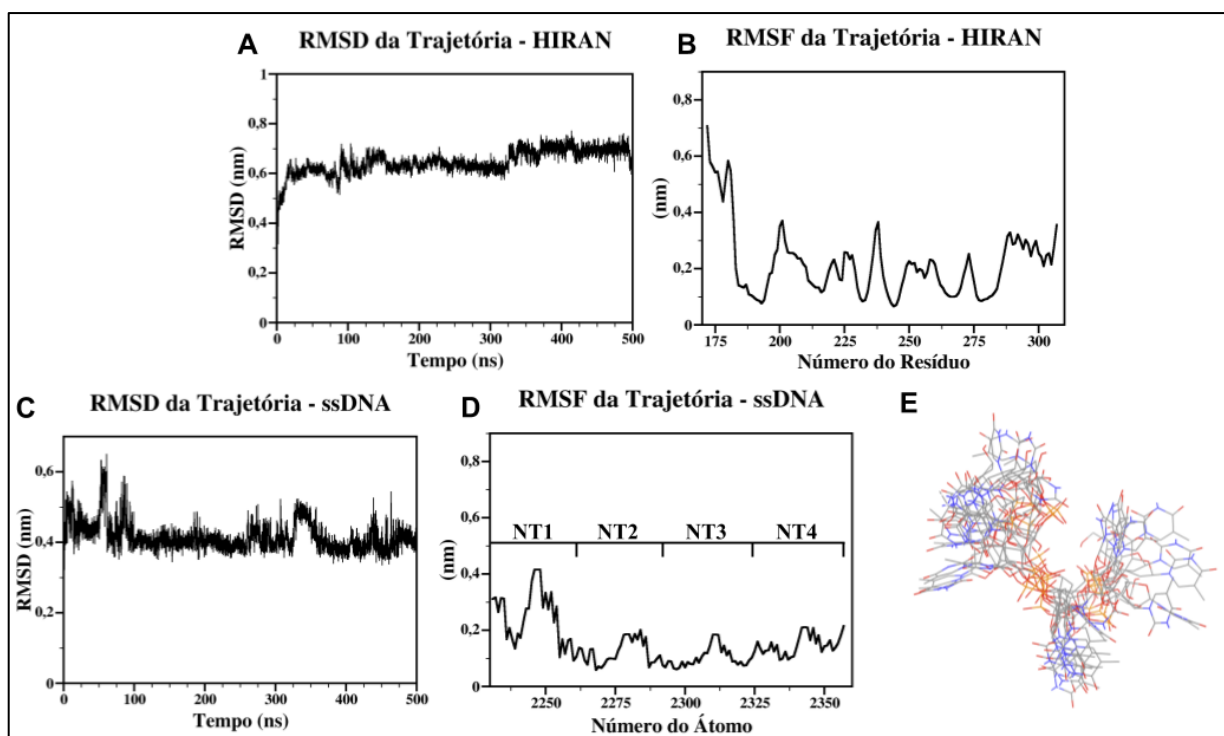
5.5.3) DOMÍNIO HIRAN DA PROTEÍNA RAD5

O modelo scRad5_{ssDNA} foi submetido a dinâmica molecular com a ssDNA a fim de observar o comportamento estrutural dos resíduos que formam o sítio de interação, hipotetizados a partir da modelagem comparativa, e compará-lo ao domínio HIRAN da HLTF e ao modelo scRad5 livre.

As análises de dinâmica, estabilidade e mobilidade foram feitas a partir dos gráficos de RMSD, RMSF e agrupamento de estruturas representativas (**Figura 33**). Em seguida, os

gráficos de DSSP de formação de estruturas secundárias (**Figura 37**) e a contagem das mesmas foram feitas (**Figura 38**).

Figura 33: Dinâmicas estruturais das porções proteica e nucleotídica do modelo scRad5_{ssDNA}.



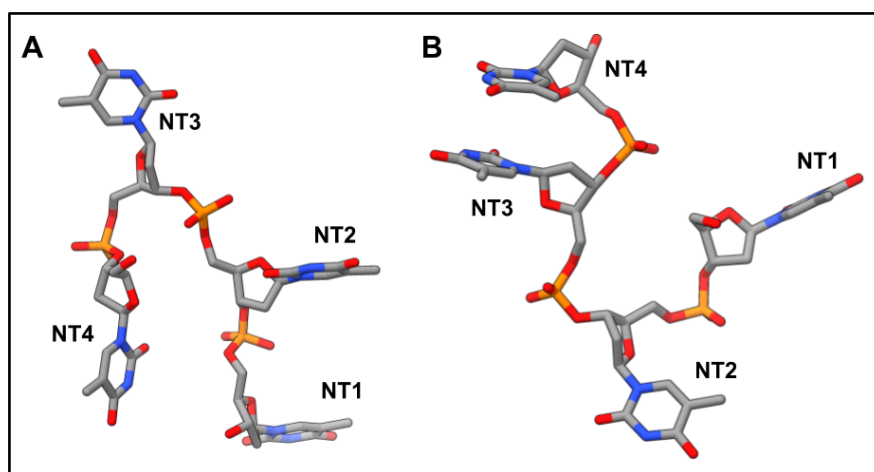
Análises das dinâmicas estruturais do domínio HIRAN da HLTF interagindo com a ssDNA, no modelo scRad5_{ssDNA}. Em A, O gráfico RMSD e, em B, o de RMSF de todos os resíduos do modelo. Em C e D, os gráficos de RMSD e RMSF da ssDNA, respectivamente. Em E, as 7 estruturas representativas da etapa de produção da ssDNA extraída a partir do algoritmo de agrupamento.

O gráfico de RMSD de todos os resíduos da porção proteica (**Figura 33 A**) demonstrou uma estrutura equilibrada sem grande variação nas distâncias dos átomos, indicando que a fita de DNA interferiu na estrutura terciária do domínio HIRAN da Rad5, quando comparado com o mesmo na forma livre (**Figura 15 D**). O RMSF (**Figura 33 B**) demonstrou que os resíduos possuem grande mobilidade e indica que o sistema não encontrou uma conformação estável, tal como observado no modelo hsHLTF_{APOssDNA}, tendo o modelo hsHLTF_{HOLO/ssDNA} como referência. O RMSD médio dos resíduos foi de $0,54 \pm 0,05$ nm, similar ao das réplicas do mesmo modelo sem a ssDNA, mostrando que a mesma não contribuiu para a diminuição da motilidade dos resíduos, apesar da restrição espacial. Este dado indica que a estrutura do HIRAN não encontrou uma conformação, junto a fita de DNA, com um mínimo energético, tal como observado no tópico 5.5.1 (modelo hsHLTF_{HOLO/ssDNA}).

As análises estruturais dos nucleotídeos revelaram que, apesar de um valor menor de RMSD (**Figura 33 C**) e RMSF (**Figura 33 D**), quando comparado com os modelos hsHLTF_{HOLO} e hsHLTF_{APO}, a ssDNA não encontrou uma região de conformação que apresentasse uma homogeneidade estrutural (**Figura 33 E**), a partir de uma inspeção visual. Porém, o raio de movimentação espacial da fita é menor, o que resultou em estruturas mais agrupadas. A não formação da cavidade de interação e a internalização da Y214 dentro do próprio domínio, durante a etapa de produção, contribuíram para que os nucleotídeos não encontrassem um mínimo energético local próximo à suposta região de interação. Isto implica que a estrutura tridimensional obtida pela modelagem comparativa não se apresenta como domínio HIRAN da Rad5 na sua conformação complexada com o DNA.

O gráfico de interações (**Figura 35**) do modelo scRad5_{ssDNA} revelou que, por mais que a Y214 esteja exposta para o possível sítio de interação, seu contato predominante (maior que 75% do tempo de simulação) foi com a porção do açúcar do NT3 (O5' e O4'). Este mesmo resíduo possui algumas interações com o N1 e C2 do NT2 evidenciando a mobilidade da ssDNA. As interações observadas não podem ser caracterizadas como empilhamento de elétrons π , tendo em vista que foram contatos com átomos específicos e o mesmo comportamento não foi observado para os outros átomos da cadeia lateral da tirosina. A fita de DNA apresentou uma configuração espacial diferente quando comparada com as encontradas no modelo hsHLTF_{HOLO/ssDNA} (**Figura 34**). Apesar da alta motilidade e da diferente conformação, a cadeia fosfato se tornou mais exposta e mais estável. A Y195 não internalizada para o sítio de interação possui somente interações momentâneas (menos de 10% do tempo de simulação) com o NT1, evidenciando novamente a motilidade dos resíduos expostos para o solvente e da ssDNA.

Figura 34: A diferença conformacional da ssDNA no modelo scRad5_{ssDNA}.



Após a etapa de *docking* molecular e ao longo da trajetória da dinâmica molecular, a ssDNA presente no sistema scRad5_{ssDNA} adquiriu uma conformação diferente a aquela apresentada no hsHLTF_{HOLO/ssDNA}. Em A, representado o modelo de ssDNA do scRad5_{ssDNA} e em B o do hsHLTF_{HOLO/ssDNA}. Em laranja está representado os átomos de fósforo, em vermelho os átomos de oxigênio, em azul os átomos de nitrogênio e em cinza os átomos de carbono. NT é a contração para nucleotídeo.

Figura 35: Resíduos do modelo scRad5_{ssDNA} que interagem com a ssDNA ao longo da etapa de produção da dinâmica molecular.

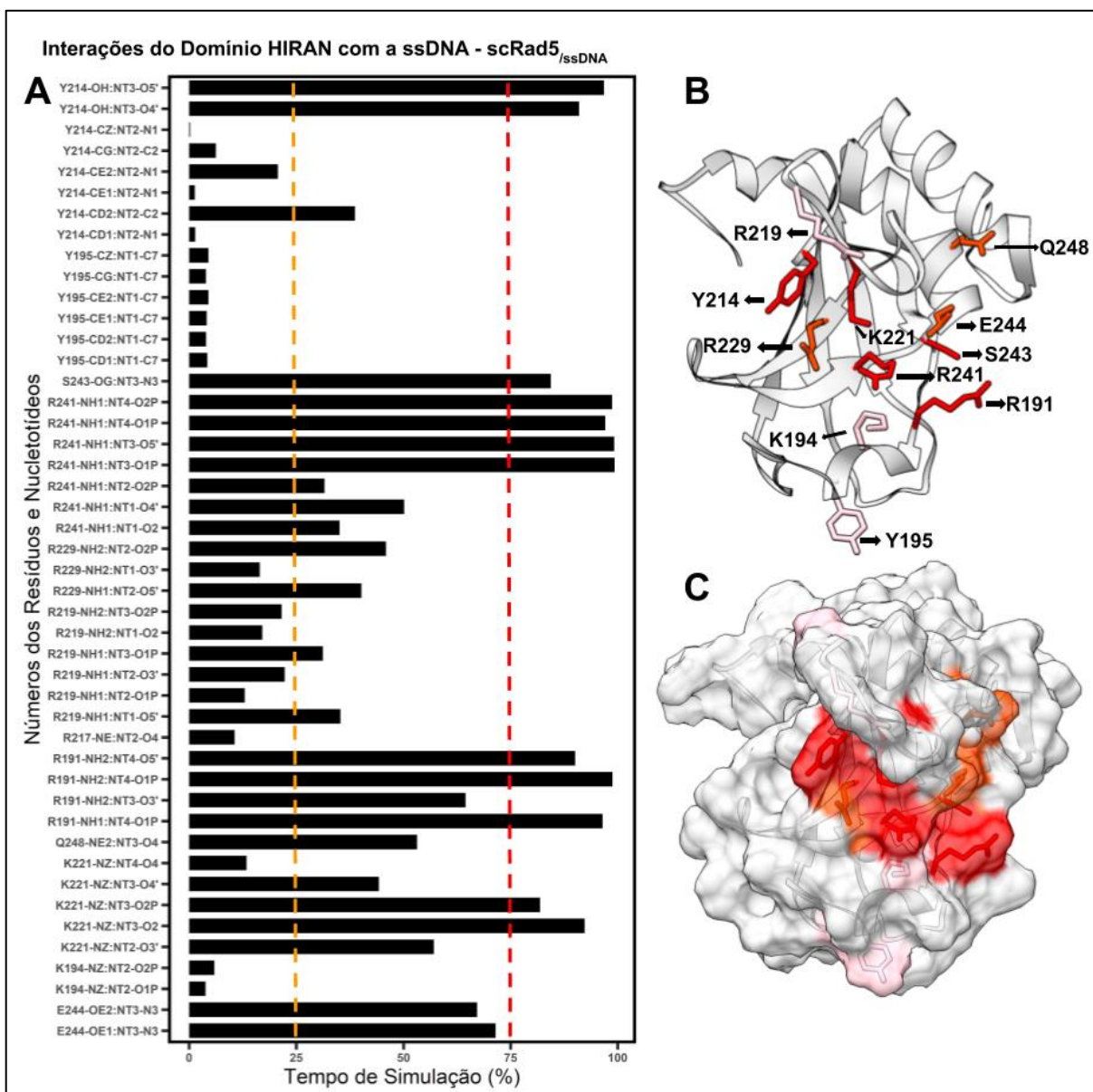


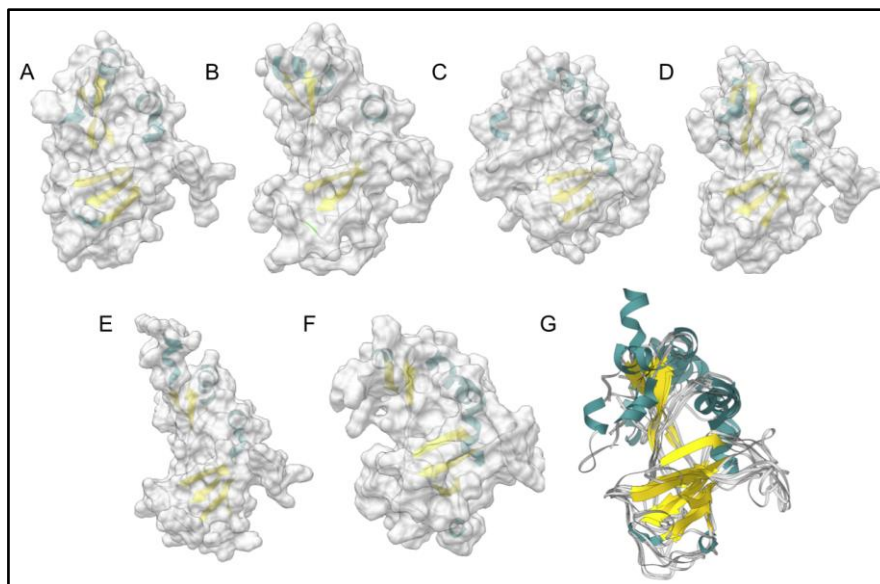
Gráfico de contatos gerado a partir das informações extraídas da ferramenta *cpptraj* e do comando *nativecontacts* do modelo scRad5_{ssDNA}. Em A, os átomos dos resíduos que interagem com seus respectivos átomos de nucleotídeos (NT), em B, a estrutura terciária do modelo e em C, a representação destes resíduos com sua superfície molecular. Resíduos que apresentaram contatos durante um período maior que 75% do tempo de simulação estão representados em vermelho, aqueles entre 25% e 75% estão em laranja e os menores que 25% estão em rosa.

A ssDNA se manteve estável nas proximidades do domínio HIRAN da Rad5 por diversas interações eletrostáticas na região de interação: os resíduos S243, R241, N191 e K221 obtiveram interações com os nucleotídeos por mais de 75% do tempo de simulação, os resíduos R229, R219, Q248 e E244 entre 25% e 75% e o resíduo K194 por aproximadamente 10%. Tal como mostrado na **Figura 14**, há a presença de diversos aminoácidos com cadeia lateral de momento positivo na região do possível sítio de interação com DNA e sugere-se uma maior estabilização da fita por parte destas interações. Os modelos hsHLTF_{HOLO} e hsHLTF_{APO}, quando se observa apenas as cadeias laterais, possuem apenas 5 resíduos que interagem eletrostaticamente com o DNA (R71, N91, K113, H110 e Y92). Em comparação, o modelo scRad5 apresentou 9 resíduos com possíveis contatos: N191, K194, R219, K221, R229, R241, S243, E244 e Q248.

Fan e colaboradores (2018), estudando a interação da proteína TcaR com a ssDNA, demonstraram por meio da dinâmica molecular a importância das interações eletrostáticas de resíduos como a arginina, serina e glutamina na estabilização da fita. Além disto, observaram a diminuição da motilidade das regiões que estão em contato com a ssDNA após a complexação com a mesma. Torabi e colaboradores (2016) também por dinâmica molecular, demonstraram que as interações eletrostáticas da proteína RBP4 são críticas para a sua interação com ssDNA.

O agrupamento das estruturas representativas da porção proteica do domínio HIRAN da Rad5 (**Figura 36**) mostra que, com a presença da ssDNA, os resíduos apresentam maior estabilidade (**Figura 37**) e maior número de estruturas secundárias (**Figura 38**). Porém, a provável cavidade de interação com a ssDNA não se forma, se comportando de forma similar ao modelo hsHLTF_{APO}, corroborando com os dados observados na **Figura 33**.

Figura 36: Estruturas representativas da etapa de produção do modelo scRad5_{ssDNA}.



Estruturas representativas extraídas através do algoritmo de agrupamento e valores de corte baseado no RMSD global do modelo scRad5_{ssDNA}. De A a F estão representadas as estruturas obtidas com sua superfície molecular a mostra. Em G, a sobreposição de todos os modelos sem a superfícies molecular. Em amarelo, estão representadas as folhas betas, em ciano as alfa-hélices e em cinza claro a superfícies molecular das estruturas

Figura 37: Estruturas secundárias formadas ao longo da etapa de produção do modelo scRad5_{ssDNA}.

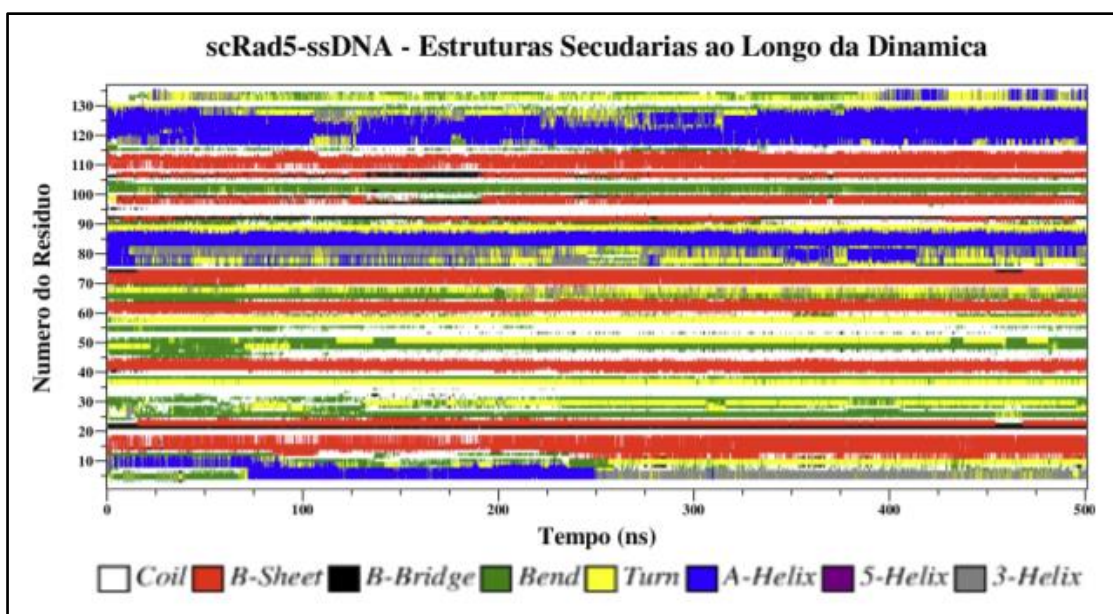


Gráfico de análise de estruturas secundárias formadas, gerado pelo programa DSSP, do modelo scRad5_{ssDNA}, com a presença da ssDNA ao longo da trajetória. Em branco, estruturas do tipo *Coil*, em vermelho do tipo Folhas Beta, em preto do tipo *Beta-Bridge*, em verde do tipo *Bend*, em amarelo do tipo *Turn*, em azul do tipo Alfa-Hélice, em roxo e em cinza do tipo Alfa-Hélice de 5 e 3 resíduos, respectivamente. As numerações dos resíduos estão normalizadas, contando como resíduo 1 o primeiro resíduo da estrutura.

Figura 38: Número de resíduos com estruturas secundárias ao longo da etapa de produção do modelo scRad5_{ssDNA}.

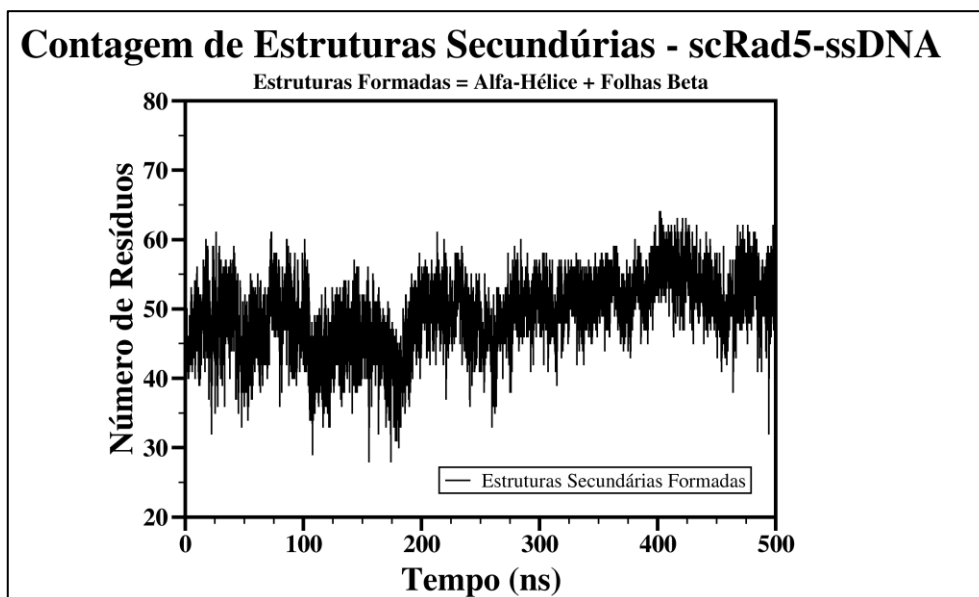


Gráfico de contagem dos resíduos que se estruturam em estruturas secundárias de alfa-hélice e folha beta realizada pelo programa DSSP do modelo scRad5_{ssDNA}.

As análises das estruturas secundárias sugerem que o modelo scRad5_{ssDNA}, se comporta de forma mais estável, quando comparado o tópico 5.2 (DM com o modelo livre), mantendo suas estruturações e aumentando a sua contagem. De acordo com a **Figura 37**, a alfa-hélice $\alpha 1$ (entre os resíduos 70 e 90) foi a que apresentou maior desestruturação. A contagem média do número de resíduos com estruturas secundárias foi de $49,43 \pm 5,35$, maior que as réplicas do modelo scRad5 livre: (a) Réplica 1: $42,30 \pm 2,70$; (b) Réplica 2: $42,84 \pm 5,21$; Réplica 3: $44,06 \pm 5,35$. Juntamente com a maior homogeneidade do agrupamento das estruturas representativas, observadas pela inspeção visual, os dados sugerem que o domínio HIRAN da Rad5 possui uma maior estabilidade na presença da ssDNA.

6) CONCLUSÕES E PERSPECTIVAS

Os métodos computacionais aplicados neste trabalho se mostraram eficientes e obtiveram resultados de acordo com dados encontrado em publicações na literatura. Com isso, informações moleculares a respeito do mecanismo de ação do domínio HIRAN foram adquiridas e poderão contribuir para o entendimento das proteínas envolvidas no processo de reconhecimento da forquilha de replicação parada.

Os resultados de modelagem comparativa permitem concluir que foi obtido um modelo tridimensional do domínio HIRAN da Rad5 satisfatório, sem aminoácidos em posições irregulares e sobreposições de nuvens eletrônicas. Modelo este que foi utilizado na dinâmica e *docking* molecular e que permitiu inferências comparativas com o domínio HIRAN da HLTF para o início do estudo evolutivo do mesmo.

A dinâmica molecular mostrou que o modelo scRad5 possui um maior grau de liberdade quando comparado com o da HLTF (hsHLTF_{APO}), em sua conformação livre. Ambos os modelos apresentaram regiões similares com alta dinâmica estrutural e sugerem que a conformação de ligação a ssDNA só é adquirida na presença da mesma. O HIRAN da HLTF apresentou uma maior estabilidade estrutural, enquanto o da Rad5 sofreu algumas desnaturações de estruturas secundárias. A instabilidade tridimensional do modelo scRad5 sugere a presença de domínios adjacentes para a sua estabilização. O domínio HIRAN da HLTF na sua conformação de interação ao ssDNA, modelo hsHLTF_{HOLO/ssDNA}, possui uma maior estabilidade conformacional e todos os modelos se tornaram mais estáveis na presença da fita de DNA.

As análises estruturais do sítio de interação com ssDNA, em todos os modelos analisados, revelaram a presença de 2 estados: um aberto e um fechado. Na HLTF, a conformação espacial das cadeias laterais dos resíduos R71 e Y72 são diferentes, enquanto na Rad5, isto ocorre devido aos resíduos K194 e Y195. Durante todas as etapas de produção das dinâmicas, não foi observado esta transição de estados deste sítio, o que sugere a realização de DM na escala de microssegundos ou utilizar outras metodologias, como a simulação de dinâmica molecular acelerada, para obter mais dados a respeito deste fenômeno.

O trabalho sugere um mecanismo de indução de conformação por parte da ssDNA na transição do estado livre do domínio HIRAN para o seu estado complexado. Na HLTF, este reconhecimento inicial pode ser atribuído a Y93, enquanto na Rad5 não se pôde obter conclusões assertivas.

As análises de contatos atômicos de cadeias laterais dos resíduos com os nucleotídeos revelaram uma maior presença de interações eletrostáticas do domínio HIRAN da Rad5. Na HLTF, tal como observado em trabalhos anteriores, os resíduos responsáveis pela estabilização da fita de ssDNA observados foram: R71, Y72, N91, Y92, H110, K113 e F142. Na Rad5, além da possível interação empilhamento de elétrons π das tirosinas Y195 e Y214, sugere-se os resíduos N191, K194, R219, K221, R229, R241, S243, E244 e Q248, totalizando um número maior de aminoácidos envolvidos na interação com a ssDNA.

Os domínios HIRAN da Rad5 e da HLTF apresentam comportamentos dinâmicos estruturais similares, sugerindo uma conservação evolutiva. Porém, as análises de interações com a ssDNA apontam para que a Rad5 consiga interagir com diferentes conformações de nucleotídeos, indicando o envolvimento de mais resíduos nesta estabilização. Essa diferença pode ter sido o motivo da SHPRH dividir funções com a HLTF no reconhecimento da forquilha de replicação parada. Dadas estas hipóteses, caso o domínio HIRAN não consiga interagir tempo suficiente com a ssDNA ou necessite de uma estrutura específica de nucleotídeos, a SHPRH pode atuar como forma de impedir que a forquilha de replicação entre em colapso, iniciando então a sinalização para a TLS. A Rad5, por ser possivelmente mais permissiva, é capaz de reconhecer diferentes tipos de ssDNA e sinalizar tanto a TLS quanto a forquilha de replicação reversa.

As perspectivas deste trabalho, e as perguntas deixadas pelo mesmo, estão relacionadas a novos experimentos e análises. Indagações como: a sinalização da TLS e da forquilha de replicação reversa, são tempo dependentes? O tempo de interação do domínio HIRAN com a ssDNA é crucial para este processo? As proteínas HLTF e SHPRH são realmente ortólogas à Rad5? Há uma conservação estrutural apesar das diferentes estruturas primárias, ao longo da evolução? Os dados *in silico* são compatíveis com as análises *in vitro*? Estas respostas podem ser obtidas por meio de: (a) construção de árvores filogenéticas das proteínas Rad5, HLTF e SHPRH; (b) modelagem e análises estruturais de domínios HIRAN de outros organismos, a partir de um caminho evolutivo; (c) simulações de *docking* e dinâmica molecular com outros modelos de ssDNA, não sendo somente uma poli-T; (d) expressão heteróloga do HIRAN e outros domínios da Rad5, HLTF e SHPRH para resolução das estruturas; (e) ensaios *in vitro* de afinidade com diferentes conformações e estruturas primárias de ssDNA; (f) correlacionar, de forma evolutiva, as diferenças nas estruturas primárias dessas proteínas com suas estruturas terciárias; (g) estudos de regulação de expressão gênica a partir de dados de RNAseq e regiões promotoras.

7) REFERÊNCIAS

- ANDREWS, C. T.; CAMPBELL, B. A.; ELCOCK, A. H. Direct Comparison of Amino Acid and Salt Interactions with Double-Stranded and Single-Stranded DNA from Explicit-Solvent Molecular Dynamics Simulations. **Journal of Chemical Theory and Computation**, v. 13, n. 4, p. 1794–1811, 2017.
- ALLEN, M. P.; FRENKEL, D. TALBOT, J. Molecular dynamics simulation using hard particles. **Computer Physics Reports**, v. 9, p. 301–353, 1989.
- BAVI, R. et al. Molecular interactions of UvrB protein and DNA from *Helicobacter pylori*: Insight into a molecular modeling approach. **Computers in Biology and Medicine**, v. 75, p. 181–189, 2016.
- BERENDSEN, H.; SPOEL, D. V. D.; DRUNEN, R. V. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1-3, p. 43–56, 1995.
- BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684–3690, 1984.
- BERMAN, H.; HENRICK, K.; NAKAMURA, H. Announcing the worldwide Protein Data Bank. **Nature Structural Biology**, v. 10, n. 12, p. 980, 2003.
- BITTENCOURT-FERREIRA, G; AZEVEDO, W. F. Homology Modeling of Protein Targets with MODELLER. In: AZEVEDO, W. F. **Docking Screens for Drug Discovery**. Humana Press, 2019, p. 231-250.
- BLASTYAK, A. et al. Role of Double-Stranded DNA Translocase Activity of Human HLTf in Replication of Damaged DNA. **Molecular and Cellular Biology**, v. 30, n. 3, p. 684–693, 2010.
- CAPORUSCIO, F.; RASTELLI, G. Exploiting computationally derived out-of-the-box protein conformations for drug design. **Future Medicinal Chemistry**, v. 8, n. 15, p. 1887–1897, 2016.
- CASE, D.A. et al.; **AMBER 2018 Reference Manual**. São Francisco, Universidade da Califórnia, 2018, 923 p.
- CEBALLOS, S. J.; HEYER, W.-D. Functions of the Snf2/Swi2 family Rad54 motor protein in homologous recombination. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, v. 1809, n. 9, p. 509–523, 2011.
- CHANG, C.-E. et al. Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools. **Catalysts**, v. 6, n. 6, p. 82, 2016.
- CHAPLIN, M. **Water Models**. London South Bank University, 2019. Disponível em: http://www1.lsbu.ac.uk/water/water_models.html.
- CHAVEZ, D. A.; GREER, B. H.; EICHMAN, B. F. The HIRAN domain of helicase-like transcription factor positions the DNA translocase motor to drive efficient DNA fork regression. **Journal of Biological Chemistry**, v. 293, n. 22, p. 8484–8494, 2018.
- CHENG, X.; IVANOV, I. Molecular Dynamics. **Methods in Molecular Biology Computational Toxicology**, p. 243–285, 2012.
- CIEMNY, M. et al. Protein–peptide docking: opportunities and challenges. **Drug Discovery Today**, v. 23, n. 8, p. 1530–1537, 2018.
- COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, v. 2, n. 9, p. 1511–1519, 1993.
- CORNELL, W. D. et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179–5197, 1995.
- CORTEZ, D. Replication-Coupled DNA Repair. **Molecular Cell**, v. 74, n. 5, p. 866–876, 2019.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, 1993.
- DE VIVO, M. et al. Role of Molecular Dynamics and Related Methods in Drug Discovery. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4035–4061, 2016.
- DU, X. et al. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. **International Journal of Molecular Sciences**, v. 17, n. 2, p. 144, 2016

- EKUNDAYO, B.; BLEICHERT, F. Origins of DNA replication. **PLoS Genetics**, v. 15, n. 9, p. 1–21, 2019.
- ELSERAFY, M. et al. Rad5, HLTF, and SHPRH: A Fresh View of an Old Story. **Trends in Genetics**, v. 34, n. 8, p. 574–577, 2018.
- FAN, J.-R. et al. Studying the recognition mechanism of TcaR and ssDNA using molecular dynamic simulations. **Journal of Molecular Graphics and Modelling**, v. 80, p. 67–75, 2018.
- FAN, Q. et al. Rad5 coordinates translesion DNA synthesis pathway by recognizing specific DNA structures in *saccharomyces cerevisiae*. **Current Genetics**, v. 64, n. 4, p. 889–899, 2018.
- FERREIRA, L. G. et al. Molecular docking and structure-based drug design strategies. **Molecules**, v. 20, n. 1, p. 13384–13421, 2015.
- FINN, R. D.; et al. Pfam: the protein families database. **Nucleic Acids Research**, v. 42, n. 1, 2014.
- FORNANDER, L. H. et al. Ca²⁺ improves organization of single-stranded DNA bases in human Rad51 filament, explaining stimulatory effect on gene recombination. **Nucleic Acids Research**, v. 40, n. 11, p. 4904–4913, 2012.
- GANESAN, A.; COOTE, M. L.; BARAKAT, K. Molecular dynamics-driven drug discovery: leaping forward with confidence. **Drug Discovery Today**, v. 22, n. 2, p. 249–269, 2017.
- GIBBS, D. R.; DHAKAL, S. Homologous recombination under the single-molecule fluorescence microscope. **International Journal of Molecular Sciences**, v. 20, n. 23, 2019.
- GILDENBERG, M. S.; TODD WASHINGTON, M. Conformational flexibility of fork-remodeling helicase Rad5 shown by full-ensemble hybrid methods. **PLoS ONE**, v. 14, n. 10, p. 1–16, 2019.
- GORDON, J. C. et al. H⁺⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules. **Nucleic Acids Research**, v. 33, n. 1, p. W368–W371, 2005.
- GRASSO, G. et al. Cell penetrating peptide modulation of membrane biomechanics by Molecular dynamics. **Journal of Biomechanics**, v. 73, p. 137–144, 2018.
- GUEDES, I. A.; DE MAGALHÃES, C. S.; DARDENNE, L. E. Receptor-ligand molecular docking. **Biophysical Reviews**, v. 6, n. 1, p. 75–87, 2014.
- HAMED, M. Y. Role of protein structure and the role of individual fingers in zinc finger protein–DNA recognition: a molecular dynamics simulation study and free energy calculations. **Journal of Computer-Aided Molecular Design**, v. 32, n. 6, p. 657–669, 2018.
- HARGREAVES, D. C.; CRABTREE, G. R. ATP-dependent chromatin remodeling: Genetics, genomics and mechanisms. **Cell Research**, v. 21, n. 3, p. 396–420, 2011.
- HAYASHI, M.; KEYAMURA, K.; HISHIDA, T. Cyclin-dependent kinase modulates budding yeast Rad5 stability during cell cycle. **PLoS ONE**, v. 13, n. 9, p. 1–14, 2018.
- HISHIKI, A.; SATO, M.; HASHIMOTO, H. Structure of HIRAN domain of human HLTF bound to duplex DNA provides structural basis for DNA unwinding to initiate replication fork regression. **The Journal of Biochemistry**, v. 167, n. 6, p. 597–602, 2020.
- HOOFT, R. W. W.; VRIEND, G. Errors in protein structures. **Nature**, v. 381, n. 1, p. 272, 1996.
- HOPFNER, K. P. et al. Swi2/Snf2 remodelers: Hybrid views on hybrid molecular machines. **Current Opinion in Structural Biology**, v. 22, n. 2, p. 225–233, 2012.
- HOSPITAL, A. et al. Molecular dynamics simulations: Advances and applications. **Advances and Applications in Bioinformatics and Chemistry**, v. 8, n. 1, p. 37–47, 2015.
- HUANG, D.; CAFLISCH, A. Small Molecule Binding to Proteins: Affinity and Binding/Unbinding Dynamics from Atomistic Simulations. **ChemMedChem**, v. 6, n. 9, p. 1578–1580, 2011.
- IYER, L. M.; BABU, M. M.; ARAVIND, L. The HIRAN domain and recruitment of chromatin remodeling and repair activities to damaged DNA. **Cell Cycle**, v. 5, n. 7, p. 775–782, 2006.
- JO, S. et al. CHARMM-GUI: A web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, v. 29, n. 11, p. 1859–1865, 2008.

- JORGENSEN, W. L.; TIRADO-RIVES, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. **Journal of the American Chemical Society**, v. 110, n. 6, p. 1657–1666, 1988.
- KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577-2637, 1983.
- KAMARAJ, B.; BOGAERTS, A. Structure and function of p53-DNA complexes with inactivation and rescue mutations: A molecular dynamics simulation study. **PLoS ONE**, v. 10, n. 8, p. 1–16, 2015.
- KHAN, F. I. et al. Current updates on computer aided protein modeling and designing. **International Journal of Biological Macromolecules**, v. 85, p. 48–62, 2016.
- KILE, A. C. et al. HLTF's Ancient HIRAN Domain Binds 3' DNA Ends to Drive Replication Fork Reversal. **Molecular Cell**, v. 58, n. 6, p. 1090–1100, 2015.
- KIM, J. M. et al. CHFR functions as a ubiquitin ligase for HLTF to regulate its stability and functions. **Biochemical and Biophysical Research Communications**, v. 395, n. 4, p. 515–520, 2010.
- KORZHNEV, D. M. et al. Solution NMR structure of the HLTF HIRAN domain: a conserved module in SWI2/SNF2 DNA damage tolerance proteins. **Journal of Biomolecular NMR**, v. 66, n. 3, p. 209–219, 2016.
- KRESSE, G.; HAFNER, J. Ab initio molecular dynamics for liquid metals. **Physical Review B**, v. 47, n. 1, p. 558–561, 1993.
- KRIJGER, P. H. et al. HLTF and SHPRH are not essential for PCNA polyubiquitination, survival and somatic hypermutation: Existence of an alternative E3 ligase. **DNA Repair**, v. 10, n. 4, p. 438–444, 2011.
- LAKHANI, B. et al. Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allostery in the MutS DNA Mismatch Repair Protein. **Journal of Physical Chemistry B**, v. 121, n. 9, p. 2049–2061, 2017.
- LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1993.
- LEACH, A. R. Ligand docking to proteins with discrete side-chain flexibility. **Journal of Molecular Biology**, v. 235, n. 1, p. 345–356, 1994.
- LEE, D. et al. SHPRH as a new player in ribosomal RNA transcription and its potential role in homeostasis of ribosomal DNA repeats. **Transcription**, v. 9, n. 3, p. 190–195, 2018.
- LEE, D. et al. SHPRH regulates rRNA transcription by recognizing the histone code in an mTOR-dependent manner. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 17, p. 3424–3433, 2017.
- LENNARD-JONES, J. E. Cohesion. **Proceedings of the Physical Society**, v. 43, n. 5, p. 461–482, 1931.
- LEVITT, M. Accurate modeling of protein conformation by automatic segment matching. **Journal of Molecular Biology**, v. 226, n. 2, p. 507–533, 1992.
- LI, J. et al. Compound F779-0434 causes synthetic lethality in BRCA2-deficient cancer cells by disrupting RAD52-ssDNA association. **RSC Advances**, v. 8, n. 34, p. 18859–18869, 2018.
- LIN, J.-R. et al. SHPRH and HLTF Act in a Damage-Specific Manner to Coordinate Different Forms of Postreplication Repair and Prevent Mutagenesis. **Molecular Cell**, v. 42, n. 2, p. 237–249, 2011.
- LIU, X. et al. Microsecond molecular dynamics simulations reveal the allosteric regulatory mechanism of p53 R249S mutation in p53-associated liver cancer. **Computational Biology and Chemistry**, v. 84, n. 1, p. 107-194, 2020.
- LIU, X. et al. Molecular dynamics simulations and novel drug discovery. **Expert Opinion on Drug Discovery**, v. 13, n. 1, p. 23–37, 2018.
- LÜTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, n. 6364, p. 83–85, 1992.
- MACHADO, L. E. F. et al. PHD domain from human SHPRH. **Journal of Biomolecular NMR**, v. 56, n. 4, p. 393–399, 2013.

- MALLAJOSYULA, S. S. et al. Molecular Dynamics Simulations of Glycoproteins Using CHARMM. **Methods in Molecular Biology Glycoinformatics**, p. 407–429, 2015.
- MARY, V. et al. Experimental Probing and Molecular Dynamics Simulation of the Molecular Recognition of DNA Duplexes by the Flavonoid Luteolin. **Journal of Chemical Information and Modeling**, v. 57, n. 9, p. 2237–2249, 2017.
- MASUDA, Y. et al. Regulation of HLTF-mediated PCNA polyubiquitination by RFC and PCNA monoubiquitination levels determines choice of damage tolerance pathway. **Nucleic acids research**, v. 46, n. 21, p. 11340–11356, 2018.
- MCGINNIS, S.; MADDEN, T. L. BLAST: At the core of a powerful and diverse set of sequence analysis tools. **Nucleic Acids Research**, v. 32, n. 1, p. 20–25, 2004.
- MENG, X.-Y. et al. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. **Current Computer Aided-Drug Design**, v. 7, n. 2, p. 146–157, 2011.
- MORRIS, G. M.; LIM-WILBY, M. Molecular Docking. In: KUKOL, A. **Methods in Molecular Biology**, v. 443, Molecular Modeling of Proteins. Totowa, Humana Press, 2008, p. 365-382.
- MUHAMMED, M. T.; AKI-YALCIN, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. **Chemical Biology and Drug Design**, v. 93, n. 1, p. 12–20, 2019.
- NAQVI, A. A. T. et al. Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships. **Current Topics in Medicinal Chemistry**, v. 18, n. 20, p. 1755–1768, 2019.
- NEELSEN, K. J.; LOPES, M. Replication fork reversal in eukaryotes: From dead end to dynamic response. **Nature Reviews Molecular Cell Biology**, v. 16, n. 4, p. 207–220, 2015.
- NESTER, K.; GAWEDA, K.; PLAZINSKI, W. A GROMOS Force Field for Furanose-Based Carbohydrates. **Journal of Chemical Theory and Computation**, v. 15, n. 2, p. 1168–1186, 2019.
- NOCUA, P. A. et al. Leishmania braziliensis replication protein A subunit 1: Molecular modelling, protein expression and analysis of its affinity for both DNA and RNA. **Parasites and Vectors**, v. 7, n. 1, p. 1–11, 2014.
- NONGKHLAW, M. et al. Motifs Q and I Are Required for ATP Hydrolysis but Not for ATP Binding in SWI2/SNF2 Proteins. **Biochemistry**, v. 51, n. 18, p. 3711–3722, 2012.
- OPRZESKA-ZINGREBE, E. A.; SMIATEK, J. Preferential Binding of Urea to Single-Stranded DNA Structures: A Molecular Dynamics Study. **Biophysical Journal**, v. 114, n. 7, p. 1551–1562, 2018.
- O'DONNELL, M.; LANGSTON, L.; STILLMAN, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. **Cold Spring Harbor Perspectives in Biology**, v. 5, n. 7, p. 1–13, 2013.
- PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. **Biophysical Reviews**, v. 9, n. 2, p. 91–102, 2017.
- PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182–7190, 1981.
- PARULEKAR, R. S. et al. Homology modeling, molecular docking and DNA binding studies of nucleotide excision repair uvrc protein from M. tuberculosis. **Protein Journal**, v. 32, n. 6, p. 467–476, 2013.
- PATEL, D. R.; WEISS, R. S. A tough row to hoe: When replication forks encounter DNA damage. **Biochemical Society Transactions**, v. 46, n. 6, p. 1643–1651, 2018.
- PETERSON, L. E. Small molecule docking of DNA repair proteins associated with cancer survival following PCNA metagene adjustment: A potential novel class of repair inhibitors. **Molecules**, v. 24, n. 3, 2019.
- PETREY, D. et al. Using Multiple Structure Alignments, Fast Model Building, and Energetic Analysis in Fold Recognition and Homology Modeling. **Proteins: Structure, Function and Genetics**, v. 53, n. SUPPL. 6, p. 430–435, 2003.
- PETROVIĆ, D.; WANG, X.; STRODEL, B. How accurately do force fields represent protein side chain ensembles? **Proteins: Structure, Function and Bioinformatics**, v. 86, n. 9, p. 935–944, 2018.
- PEZESHKIAN, W.; KHANDELIA, H.; MARSH, D. Lipid Configurations from Molecular Dynamics Simulations. **Biophysical Journal**, v. 114, n. 8, p. 1895–1907, 2018.

- PINZI, L.; RASTELLI, G. Molecular docking: Shifting paradigms in drug discovery. **International Journal of Molecular Sciences**, v. 20, n. 18, 2019.
- PITTA, K.; KRISHNAN, M. Molecular Mechanism, Dynamics, and Energetics of Protein-Mediated Dinucleotide Flipping in a Mismatched DNA: A Computational Study of the RAD4-DNA Complex. **Journal of Chemical Information and Modeling**, v. 58, n. 3, p. 647–660, 2018.
- POL-FACHIN, L. et al. GROMOS 53A6 GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. **Journal of Chemical Theory and Computation**, v. 8, n. 11, p. 4681–4690, 2012.
- PONTIUS, J.; RICHELLE, J.; WODAK, S. J. Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structure. **Journal of Molecular Biology**, v. 264, n. 1, p. 121–136, 1996.
- POOLE, L. A.; CORTEZ, D. Functions of SMARCAL1, ZRANB3, and HLTF in maintaining genome stability. **Critical Reviews in Biochemistry and Molecular Biology**, v. 52, n. 6, p. 696–714, 2017.
- PRADHAN, S.; DAS, P.; MATTAPARTHI, V. S. K. Characterizing the Binding Interactions between DNA-Binding Proteins, XPA and XPE: A Molecular Dynamics Approach. **ACS Omega**, v. 3, n. 11, p. 15442–15454, 2018.
- QING, P. et al. USP7 regulates the stability and function of HLTF through deubiquitination. **Journal of Cellular Biochemistry**, v. 112, n. 12, p. 3856–3862, 2011.
- QUINET, A.; LEMAÇON, D.; VINDIGNI, A. Replication Fork Reversal: Players and Guardians. **Molecular Cell**, v. 68, n. 5, p. 830–833, 2017.
- REKIK, I. et al. In silico characterization and Molecular modeling of double-strand break repair protein MRE11 from *Phoenix dactylifera* v *deglet nour*. **Theoretical Biology and Medical Modelling**, v. 12, n. 1, p. 1–14, 2015.
- RICKMAN, K.; SMOGORZEWSKA, A. Advances in understanding DNA processing and protection at stalled replication forks. **Journal of Cell Biology**, v. 218, n. 4, p. 1096–1107, 2019.
- ROBERTSON, M. J. et al. Development and Testing of the OPLS-AA/M Force Field for RNA. **Journal of Chemical Theory and Computation**, v. 15, n. 4, p. 2734–2742, 2019.
- ŠALI, A.; BLUNDELL, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. **Journal of Molecular Biology**, v. 234, n. 3, p. 779–815, 1993.
- SALMASO, V.; MORO, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. **Frontiers in Pharmacology**, v. 9, 2018.
- SCHNEIDER, N. et al. Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. **Journal of Computer-Aided Molecular Design**, v. 26, n. 6, p. 701–723, 2011.
- SELVARAJ, C. et al. Molecular dynamics simulations and applications in computational toxicology and nanotoxicology. **Food and Chemical Toxicology**, v. 112, p. 495–506, 2018.
- SHAO, Q.; ZHU, W. Assessing AMBER force fields for protein folding in an implicit solvent. **Physical Chemistry Chemical Physics**, v. 20, n. 10, p. 7206–7216, 2018.
- SHRIVASTAVA, N. et al. Homology modeling of NAD⁺-dependent DNA ligase of the *Wolbachia* endosymbiont of *Brugia malayi* and its drug target potential using dispiro-cycloalkanones. **Antimicrobial Agents and Chemotherapy**, v. 59, n. 7, p. 3736–3747, 2015.
- SILVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 539, 2011.
- SKJEVIK, Å. A. et al. All-atom lipid bilayer self-assembly with the AMBER and CHARMM lipid force fields. **Chemical Communications**, v. 51, n. 21, p. 4402–4405, 2015.
- SNEHA, P.; PRIYA DOSS, C. G. Molecular Dynamics: New Frontier in Personalized Medicine. **Elsevier Inc.** v. 102, n. 1, 2016.
- SONG, Y. et al. High-resolution comparative modeling with RosettaCM. **Structure**, v. 21, n. 10, p. 1735–1742, 2013.
- TARAZI, H.; SALEH, E.; EL-AWADY, R. In-silico screening for DNA-dependent protein kinase (DNA-PK) inhibitors: Combined homology modeling, docking, molecular dynamic study followed by biological investigation. **Biomedicine and Pharmacotherapy**, v. 83, p. 693–703, 2016.

- TOMI, N. S. et al. Analysis of SHPRH functions in DNA repair and immunoglobulin diversification. **DNA Repair**, v. 24, p. 63–72, 2014.
- TORABI, R. et al. An investigation on the interaction modes of a single-strand DNA aptamer and RBP4 protein: a molecular dynamic simulations approach. **Organic & Biomolecular Chemistry**, v. 14, n. 34, p. 8141–8153, 2016.
- TORRES, P. H. M. et al. Key topics in molecular docking for drug design. **International Journal of Molecular Sciences**, v. 20, n. 18, p. 1–29, 2019.
- THOMA, N. et al. Structure of the SWI2/SNF2 chromatin remodeling domain of eukaryotic Rad54. **Nature Structural & Molecular Biology**, v. 12, n. 4, 2005.
- TRIPATHI, A.; VA, B. Molecular Docking: From Lock and Key to Combination Lock. **Journal of Molecular Medicine and Clinical Applications**, v. 2, n. 1, p. 1–19, 2018.
- TURCHICK, A. et al. A cell-penetrating antibody inhibits human RAD51 via direct binding. **Nucleic Acids Research**, v. 45, n. 20, p. 11782–11799, 2017.
- UCHIKOGA, N. et al. Re-Docking Scheme for Generating Near-Native Protein Complexes by Assembling Residue Interaction Fingerprints. **PLoS ONE**, v. 8, n. 7, 2013.
- UNK, I. et al. Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. **Proceedings of the National Academy of Sciences**, v. 105, n. 10, p. 3768–3773, 2008.
- VERLET, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. **Physical Review**, v. 159, n. 1, p. 98–103, 1967.
- VERMA, S. et al. Hydrophobic interactions are a key to MDM2 inhibition by polyphenols as revealed by molecular dynamics simulations and MM/PBSA free energy calculations. **PLoS ONE**, v. 11, n. 2, p. 1–14, 2016.
- VLACHAKIS, D. et al. Current state-of-the-art molecular dynamics methods and applications. **Elsevier Inc.** v. 94, n. 1, 2014.
- VUJANOVIC, M. et al. Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. **Molecular Cell**, v. 67, n. 5, p. 882–890.e5, 2017.
- VYAS, V. K. et al. Homology modeling a fast tool for drug discovery: Current perspectives. **Indian Journal of Pharmaceutical Sciences**, v. 74, n. 1, p. 1–17, 2012.
- WARSHEL, A.; LEVITT, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. **Journal of Molecular Biology**, v. 103, n. 2, p. 227–249, 1976.
- WATERHOUSE, A. et al. SWISS-MODEL: Homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. W1, p. W296–W303, 2018.
- WEBB, B.; SALI, A. Comparative protein structure modeling using MODELLER. **Current Protocols in Bioinformatics**. v. 47, n. 1, 2014
- WORTH, C.L.; GONG, S.; BLUNDELL, T.L. Structural and functional constraints in the evolution of protein families. **Nature Reviews Molecular Cell Biology**, v. 10, n. 10, p. 709–720, 2009.
- XIANG, N. et al. Investigation of the interaction of amyloid β peptide (11–42) oligomers with a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane using molecular dynamics simulation. **Physical Chemistry Chemical Physics**, v. 20, n. 10, p. 6817–6829, 2018.
- XU, X. et al. Involvement of budding yeast Rad5 in translesion DNA synthesis through physical interaction with Rev1. **Nucleic Acids Research**, v. 44, n. 11, p. 5231–5245, 2016.
- YAN, X. C. et al. Improved Description of Sulfur Charge Anisotropy in OPLS Force Fields: Model Development and Parameterization. **Journal of Physical Chemistry B**, v. 121, n. 27, p. 6626–6636, 2017.
- YANG, W. et al. Replisome structure suggests mechanism for continuous fork progression and post-replication repair. **DNA Repair**, v. 81, n. July, p. 102658, 2019.
- ZHANG, Y.; HUANG, S. Exploring the binding mechanism and dynamics of endoms/nucs to mismatched dsdna. **International Journal of Molecular Sciences**, v. 20, n. 20, 2019.

ANEXO

Ao longo do desenvolvimento deste trabalho, foi realizada a tentativa da expressão heteróloga do domínio HIRAN da Rad5. A expressão do domínio resultava em corpos de inclusão, na fração insolúvel, e se tornava solúvel apenas em altas concentrações de ureia. Foram feitas diversas tentativas de troca de tampão de lise e células competentes.

O plasmídeo de expressão, utilizando pET28b (+), foi construído da seguinte maneira:

Sequência de nucleotídeos, otimizada para *E. coli*:

```
CATATGGGATCCGAAAACCTGTATTTTCAGGGTAAAGTTACCTGGAAACGTTTTA
TTGGTGCACCTGCAGGTTACCGGTATGGCAACCCGTCCGACCGTTCGTCCGCTGAA
ATATGGTAGCCAGATGAAACTGAAACGTAGCAGCGAAGAAATTAGCGCCACCAA
AGTTTATGATAGCCGTGGTCGTAAAAAAGCAAGCATGGCAAGCCTGGTTCGCATT
TTTGATATTCAGTATGATCGTGAAATTGGTCGCGTGAGCGAAGATATTGCACAGA
TTCTGTATCCGCTGCTGAGCAGCCATGAAATTAGTTTTGAAGTGACCCTGATCTTT
TGCGATAACAAACGTCTGAGCATTGGCGATAGCTTTATTCTGCAGCTGGATTGTTT
TCTGACCAGCCTGATTTTTGAAGAACGCAATGATGGTGAAAGCCTGATGAAACGT
TAACTCGAGGAATTC
```

O que resulta na seguinte proteína:

Sequência de aminoácidos, contendo o sítio de clivagem pela enzima TEV e a região 169 a 304 da Rad5:

```
ENLYFQG
KVTWKRFIGALQVTGMATRPTVRPLKYGSQMKLKRSSEEISATKVYDSRGRKKASM
ASLVRIFDIQYDREIGRVSEIDIAQILYPLLSSHEISFEVTLIFCDNKRLSIGDSFILQLDCF
LTSLIFEERNDGESLMKR
```

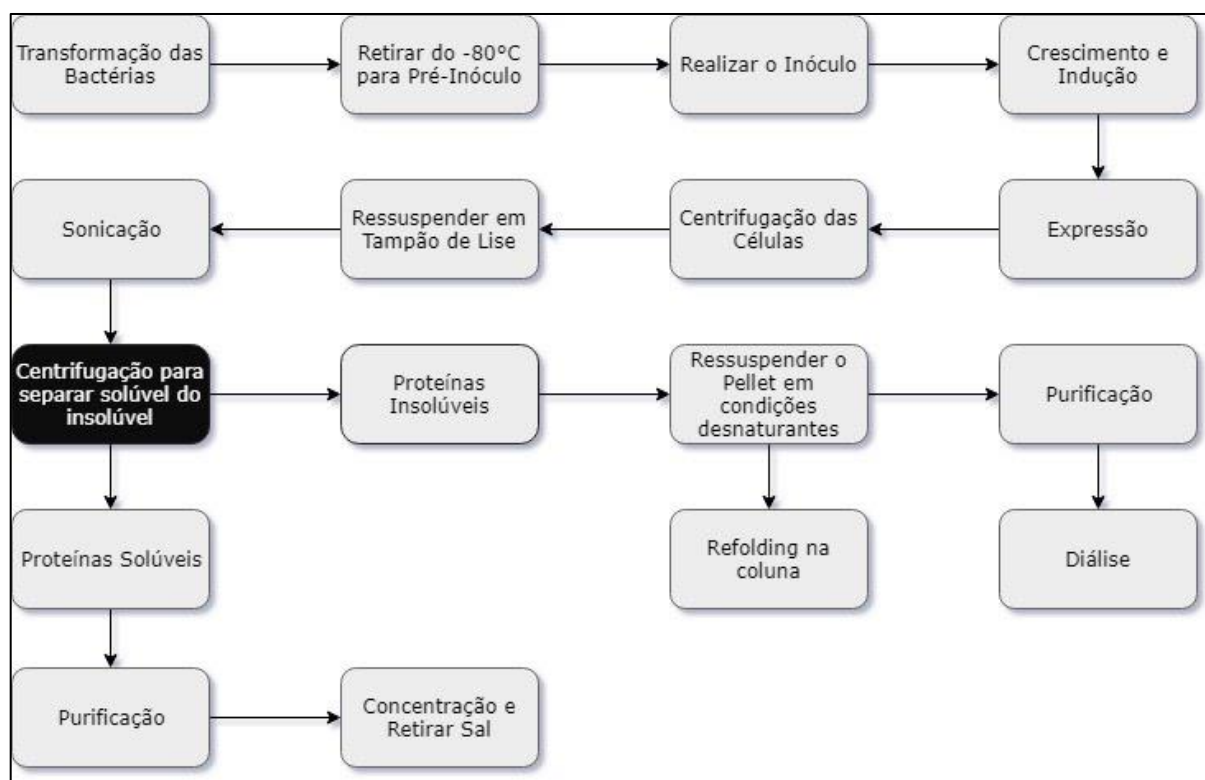
As soluções utilizadas ao longo desta etapa foram:

- Tampão Fosfato: 50mM NaPi pH 7.8, 500-350mM NaCl, 10mM Imidazol;
- Tampão de Eluição: 50mM NaPi pH 7.8, 500-350mM NaCl, 300mM Imidazol;
- Tampão Ureia 8M: Tampão Fosfato + 8M Ureia;
- Tampão Guanidina 6M: Tampão Fosfato + 6M Guanidina;
- TES: 10 mM Tris pH 7.5, 10mM EDTA pH 8.0, 0.5% SDS;
- Tampão Lise Glicerol e Triton X-100: Tampão Fosfato + 1% Glicerol e 0.5% Triton X-100;
- Meios de cultura LB e M9.

A metodologia da expressão heteróloga consistiu em transformar as células de *E. coli* competentes com os plasmídeos e estocá-las a -80 °C para futuros usos. Um pré-inóculo é realizado para crescimento das células que serão inoculadas em um volume maior, até atingir a densidade ótica entre 0,6 e 0,8 a 600nm de comprimento de onda UV. A indução é realizada

com IPTG e tanto o tempo de expressão, quanto a temperatura e a rotação, foram modificados a cada expressão, a fim de encontrar um protocolo favorável. Em seguida estas células são centrifugadas para concentra-las, ressuspensas em um tampão de lise, para liberar o conteúdo intracelular e sonicadas para diminuir a viscosidade e a agregação de dejetos na solução. A verificação do perfil de expressão proteico é feita através do gel de acrilamida SDS-PAGE. A purificação foi realizada utilizando uma coluna de afinidade e concentradas utilizando filtros de 20kDa (**Figura A1**).

Figura A1: Esquema sequencial dos processos utilizados da etapa de expressão heteróloga do domínio HIRAN da Rad5.



Todas as etapas utilizadas para a expressão do domínio HIRAN da Rad5 em *E. Coli* de forma sequencial.

De forma descritivas, as etapas da **Figura A1** estão listadas a seguir:

Transformação Bacteriana: Inicialmente é adicionado 1 μL do plasmídeo em 50 μL de células termocompetentes. Em seguida, a amostra é colocada no gelo, na temperatura de 4 $^{\circ}\text{C}$, durante 30 minutos, depois colocada em um termobloco à 42 $^{\circ}\text{C}$ durante 20 a 40 segundos e por fim durante 3 minutos no gelo novamente. É adicionado 1 mL de meio de cultura e deixado em uma estufa à 37 $^{\circ}\text{C}$ por 1 hora. Após esta etapa, todo o meio de cultura é adicionado a 10 mL de meio cultura com antibiótico e deixado sob agitação, de 180 rpm, a 37 $^{\circ}\text{C}$ durante 14 a 16 horas. Após o crescimento, as células são estocadas a -80 $^{\circ}\text{C}$ em uma solução contendo 0.3 mL de células e 0,6 mL de solução Glicerol 30%.

Metodologia de Expressão em *E. coli* CodonPlus (DE3):

- Pré-Inóculo: 10 µL de células estocadas a -80 °C são adicionadas a 10 mL de meio de cultura contendo 10 µL de canamicina 100 mg/mL e colocado sob agitação, de 180 rpm, a 37 °C durante 14 a 16 horas.
- Inóculo: Todo o volume presente na etapa de pré-inóculo é adicionado à um volume de 250 mL de meio de cultura contendo 250 µL de canamicina 100 mg/mL e deixado sob agitação, de 180 rpm, a 37 °C até a densidade ótica atingir 0,6 (U.V. 660nm).
- Indução: Após atingir a densidade ótica necessária, é adicionado ao meio de cultura a concentração necessária de IPTG (1mM ou 0,4 mM) e deixado sob agitação, de 180 rpm, a 20 °C durante o tempo desejado.

Metodologia de Expressão em *E. coli* Lemo 21 (DE3):

- Pré-Inóculo: 10 µL de células estocadas a -80 °C são adicionadas a 10 mL de meio de cultura contendo 10 µL de canamicina 100 mg/mL e 10 µL de cloranfenicol 50 mg/mL e colocado sob agitação, de 220 rpm, a 37 °C durante 14 a 16 horas.
- Inóculo: Todo o volume presente na etapa de pré-inóculo é adicionado à um volume de 250 mL de meio de cultura contendo 250 µL de canamicina 100 mg/mL e 250 µL de cloranfenicol 50 mg/mL e deixado sob agitação, de 220 rpm, a 37 °C até a densidade ótica atingir 0,6 (U.V. 660nm).
- Indução: Após atingir a densidade ótica necessária, é adicionado ao meio de cultura a concentração de 0,4 mM de IPTG e L-Ramnose, em diferentes concentrações, e deixado sob agitação, de 220 rpm, a 30 °C durante 5 horas.

Metodologia de Expressão em *E. coli* BL21 (DE3) plysS:

- Pré-Inóculo: 10 µL de células estocadas a -80 °C são adicionadas a 10 mL de meio de cultura contendo 10 µL de canamicina 100 mg/mL e colocado sob agitação, de 220 rpm, a 37 °C durante 14 a 16 horas.
- Inóculo: Todo o volume presente na etapa de pré-inóculo é adicionado à um volume de 250 mL de meio de cultura contendo 250 µL de canamicina 100 mg/mL e 250 µL de cloranfenicol 50 mg/mL e deixado sob agitação, de 220 rpm, a 37 °C até a densidade ótica atingir 0,6 (U.V. 660nm).
- Indução: Após atingir a densidade ótica necessária, é adicionado ao meio de cultura a concentração de 1mM de IPTG e deixado sob agitação, de 220 rpm, a 37 °C durante 4 horas.

Metodologia de Expressão em *E. coli* Arctic Express:

- Pré-Inóculo: 10 µL de células estocadas a -80 °C são adicionadas a 10 mL de meio de cultura contendo 10 µL de canamicina 100 mg/mL e 10 µL de gentamicina 100 mg/mL e colocado sob agitação, de 220 rpm, a 37 °C durante 14 a 16 horas.

- Inóculo: Todo o volume presente na etapa de pré-inóculo é adicionado à um volume de 250 mL de meio de cultura contendo 250 µL de canamicina 100 mg/mL e 250 µL de gentamicina 100 mg/mL e deixado sob agitação, de 220 rpm, a 30 °C até a densidade ótica atingir 0,6 (U.V. 660nm).

- Indução: Após atingir a densidade ótica necessária, é adicionado ao meio de cultura a concentração de 1 mM de IPTG e deixado sob agitação, de 220 rpm, a 12 °C durante 24 horas.

Verificação da Expressão: 1 mL do meio de cultura é retirado após a etapa de indução e centrifugado a 6000 rpm durante 8 minutos a 4 °C. O sobrenadante é descartado e as células concentradas são ressuspensas em 300 µL de tampão TES e deixado em repouso, a temperatura ambiente, durante 2 horas. 80 µL são retirados e adicionados a 20 µL de tampão de amostra 5X para ser aplicado no gel de acrilamida 12%.

Verificação da Solubilidade: Todo o volume do meio de cultura após a etapa de indução é centrifugado a 6000 rpm, durante 8 minutos a 4 °C para concentração das células. As células são ressuspensas em tampão de lise e levado ao sonicador em um ciclo de 10 pulsos de 30 segundos com intervalo de 60 segundos. O sobrenadante é centrifugado por 30 minutos a 14000 rpm a 12 °C, retirado 80 µL e adicionado a 20 µL de tampão de amostra 5X. O concentrado é ressuspensado em tampão TES e deixado em repouso durante 2 horas à temperatura ambiente. Em seguida, é retirado 80 µL e adicionado a 20 µL de tampão de amostra 5X.

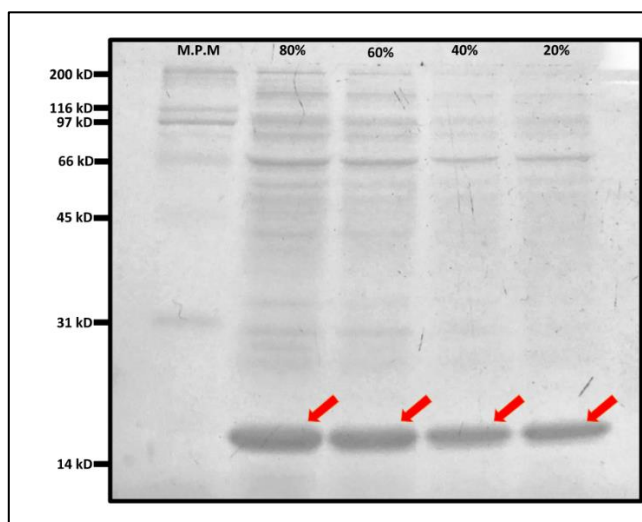
Purificação em coluna de afinidade: A purificação da proteína é feita a partir da utilização de uma coluna de afinidade de 2 mL HisTrap Ni²⁺. A primeira etapa é a equilibrção da coluna com a passagem de 40 mL de tampão de lise. A segunda etapa é a passagem da amostra. A terceira etapa é uma passagem de 20 mL do tampão de lise. A quarta etapa é a passagem do tampão de eluição.

Diálise: A amostra, após a purificação, é colocada em uma membrana com poros de 0,2 µm e deixada sob leve agitação durante 16 a 18 horas, a 18 °C em uma solução de 1 L contendo tampão fosfato. Em seguida, a amostra é centrifugada por 20 minutos, a 5000 rpm e a 12 °C e retirado 80 µL do sobrenadante e adicionado a 20 µL de tampão de amostra 5X. O concentrado é ressuspensado em 100 µL de tampão TES e deixado em repouso durante 2 horas para ser retirado 80 µL e adicionado a 20 µL de tampão de amostra 5X.

Diálise seriada: A amostra, após a purificação, é colocada em uma membrana com poros de 0,2 µm e deixada sob leve agitação durante 2 horas, a 18 °C em uma solução de 1 L contendo tampão fosfato e 8M de ureia. Em seguida, foi-se trocando a solução externa à membrana de forma gradativa (1 M) até atingir a concentração de 0 M, sendo deixado 1 hora por etapa.

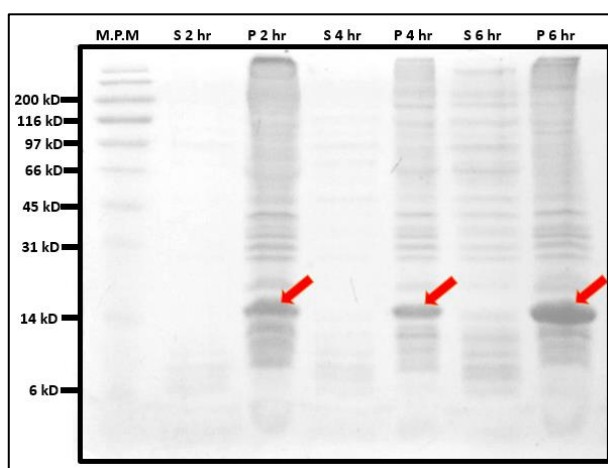
A **Figura A2** mostra que de fato há a expressão do plasmídeo contendo o domínio HIRAN da Rad5 em meio de cultura LB e células *E. coli CodonPlus* (DE3). A expressão também é observada após 2, 4 e 6 horas após a indução das células por IPTG (**Figura A3**) e em meio mínimo M9 (**Figura A4**).

Figura A2: Expressão do plasmídeo contendo o domínio HIRAN da Rad5 em meio de cultura LB utilizando células *E. coli CodonPlus* (DE3).



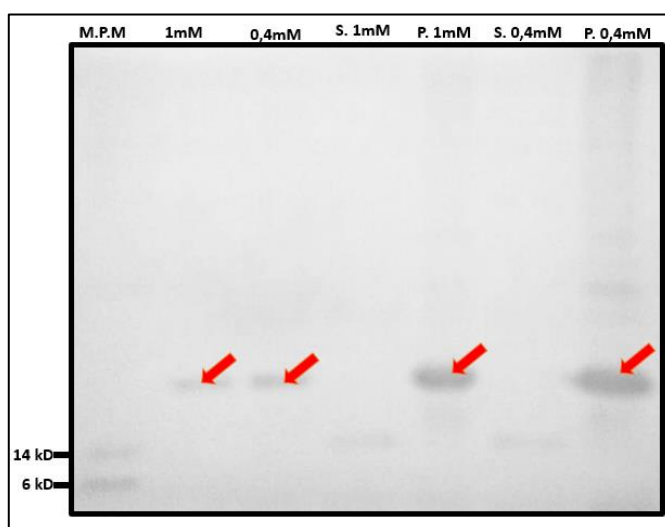
Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: MPM, Marcador de Peso Molecular; 80%, sendo 80 μ L de amostra e adição de 20 μ L de Tampão de Amostra 5X; 60%, sendo 20 μ L de amostra e adição de 20 μ L de Tampão de Amostra 5X e 20 μ L de tampão TES; 40%, sendo 20 μ L de amostra e adição de 20 μ L de Tampão de Amostra 5X e 40 μ L de tampão TES; 20%, sendo 20 μ L de amostra e adição de 20 μ L de Tampão de Amostra 5X e 60 μ L de tampão TES.

Figura A3: Expressão do plasmídeo em períodos de 2, 4 e 6 horas após indução por IPTG, em meio LB e células *E. coli CodonPlus* (DE3).



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: MPM, Marcador de Peso Molecular; Todas as letras S e P significam sobrenadante e *pellet*, respectivamente, da etapa de sonicação. O tempo, representado em algarismos e em horas (hr), é marcado após a indução por IPTG.

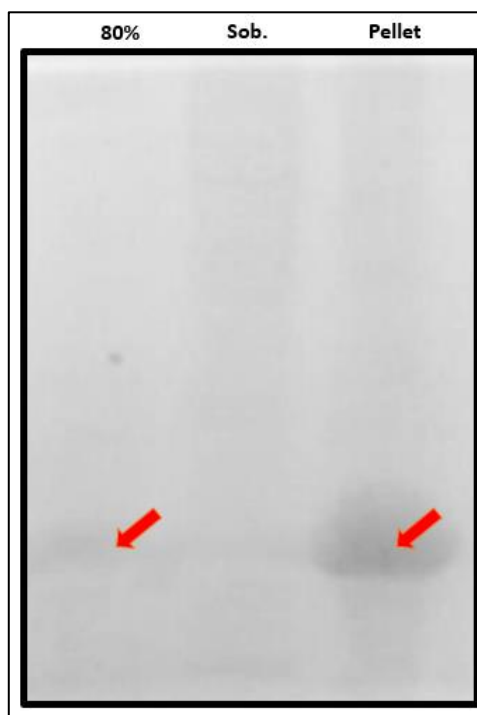
Figura A4: Expressão em meio mínimo M9 por 18 horas utilizando células *E. coli CodonPlus* (DE3)..



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: MPM, Marcador de Peso Molecular; A concentração de IPTG está representada em mM e as letras S e P representam sobrenadante e *pellet* da etapa de sonicação, respectivamente.

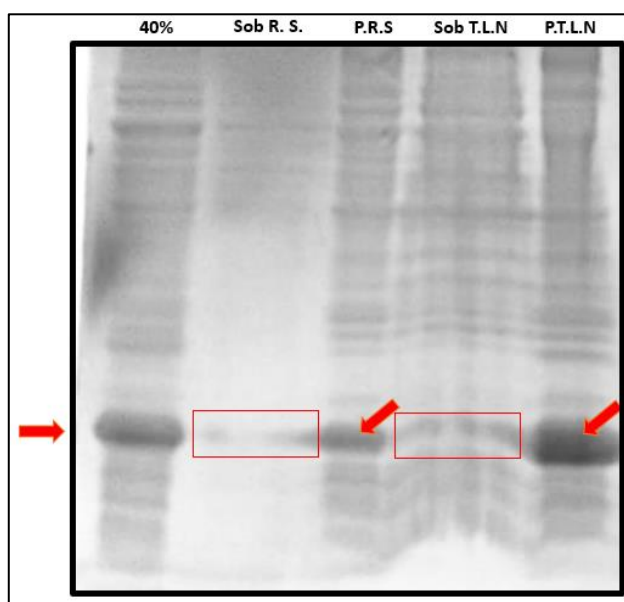
A etapa de verificação de solubilidade do domínio foi testada em tampão fosfato (**Figura A5**) e em tampão fosfato contendo Glicerol e Triton X100 (**Figura A6**). Em ambas soluções o domínio estava na fração insolúvel, indicando a formação de corpos de inclusão. A fim de verificar se a etapa de sonicação não havia sido suficiente, esta etapa foi realizada novamente para uma mesma amostra, porém o domínio continuava na fração insolúvel (**Figura A6**).

Figura A5: Etapa de verificação da solubilidade utilizando tampão fosfato.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; Sob. e Pellet representam o sobrenadante e o pellet da etapa de sonicação.

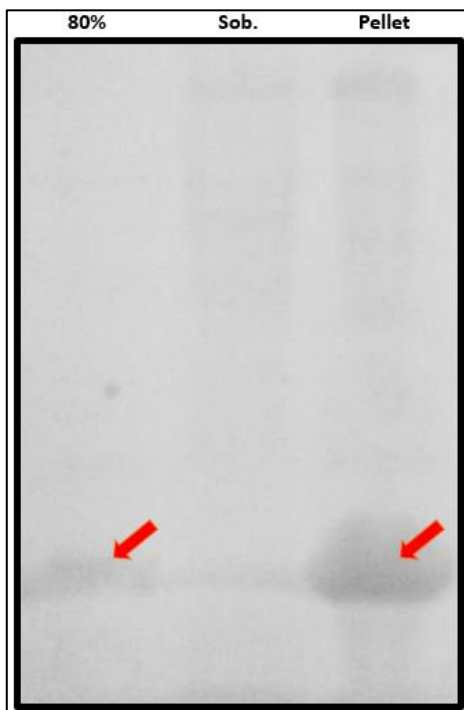
Figura A6: Etapa de verificação da solubilidade utilizando tampão fosfato contendo Glicerol e Triton X100 e amostra sonicada pela segunda vez.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. O retângulo vermelho está indicado a região da qual era esperado a expressão do plasmídeo. Legenda: 40%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; Sob R.S. e P.R.S representam o sobrenadante e o *pellet* da etapa de re-sonicação; Sob T.L.N e P.T.L.N representam o sobrenadante e o *pellet* da sonicação utilizando como tampão de lise o Tampão Fosfato com adição de Glicerol e Triton.

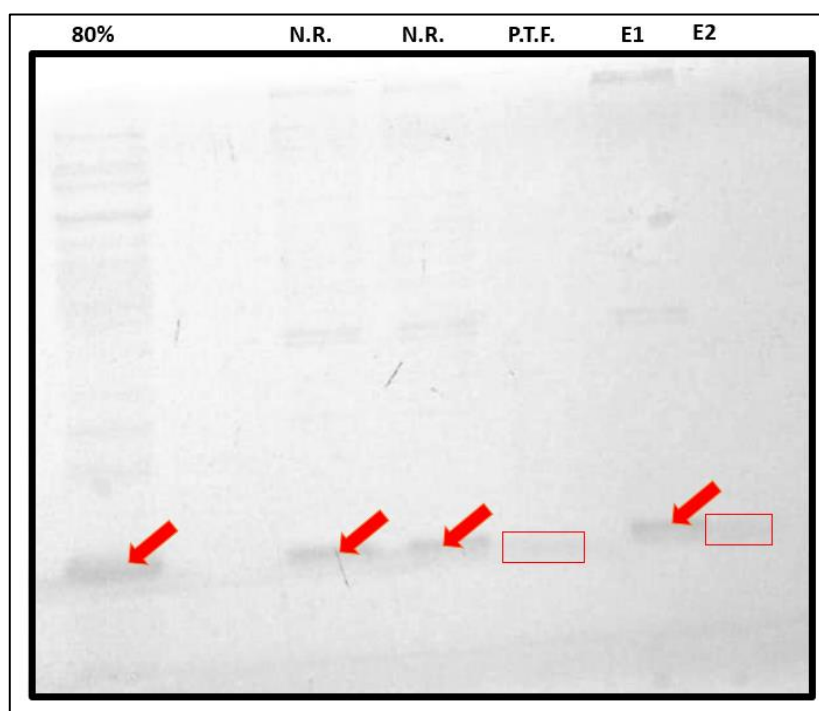
O uso do tampão fosfato com adição de 6M de Guanidina (**Figura A7**) e 8M de Ureia, na etapa de verificação de solubilidade, foi eficaz em colocar o domínio na fração solúvel somente com o incremento da ureia em meio LB e em células *E. coli CodonPlus* (DE3). A amostra foi purificada em coluna de afinidade e observado a presença no domínio na fração de eluição (**Figura A8**).

Figura A7: Solubilidade do domínio em tampão fosfato com adição de 6M de Guanidina.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; Sob. e Pellet representam o sobrenadante e o *pellet* da etapa de ressuspensão, após centrifugação, do *pellet* da sonicação.

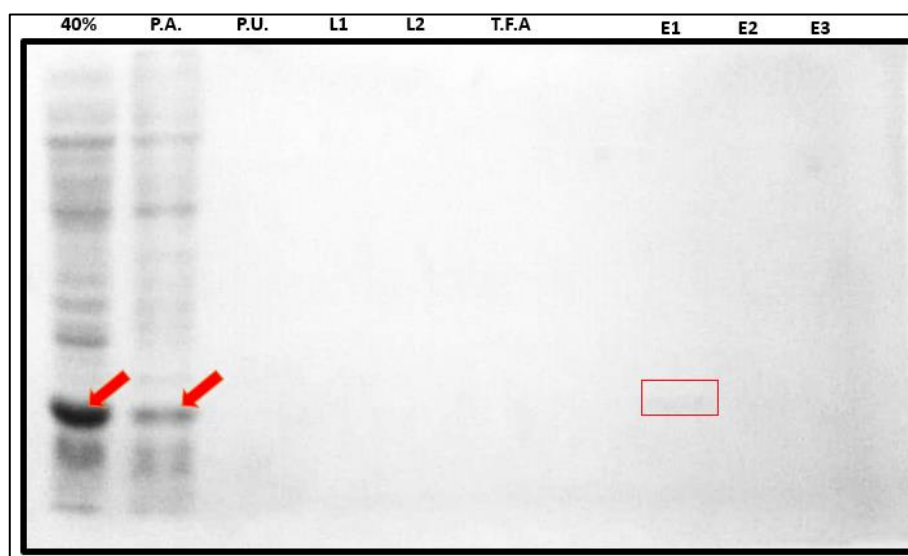
Figura A8: Purificação da amostra após a ressuspensão do pellet da centrifugação após etapa de sonicação com tampão fosfato com adição de 8M de Uréia.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. O retângulo vermelho está indicado a região da qual era esperado a expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; N.R. representa o que não foi retido na coluna após a passagem do sobrenadante da etapa de sonicação; P.T.F. representa a passagem de Tampão Fosfato com adição de 8M de Uréia para equilíbrio da coluna; E1 e E2 representam as eluições após a passagem de Tampão Fosfato com adição de Uréia e Imidazol.

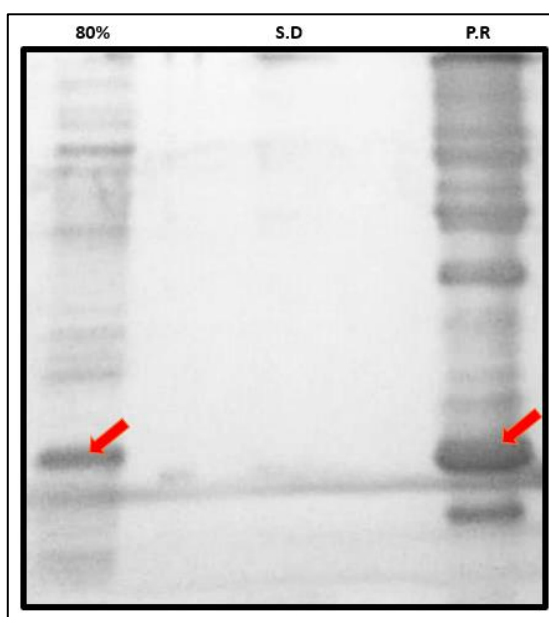
Como foi observado a presença do domínio na fração solúvel, as próximas etapas consistiram em tentar retirar o sal e a condição desnaturante das amostras. A **Figura A9** é a tentativa de reenovelamento na coluna de afinidade, aplicando concentrações gradativas de ureia 8M até 0M. A **Figura A10** é a diálise direta de uma concentração 8M de ureia para uma 0M, enquanto a **Figura A11** é a diálise seriada, com a troca do tampão com ureia em concentrações gradativas até 0M. Em todos os experimentos não foi possível recuperar o expresso em um tampão sem a condição desnaturante. Todos os experimentos foram feitos em meio LB e em células *E. coli CodonPlus* (DE3).

Figura A9: Tentativa de realizar o reenovelamento direto na coluna utilizando diluições na concentração de ureia a cada aplicação.



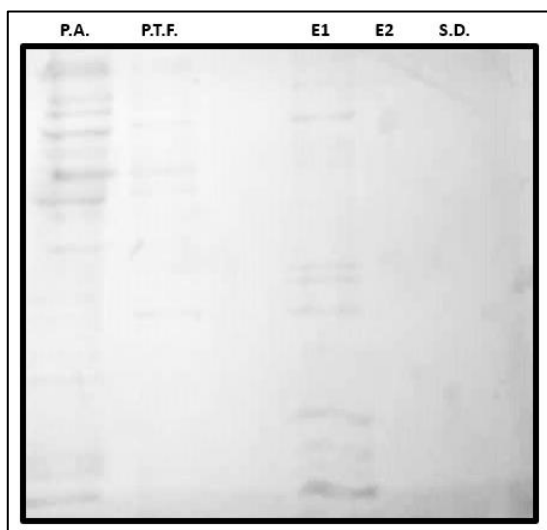
Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. O retângulo vermelho está indicado a região da qual era esperado a expressão do plasmídeo. Legenda: 40%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; P.A. representa o que não foi retido após a passagem da amostra pela coluna; P.U. representa a passagem de Tampão Fosfato com adição de 8M de ureia para equilíbrio da coluna; L1 e L2 representam a passagem de diferentes diluições na concentração da de ureia; T.F.A representa a passagem de Tampão Fosfato pela coluna; E1, E2 e E3 representam as eluição após a passagem de Tampão Fosfato com adição de Imidazol.

Figura A10: Diálise direta de uma concentração 8M de ureia para uma 0M.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; S.D representa o sobrenadante da centrifugação após a diálise; P.R. representa o *pellet* ressuspendido após a centrifugação na etapa de diálise.

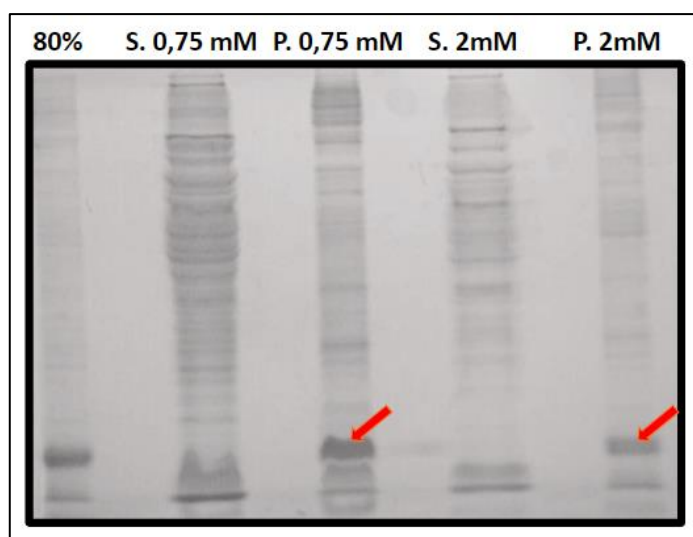
Figura A11: Diálise feita em diluições gradativas na concentração de ureia e a purificação da mesma.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. Legenda: P.A. representa o que não foi retido pela coluna após a passagem do sobrenadante da etapa de sonicação; P.T.F representa a passagem de Tampão Fosfato para equilíbrio da coluna; E1 e E2 representam a eluição e o que foi retido pela coluna após a passagem de Tampão Fosfato com adição de Imidazol; S.D. representa o sobrenadante da centrifugação após a etapa da diálise seriada.

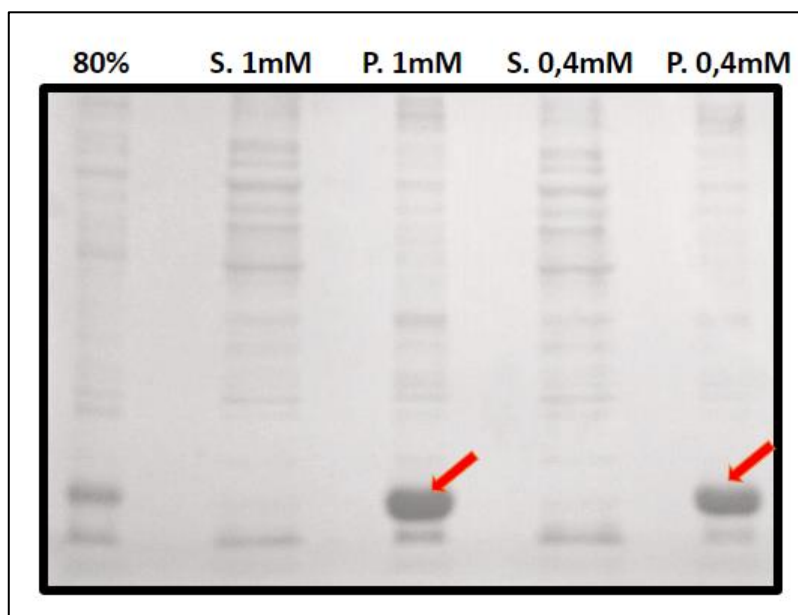
Por fim, os últimos experimentos feitos foram a expressão do plasmídeo em células *E. coli* Lemo 21 (DE3) (**Figura A12 e Figura A14**), *E. coli* BL21 (DE3) *plysS* (**Figura A13**) e *E. coli* Arctic Express (**Figura A14**). Em todas as células o domínio também estava presente na fração insolúvel.

Figura A12: Expressão do plasmídeo em células *E. coli* Lemo 21 (DE3).



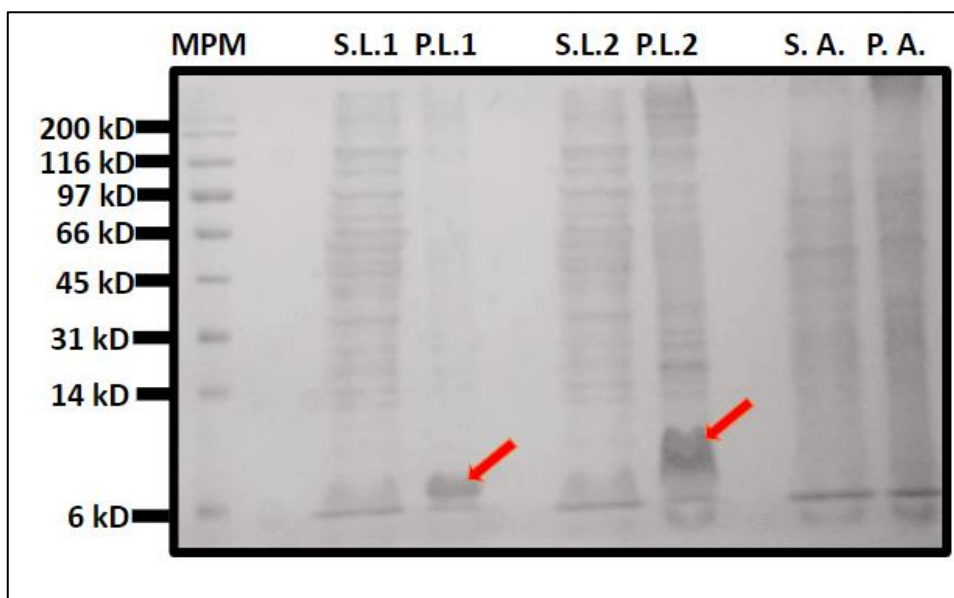
Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; A concentração de L-Ramnose está representada em mM e as letras S e P representam sobrenadante e *pellet* da etapa de sonicação, respectivamente.

Figura A13: Expressão do plasmídeo em células *E. coli* BL21 (DE3) *plysS*.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: 80%, amostra da Figura A2 utilizada como padrão e marcador de peso molecular; A concentração de IPTG está representada em mM e as letras S e P representam sobrenadante e *pellet* da etapa de sonicação, respectivamente.

Figura A14: Expressão do plasmídeo em células *E. coli* Lemo 21 (DE3) contendo 0,5 mM de L-Ramnose e em células *E. coli* Artic Express.



Todas as amostras foram aplicadas com 20 μ L em cada canaleta. As setas vermelhas indicam a região da proteína resultante da expressão do plasmídeo. Legenda: MPM representa o Marcador de Peso Molecular; S.L. e P.L. representam o sobrenadante e o *pellet* da etapa de centrifugação após a sonicação da expressão em células *E. coli* Lemo 21 (DE3), respectivamente, com adição de 0,5 mM de L-Ramnose e o algarismo representa o número da réplica; S.A. e P.A. representam o sobrenadante e o *pellet* da etapa de centrifugação após a sonicação da expressão em células *E. coli* Artic Express, respectivamente.

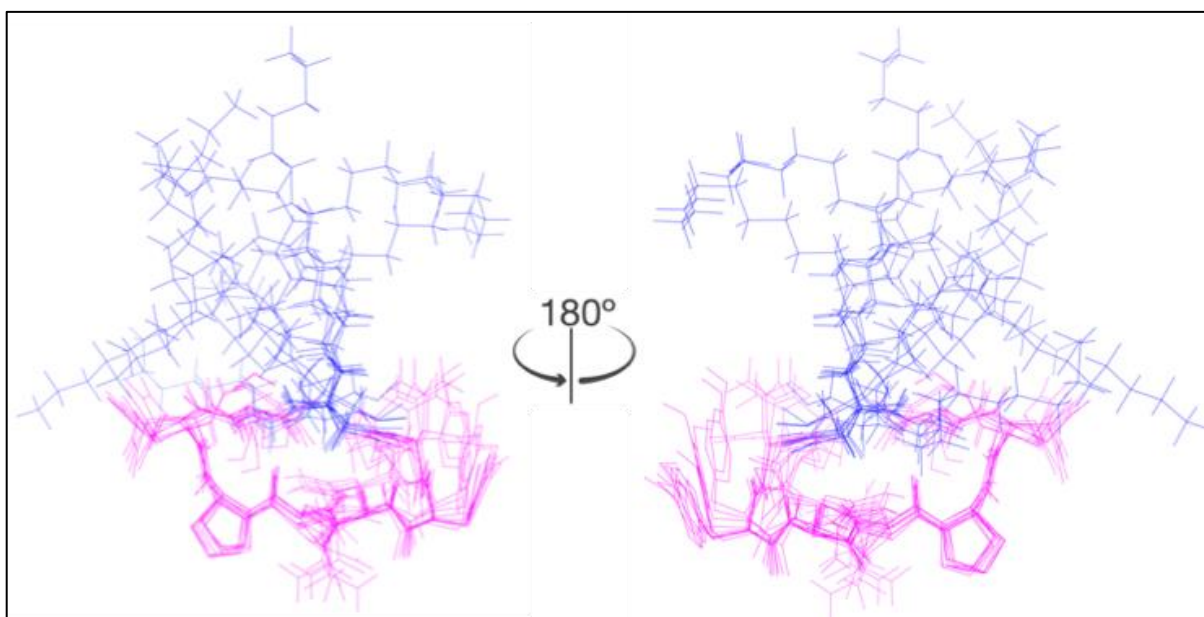
A partir destes resultados podemos concluir que o domínio HIRAN da Rad5 possui dificuldades em se manter na fração solúvel e ocorre sua agregação em forma de corpos de inclusão. Esses dados contribuem com a hipótese de que este domínio necessita de seus domínios adjacentes para se estabilizar estruturalmente.

APÊNDICE

Particpei na caracterização da estrutura primária e terciária de um lipopeptídeo com potencial biotecnológico denominado Iturina que resultou em um pedido de patente. Este trabalho foi realizado em colaboração com o professor Paulo Granjeiro da Universidade Federal de São João Del Rei e conta com Julia Rodrigues, Paula Torres, Diego Livio e Mariana de Magalhães como autores.

Os lipopeptídeos são pequenas moléculas produzidas por uma ampla variedade de microrganismos e podem funcionar como biossurfactantes, capazes de formar micelas e poros na membrana celular dos patógenos. Devido a essas propriedades, têm sido descritas como importantes moléculas bioativas com potencial antimicrobiano. Estas moléculas podem ser lineares ou circulares, com uma cabeça hidrofílica, em uma mistura de conformações D- e L- α -aminoácidos, e uma cauda de hidrocarboneto ligada a um β -aminoácido. Esse tipo de estrutura aumenta sua estabilidade e resistência às enzimas proteolíticas, aumentando o potencial da biotecnologia e aplicabilidade em substituição aos antibióticos.

Foi feita a caracterização do extrato purificado de *Bacillus sp* por espectrometria de massa MS e MS/MS e sequenciamento *de novo* de seus componentes, análises de bioinformática para a classificação deste lipopeptídeo no grupo das Iturinas e determinação da estrutura terciária dos peptídeos isolados por meio da espectroscopia de ressonância magnética nuclear. Para os cálculos estruturais, foi necessário a criação de diversos parâmetros de topologia e modificações nos arquivos dos programas CCPNmr, Aria e CNS, utilizados para calcular as estruturas a partir das restrições extraídas dos assinalamentos dos espectros de RMN.



Estrutura tridimensional resolvida por RMN do lipopeptídeo. Representação das 10 estruturas de menor energia. Em azul está representado o β -Aminoácido com a cauda lipídica ligada à sua cadeia lateral e em rosa estão representados os α -aminoácidos da porção peptídica