

**MONOGRAFIA**

**Curso de Ciências Biológicas - UFPR**

**Bacharelado – Ênfase em Processos Moleculares**

**Clonagem do gene hrq1 e expressão, purificação e cristalização da proteína Hrq1 de *Saccharomyces cerevisiae*, uma RecQ4 homóloga**

**Aluno:** Juliano Morimoto Borges

**Orientadoras:** Dra. Elaine Machado Benelli<sup>1</sup>

Dra. Silvia Onesti<sup>2</sup>

<sup>1</sup>Universidade Federal do Paraná

Departamento de Bioquímica e Biologia Molecular

<sup>2</sup>Structural Biology Laboratory

Elettra Sincrotrone Trieste S. p A

## AGRADECIMENTOS

Agradeço e peço desculpas ao mesmo tempo. Não pude citar nomes de cada um que fez parte da minha caminhada até agora. Mas certamente não esqueço nenhum dentro do meu coração, e meu agradecimento é eterno.

Ao CNPq, pela bolsa de estudos no exterior e oportunidade oferecida pelo programa Ciência sem Fronteiras.

Agradeço à Natureza e ao Universo.

À professora e amiga Elaine Machado Benelli, por confiar em mim e dar-me a oportunidade de estudar no exterior. Além de todas as nossas conversas de amigos no laboratório. Todo o incentivo e conselhos durante toda a minha vida acadêmica.

À professora Silvia Onesti, pela oportunidade e orientação dada ao longo deste ano. Além de uma orientadora, é uma amiga, sempre pronta para ajudar com os mais diversos assuntos. Pelo incentivo dado na minha nova estrada.

A todos os membros do laboratório de Biologia Estrutural, especialmente ao Aditya e Francesca Marino, por me ajudarem e orientarem. Nós somos a “RecQ Family”.

Ao NFN, pela ajuda com equipamentos e técnicas e a todo o departamento de Bioquímica da Universidade Federal do Paraná. A todos os professores do departamento e a Nancy, pela amizade e risos durante meu período no departamento.

À minha família, especialmente minha mãe, minha irmã Thatyane e meu cunhado e minha sobrinha/afilhada Alice. Ao meu pai, Tati, Ana Luiza e meus irmãos Vinicius e Felipe, pelo apoio e sacrifício durante esse tempo no exterior. Aos meus primos, que são como irmãos, meu muito obrigado. A todos os membros da minha família, vocês me dão o suporte necessário para seguir em frente.

A todos os meus amigos, em especial Gabriela Kinaske, Vinicius de Moraes, Lucas Moreira, Inaê Correa, Rodrigo Almeida e Prof. Luís Fernando Fávaro. Sem meus amigos eu não teria caráter nem face. A amizade é o segundo maior bem de um homem e eu posso dizer que sou rico por ter vocês.

Um agradecimento especial à Bruna Fernanda B. Barbosa pelo amor, carinho e companheirismo, e a toda sua família, pelo apoio, alegria e diversão que vocês me

proporcionam todos os dias. Fui acolhido como filho em sua casa e espero um dia retribuir tudo o que vocês fazem por mim.

À minha turma de Biologia da UFPR e a todos envolvidos no curso, em especial à Rô. Como disse durante o curso inteiro, você é o anjo da Biologia. Sem você eu não conseguiria metade do que eu consegui. Muito Obrigado.

A todos que me apoiaram durante minha caminhada. Tia Ivone, Ângela e Gisele Marchetti, vocês foram como parte da minha família e me ensinaram muito, não tenho palavras para agradecer o quanto vocês são importantes para mim.

Em geral, agradeço a todos que fizeram parte da minha vida. Cada etapa cumprida é em agradecimento a tudo o que vocês fazem por mim. Eu sou eternamente grato.

Muito Obrigado.

*Dedico esta monografia ao meu avô, Minoru Morimoto.*

*Que sua história continue me inspirando, mesmo que em memória, todos os dias.*

# ÍNDICE

<b>RESUMO</b> .....	<b>7</b>
<b>1. INTRODUÇÃO</b> .....	<b>8</b>
<b>2. REVISÃO BIBLIOGRÁFICA</b> .....	<b>12</b>
2.1 Síndrome de Werner (WS).....	17
2.2 Síndrome de Bloom (BS).....	17
2.3 Síndrome de Rothmund-Thomson (RTS).....	18
2.4 Características Gerais da família RecQ .....	19
2.5 Funções desempenhadas pelas proteínas da família RecQ.....	24
2.5.1. <i>Papel na replicação do DNA</i> .....	24
2.5.2. <i>Papel na recombinação homóloga (HR)</i> .....	25
2.5.3. <i>Papel no metabolismo e manutenção dos telômeros</i> .....	25
2.6 Hrq1: uma homóloga da proteína RecQ4 .....	26
<b>3. OBJETIVOS</b> .....	<b>29</b>
3.1 Objetivo Geral .....	29
3.2 Objetivo Específico .....	29
<b>4. MATERIAIS E MÉTODOS</b> .....	<b>30</b>
4.1 Análise estrutural <i>in silico</i> da proteína Hrq1 de <i>S. cerevisiae</i> .....	30
4.2 Clonagem do gene <i>hrq1</i> no vetor pET-SUMO/CAT .....	30
4.3 Purificação de plasmídeo em pequena escala (MiniPrep) .....	32
4.4 Transformação .....	33
4.5 Teste de expressão .....	34
4.6 Condições de auto-indução .....	34
4.7 Obtenção do extrato bruto de células .....	35
4.8 Purificação dos domínios $\alpha$ e $\omega$ da proteína Hrq1.....	35
4.9 Determinação das condições de cristalização do domínio $\omega$ .....	36
<b>5. RESULTADOS E DISCUSSÃO</b> .....	<b>37</b>
5.1 Análise estrutural <i>in silico</i> da proteína Hrq1 de <i>S. cerevisiae</i> .....	37
5.2 Clonagem do gene <i>hrq1</i> de <i>S.cerevisiae</i> .....	39
5.3 Otimização do protocolo de expressão .....	42
5.4 Purificação dos domínios $\alpha$ e $\omega$ .....	46
5.4.1. <i>Purificação do domínio <math>\alpha</math></i> .....	46

5.4.2. <i>Purificação do domínio <math>\omega</math></i> .....	49
5.5 Determinação das condições de cristalização do domínio $\omega$ .....	51
<b>6. CONCLUSÕES</b> .....	<b>54</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>55</b>
<b>8. APÊNDICE</b> .....	<b>58</b>

## RESUMO

Helicases são enzimas necessárias para a abertura da dupla fita de ácidos nucleicos para replicação e reparo. As helicases da família RecQ são consideradas as guardiãs do genoma (Hickson *et. al.* 2003). Mutações em três genes de proteínas RecQ em humanos estão associadas com síndromes que tem como característica geral predisposição a vários tipos de câncer. São elas a Síndrome de Bloom (envolvido com mutações na helicase BLM), Síndrome de Werner (mutações na helicase WRN) e Síndrome de Rothmund-Thomson (envolvido com mutações na helicase RecQ4). As helicases BLM e WRN são bem estudadas enquanto que RecQ4 e seu papel na síndrome de Rothmund-Thompson permanece desconhecido. Em seres humanos há outra duas proteínas da família RecQ, RecQ5 e RecQ1, que não estão associadas a síndromes. *Saccharomyces cerevisiae* era conhecido por possuir apenas uma proteína da família RecQ, a helicase Sgs1. Entretanto, Barea *et. al.* (2008) propuseram a existência de uma segunda proteína da família RecQ, chamada Hrq1, que seria ortóloga à proteína RecQ4 em humanos. Tendo em vista as dificuldades de trabalho encontradas com a proteína RecQ4 pela baixa solubilidade e alto nível de proteólise, a proteína Hrq1 apresenta-se como uma boa alternativa para a elucidação das funções de RecQ4 na síndrome de Rothmund-Thompson. Portanto, este trabalho tem como objetivo estabelecer um protocolo eficiente de clonagem, expressão, purificação e cristalização da proteína Hrq1 de *S.cerevisiae* para estudos estruturais. Através de análise *in silico* foi possível determinar os domínios de clonagem que foram realizados utilizando a técnica de *RF-cloning*. O protocolo de autoindução foi utilizado em duas estirpes de *E.coli* (BL21Star e Rosetta-2) para a expressão das construções. A purificação foi centrada nos domínios Cdt-1 e DUF da proteína Hrq1. O *screening* das condições de cristalização foi realizado apenas para o domínio DUF e alguns cristais foram obtidos embora a otimização das condições de cristalização e a coleta dos dados são as próximas etapas deste estudo.

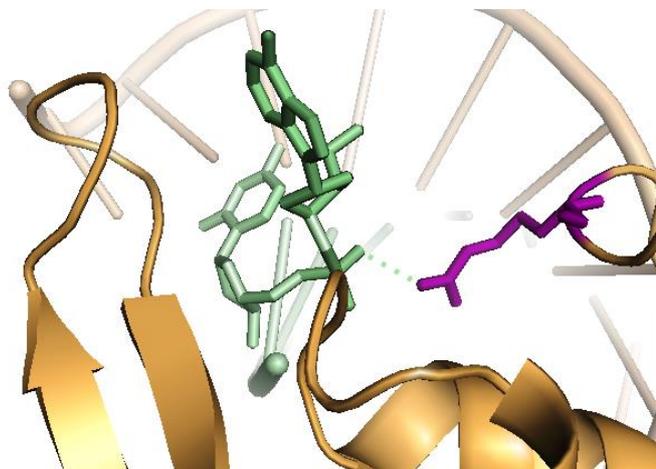
## 1. INTRODUÇÃO

Recentemente, o papel de proteínas envolvidas na replicação e reparo do DNA tem sido exhaustivamente investigado. Isso se deve ao fato de que algumas síndromes são resultado de mutações em genes codificantes para enzimas envolvidas nessas vias. A grande maioria dessas síndromes é responsável por uma elevada predisposição aos mais diversos tipos de câncer. Câncer é uma doença que acomete grande parte da população mundial e que pouco se conhece sobre aspectos multifatoriais que o desencadeiam e como tratá-los. Diferentes tipos de câncer possuem diferentes características morfológicas, bioquímicas, estruturais, o que torna ainda mais difícil a padronização dos mecanismos responsáveis pela doença.

Mutações em genes de enzimas envolvidas no metabolismo de ácidos nucleicos, tais como helicases, translocases, polimerases, telomerasas, são encontrados em síndromes ou doenças com elevada incidência de câncer.

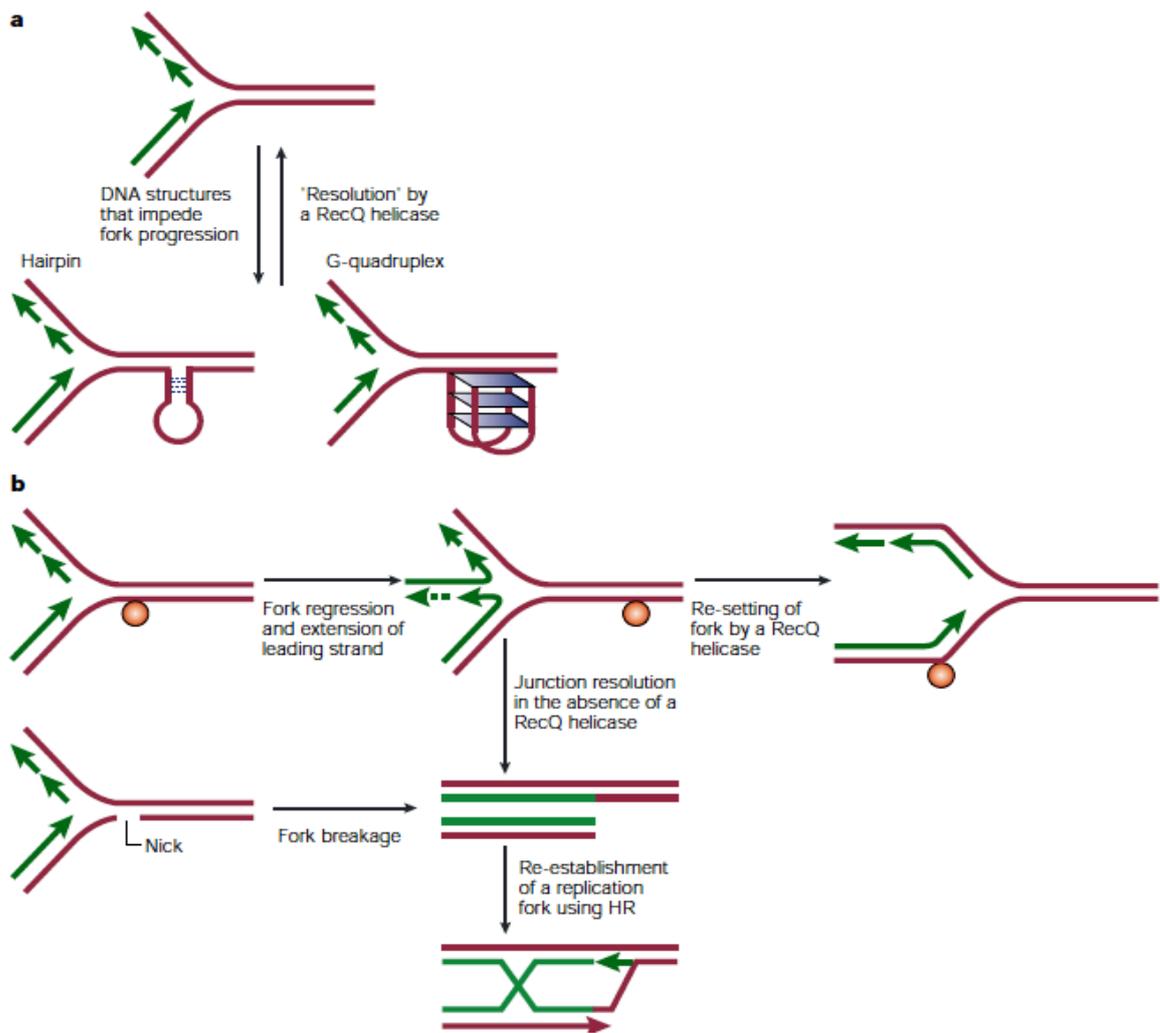
Helicases são enzimas necessárias para a abertura da dupla fita de ácidos nucleicos (Figura 01) e estão envolvidas tanto em vias de reparo como na própria replicação do DNA. Já as translocases são se ligam e movem-se com polaridade característica, ao longo das fitas de DNA. São reconhecidas seis super-famílias de helicases/translocases (Singleton *et. al.* 2007) e cada uma apresenta características próprias e as maiores super-famílias são a superfamília-1 (SF1) e a superfamília-2 (SF2). Geralmente, é difícil estabelecer a diferença entre a atividade de uma proteína helicase de uma translocase restrita, uma vez que estudos estruturais devem ser realizados para verificar se há, de fato, uma atividade de abertura da dupla fita ou simplesmente um “caminhar” sobre ela. Além disso, helicases são translocases capazes de abrir a dupla fita de DNA durante sua translocação. As enzimas da subfamília RecQ pertencem a

superfamília SF2, a maior superfamília dentre as helicases e translocases. (Singleton *et. al.* 2007).



**Figura 01** – WRN helicase ligado ao DNA (verde). A arginina conservada (roxo) interage com o fosfato do esqueleto do DNA que é aberto. *PDB code: 3AAF*

A família das RecQ-helicases é responsável pela manutenção da integridade do genoma, atuando diretamente na via de Recombinação Homóloga (HR) (Figura 02). Essas proteínas são responsáveis por evitarem recombinações ao acaso. (Gangloff *et. al.*, 1994 e Watt *et. al.*, 1995). Portanto, mutações em três RecQ-helicases de humanos, BLM, WRN e RecQ4 (ou RecQL4), são responsáveis por três diferentes síndromes, Síndrome de Bloom, Síndrome de Werner e Síndrome de Rothmund-Thompson, respectivamente. (Ellis *et. al.* 1995; Imbert *et. al.* 1996; Yu *et. al.* 1996; Kitao *et. al.*, 1999). Todas essas síndromes estão relacionadas com a elevada predisposição a cânceres. (revisado por Monnat Jr., 2010).



**Figura 02** – Vias de atuação das proteínas da família RecQ. Fonte: Hickson (2003)

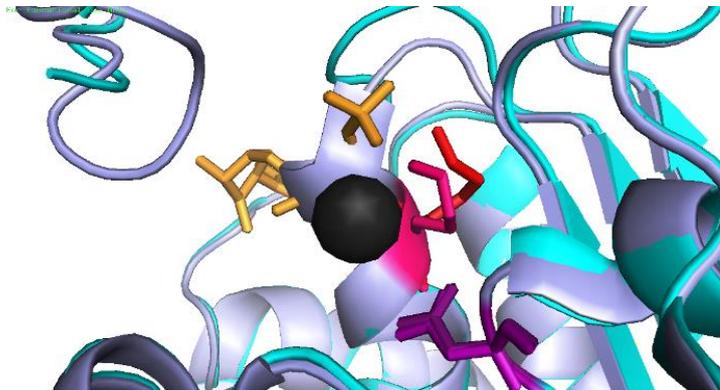
As enzimas WRN e BLM são as mais estudadas e por esse motivo suas características bioquímicas e estruturais são mais bem compreendidas. Por outro lado, pouco se conhece a respeito da proteína RecQ4 e de enzimas homólogas a esta em outros organismos. A proteína RecQ4 é de difícil obtenção tanto pela insolubilidade quanto pela susceptibilidade à proteólise. (Marino, F. *comunicação pessoal*). Por esse motivo, há necessidade de estudos em outros organismos para contribuir com a elucidação das funções bioquímicas e características estruturais da proteína RecQ4.

Através de análises bioinformáticas, Barea *et al* (2008) propôs que Hrq1 em *Saccharomyces cerevisiae* atuasse como ortólogo de RecQ4 e Grocock *et. al.* (2012) caracterizaram bioquimicamente as funções da proteína Hrq1 em *Schizosaccharomyces pombe* sugerindo que essa proteína é funcionalmente semelhante à RecQ4. Estes achados abriram novos caminhos para o entendimento da função dessas enzimas na síndrome de Rothmund-Thompson (Kitao *et. al.*,1999). Este estudo visa contribuir obtenção de informações estruturais da proteína Hqr1 estabelecendo protocolos de expressão, purificação e cristalização de domínios funcionais desta proteína. Esses resultados possibilitarão o estudo dos aspectos bioquímicos e estruturais da proteína Hrq1, auxiliando a elucidar os papéis da proteína RecQ4 em humanos e seu envolvimento no desenvolvimento das síndromes.

## 2. REVISÃO BIBLIOGRÁFICA

Helicases são enzimas que participam em diferentes fases do metabolismo de ácidos nucleicos em diferentes aspectos. Essas enzimas operam quando RNA:RNA, DNA:DNA ou DNA:RNA precisam ser abertos para que ocorra expressão gênica, recombinação e replicação. Entretanto, essa atividade de abertura somente é permitida mediante consumo de nucleotídeo tri-fosfatado, ou NTP. Por esse motivo, helicases são *NTP-binding proteins*, e é a hidrólise de NTPs que fornece a energia necessária para que a proteína desempenhe sua função. (Matson *et. al.* 1994) (Singleton & Wigley 2002) (von Hippel 2004). Translocases, por sua vez, são enzimas com capacidade de se ligarem à ácidos nucleicos, porém, sem atividade de helicase, isto é, não desemparelham as fitas de DNA. Translocases acoplam a energia proveniente da hidrólise de NTPs para o movimento *ao longo* da molécula de DNA.

As helicases são classificadas em 6 superfamílias (SF1, SF2, SF3, SF4, SF5 e SF6) sendo que as maiores são SF1 e SF2. Apesar da alta diversidade e variabilidade estrutural, são reconhecidas “assinaturas” que caracterizam proteínas dessas superfamílias. Essas “assinaturas” são responsáveis pela formação de motivos chamados “*RecA-like folds*” devido a semelhança estrutural à enzima RecA de *E. coli*. Esses motivos *in tandem* compõe um núcleo que é responsável pela atividade de hidrólise de NTP. (Ye 2004). As características dessa região são: 1) resíduos conservados que são responsáveis pela ligação e hidrólise de NTP, semelhante aos domínios Walker A e Walker B de ATPases. (Walker *et. al.* 1982) (Figura 2 e 3) um domínio chamado de “*dedo de arginina*”. (Scheffzek *et.al.* 1997). Em geral, proteínas SF1 e SF2 possuem dois motivos *RecA-like* por cadeia polipeptídica. (Singleton *et. al.* 2007).



**Figura 03** – RecQ de *E. coli* (PDB code: 1OYW). Sobreposição dos motivos Walker A e Walker B na estrutura não ligada a ATP (cinza) e ligada a ATP (azul). É possível verificar a mudança conformacional da lisina quando não há interação com ATP (rosa) e com interação com ATP (vermelho). O aspartato conservado no motivo Walker B praticamente não se altera devido à interação com ATP. Molécula de ATP é representada na figura pela cor amarela.

As Helicases/Translocases podem ainda serem classificadas em relação a sua polaridade. O DNA é um polímero polarizado e suas fitas anti-paralelas conferem a possibilidade de dois possíveis sentidos de polarização à proteína: 5'->3' ou 3'->5'. Proteínas com polaridade 5'->3' são consideradas de **tipo B**, enquanto que proteínas com polaridade 3'->5', **tipo A**. As superfamílias SF1 e SF2 possuem representantes em ambas as categorias, contudo, isso é uma exceção. Proteínas da família SF3, por exemplo, são todas pertencentes ao tipo A (Monnart Jr. 2010). Além disso, ainda classifica-se helicases/translocases podem ser classificadas quanto à atividade desempenhada no DNA. A enzima pode ligar-se ao DNA de duas maneiras: em uma das fitas ou em ambas. Enzimas que ligam e desempenham sua atividade catalítica em uma única fita do DNA são chamadas de **tipo  $\alpha$** . Por outro lado, enzimas que reconhecem,

ligam-se e atuam em ambas as fitas do DNA são chamadas de **tipo  $\beta$** . (Singleton *et. al.* 2007). Portanto, SF2A $\alpha$  uma helicase hipotética é membro da superfamília (SF2), apresenta polaridade 3'->5' e seu substrato é DNA simples fita.

Além do núcleo, as helicases/translocases possuem domínios acessórios localizados na porção N-terminal tanto quanto na porção C-terminal. Esses domínios acessórios possuem diferentes papéis, atuando desde reconhecimento e afinidade por diferentes substratos, interação proteína:proteína e modulação da atividade do núcleo. (Singleton & Wigley, 2002).

A superfamília-2 (SF2) é a maior dentre as helicases/translocases e, portanto, apresenta diversas funções celulares e suas subfamílias são alvos de intensos estudos. Nesta superfamília encontram-se as subfamílias das DEAD(H)box RNA-helicases, RecQ-helicases e enzimas Sfn2. Essas proteínas possuem atividade de translocase/helicase pela hidrólise de ATP, entretanto, o substrato dessas proteínas (dupla fita ou simples fita de DNA) ainda é desconhecido. Por esse motivo, não se pode classificar todos os membros dessa família seguindo a classificação descrita acima. O movimento dessas enzimas é dado pela hidrólise de ATP, o qual se liga ao motivo I, na porção N-terminal, induzindo mudanças conformacionais as quais determinam a enzima no estado "fechado". Além disso, a região denominada "dedo de Arginina" no motivo VI interage com os nucleotídeos. A hidrólise do ATP induz novamente mudanças conformacionais que permitem a abertura da dupla fita e/ou o deslocamento ao longo delas. Aparentemente, muitas das enzimas da família SF2 possuem polaridade 3'->5'. É esse o caso das proteínas pertencentes a família das RecQ-helicases. (Singleton *et. al.* 2007).

RecQ helicases são essenciais para o metabolismo do DNA atuando de diferentes maneiras.. As RecQ helicases são capazes de atuarem em estruturas de DNA não canônicos (formas diferentes da forma B-DNA) tais como *D-loops*, tripla- e quadrupla-

hélices, *Holliday-Junctions* entre outros (Frei & Gasser 2000; Bachrati & Hickson, 2008; Bohr, 2008; Chu & Hickson 2009; Hickson, 2003). Além disso, é provável que RecQ helicases atuem na forquilha de replicação (Opresko *et. al.* 2004). Devido a diversidade funcional relacionada a essas helicases, diferentes tipo de doenças tem sido extensivamente relacionadas a algumas mutações em genes codificantes de proteínas da família das RecQ. Humanos possuem cinco tipos diferentes de enzimas da família RecQ e mutações em três desses cinco genes estão associados ao fenótipo de doenças autossômicas recessivas caracterizadas por instabilidade genômica e elevada predisposição ao câncer. Mutações em WRN, BLM e RecQ4 (ou RecQL4) são correlacionadas com Síndrome de Werner, Síndrome de Bloom e a Síndrome de Rothmund-Thompson, respectivamente. As demais proteínas da família RecQ (RecQ1 e RecQ5) não são relacionadas a nenhuma doença. (Monnat Jr. 2010).

### 2.1 Síndrome de Werner(WS)

A síndrome de Werner é provocada por mutações no gene codificante para a helicase WRN. As principais características associadas a essa síndrome são baixa estatura, envelhecimento precoce (que pode ser observado pelo fato dos cabelos grisalhos e calvice precoce), escleroderma e cataratas em ambos os olhos. Portadores dessa síndrome possuem um elevado risco de câncer. Entretanto, somente alguns tipos específicos de câncer são mais propícios tais como sarcomas, melanomas e osteosarcomas. (Goto *et. al.* 1996; Monnat Jr. 2001; Monnat Jr. 2002). Além de câncer, pacientes ainda podem desenvolver arterosclerose e doenças cardíacas sendo essas doenças cardiovasculares precoces as principais causas de morte em portadores dessa síndrome. (Monnat Jr. 2010). Pacientes portadores da síndrome de Werner possuem comprometimentos na via de recombinação homóloga. (Prince *et. al.* 2001; Saintigny *et. al.*

2002). O comprometimento dessa via pode ser responsável pelo alto índice de instabilidade cromossômica observado nas células desses pacientes. (Monnart Jr. 2010). Não obstante, células desses pacientes apresentam baixa taxa de crescimento em cultura, elevada taxa de morte celular e envelhecimento celular precoce. Ao contrário dos pacientes portadores da síndrome de Bloom, as células não apresentam elevadas taxas de troca de material entre cromátides irmãs, entretanto, acumulam deleções e translocações cromossômicas. (Salk *et. al.* 1985).

## 2.2 Síndrome de Bloom (BS)

A síndrome de Bloom é provocada por mutações no gene codificante para a helicase BLM. As principais características dos portadores dessa síndrome são baixa estatura congênita, retardo no crescimento pós-natal e manchas avermelhadas no nariz e nas pálpebras. Essa mancha forma-se nos primeiros meses de vida e provavelmente é devido a exposição ao sol. Além disso, é observada uma diminuição da resposta imune celular e humoral e essa depleção do sistema imune tem sido associada ao elevado risco de otites e pneumonia que acomete os portadores dessa síndrome. Assim, portador da síndrome de Bloom tem elevados riscos de câncer, entretanto, diferentemente dos portadores da síndrome de Werner, pacientes portadores dessa síndrome possuem susceptibilidade a uma grande variedade de cânceres.

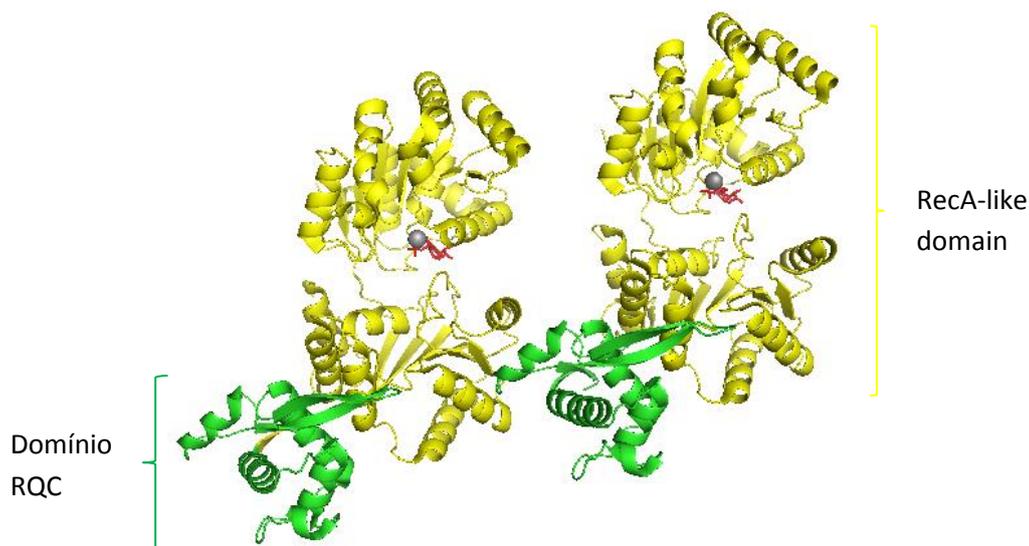
Os portadores da síndrome de Bloom possuem uma elevada taxa de trocas de material entre cromátides irmãs e isso foi atribuído a distúrbios na via de recombinação homóloga. (German, 1993). Assim como pacientes portadores da síndrome de Werner, células de paciente portadores da síndrome de Bloom apresentam crescimento lento em cultura, alta taxa de morte celular e senescência precoce. Entretanto, essas células possuem um retardo ainda maior no crescimento quando comparado às células da síndrome de Werner. (Sharman *et. al.* 2007).

### 2.3 Síndrome de Rothmund-Thompson (RTS)

Essa síndrome é provocada por mutações associadas ao gene codificante da helicase RECQ4. Gene *rec4* está localizado no braço longo do cromossomo 8 (8q). Pacientes portadores da síndrome de Rothmund-Thompson apresentam trissomia do cromossomo 8 ou isocromossomo 8q. (Grant *et. al.* 2000; Maire *et. al.* 2009). Portadores dessa síndrome possuem mudanças características na pele devido a susceptibilidade ao sol. Essa sensibilidade promove o aparecimento de manchas com características avermelhadas e inchadas na face que pode espalhar-se para as bochechas, peito, costas e abdômen. Essas manchas tornam-se diferentemente pigmentadas ao longo do tempo. (Monnat Jr. 2010). Outras características são a ausência de cabelos, cílios e supercílios, anormalidades nos dentes e ossos, elevado risco de osteosarcomas e cataratas, muito embora cataratas acometam somente uma pequena porção dos portadores dessa síndrome. (Wang *et. al.* 2001; Siitonen *et. al.* 2009). Outras duas disfunções são relacionadas a síndrome de Rothmund-Thompson: a síndrome RAPALADINO e a síndrome de Baller-Gerold. (Monnat Jr. 2010). As células de pacientes portadores dessa síndrome são altamente sensíveis a radiação ionizante. (Kerr *et. al.* 1996).

### 2.4 Características gerais da família RecQ

As proteínas da família RecQ possuem um domínio chamado de “domínio helicase” formado por dois domínios “*RecA-like domains*” cada um contendo aproximadamente 200 aminoácidos. Esse domínio é chamado de “*core das helicases*” e é responsável pelo movimento ao longo da molécula de ácido nucléico (Figura 04). Esse movimento é dependente de hidrólise de ATP como fonte de energia. Entretanto, a atividade de helicase *per se* necessita de estruturas adicionais. (Vindigni *et. al.* 2010).



```

MASVSALTEELDSITSELHAVEIQIQELTERQQELIQKKVLTKKIKQCLESDAGASNEYDSSPAAWNK
EDFPWSGKVKDILQNVFKLEKFRPLQLETINVTMAGKEVFLVMPTGGGKSLCYQLPALCSDGFTLVICPL
ISIMEDQLMVLKQLGISATMLNASSSKEHVKVVHAEMVNKNSLKLIVTPEKIAKSKMFMRSLEKAYEA
RRFTRIAVDEVHCCSQWGHDFRPDYKALGILKRQFPNASLIGLTATATNHVLTDAQKILCIEKCFFFTAS
FNRPNLYEVRQKPSNTEDFIEDIVKLINGRYKQSGI IYCFSQKDSEQVTVSLQNLGIHAGAYHANLEP
EDKTTVHRKWSANEIOVVVATVAFGMGIDKPDVRFV IHHSMKSMENYYOESGRAGRDMKADCILYYGF
GDI FRIS SMVVMENVGQOKLYEMVSYQNISKRRVIM AQHFDEVWNSEACNKMCNCCCKDSAFERKNIT
EYCRDLIKILKQAEELNEKLTPLKLIDSWMGKGAAKLRVAGVVAPTLPREDLEKIIAHFLIQQYLKEDYS
FTAYATISYLKIGPKANLLNNEAHAITMQVTKSTQNSFRAESSQCHSEQGDKKMEEKNSGNFQKKAANM
LQSGSKNTGAKKRKIDDA

```

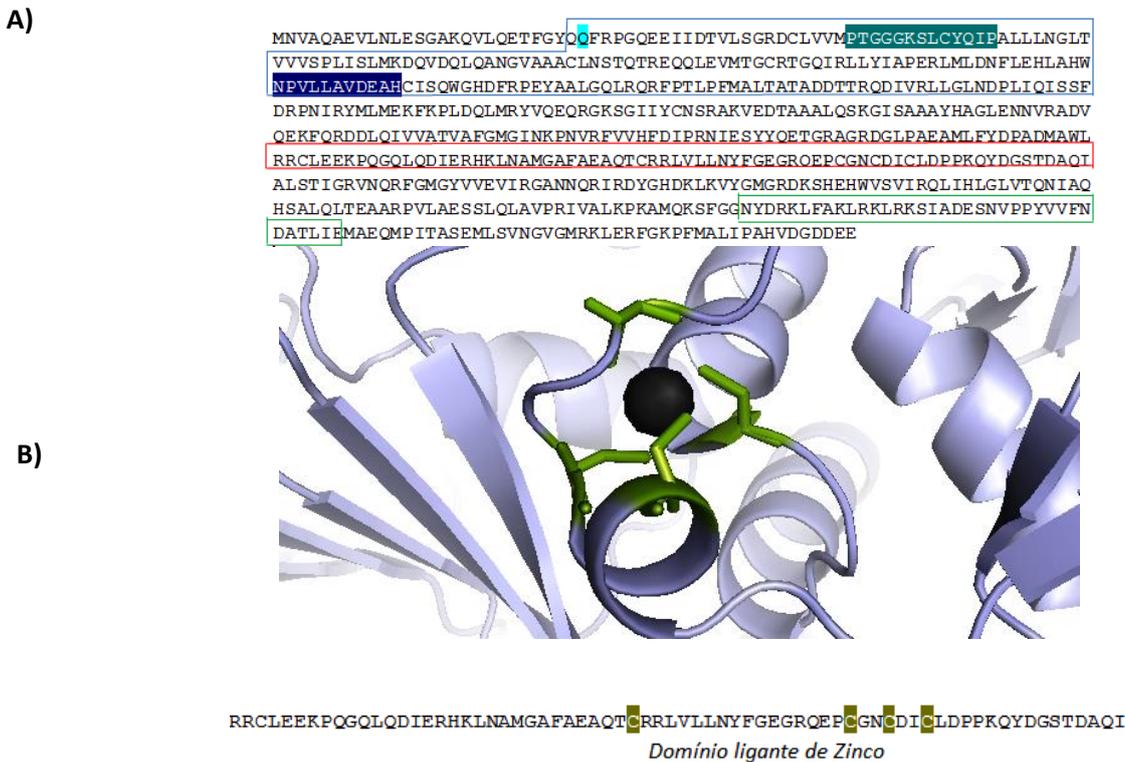
**Figura 04** – Sequência de aminoácidos e estrutura cristalográfica da proteína RecQ1 de humano. (Pike *et. al.* 2009). As proteínas da família RecQ são compostas dois domínios *RecA-like* (em amarelo), que são *ATP-binding* motifs. ATP é mostrado em vermelho. Átomo de Manganésio ( $Mg^{2+}$ ) é mostrado como uma esfera cinza. Em verde o domínio RQC, característico das proteínas RecQ. PDB code: 2V1X.

Entre os domínios *RecA-like domains* há sete motivos de helicases, denominados como “helicase motifs”. Por meio de análises sobre outras helicases de outras famílias, esses motivos são necessários para a ligação e hidrólise de nucleotídeos além da ligação a DNA simples-fita. (von Hippel, 2004; Lohman *et. al.* 2008; Singleton *et. al.* 2007). O domínio helicase da proteína RecQ de *E. coli* foi determinado a resolução de 1.8 Å por difração de raios-X. A proteína foi cristalizada em duas condições: 1) ligada a

nucleotídeos e 2) não ligada a nucleotídeos. Marino *et. al.* (2013) sugeriram que há uma pequena rotação entre os dois domínios *RecA-like* e é esse movimento que permite a ligação de nucleotídeos. Outro estudo realizado com a proteína RecQ1 humana mostrou que esses domínios possuem uma alta flexibilidade e que esse movimento é responsável pelo deslocamento da proteína ao longo da fita de DNA, movimento este dirigido pela hidrólise de ATP. (Velankar *et. al.* 1999; Lee & Yang, 2006).

RecQ helicases possuem ainda um motivo chamado de “motivo 0”, localizado na porção N-terminal do núcleo das helicases. Esse domínio é fundamental na função de helicase dessas proteínas e estudos mostram que mutantes para essa região possuem função de abertura da dupla fita de DNA comprometida. (Bahr *et. al.* 1998; Garcia *et. al.* 2004).

As proteínas da família RecQ possuem uma região contendo um domínio ligante de Zinco, ou “Zinc domain” (Figura 05) e um domínio hélice-volta-hélice chamado de “Winged-helix domains” na porção C-terminal ao núcleo helicase. A região contendo essas estruturas é chamada de “RQC domain” (RecQ-C-terminal). O domínio RQC tem um importante papel na estabilidade da proteína RecQ (Liu *et. al.* 2004; Guo *et. al.* 2005) e pode estar envolvido na interação proteína:DNA. Análises estruturais feitas por Bernstein *et al.* (2003) sugere que a dupla fita de DNA localiza-se entre Winged-helix domain e o Zinc-domain, porém, não se pode afirmar o quanto dessa interação é provocada no processo de formação do cristal ou se ela ocorre *in vivo*. (Vindigni *et. al.* 2010; Bernstein *et al.* 2003). A confirmação da interação depende de análises funcionais destas proteínas, tais como ensaios de atividade de helicase por meio de substituição de aminoácidos.

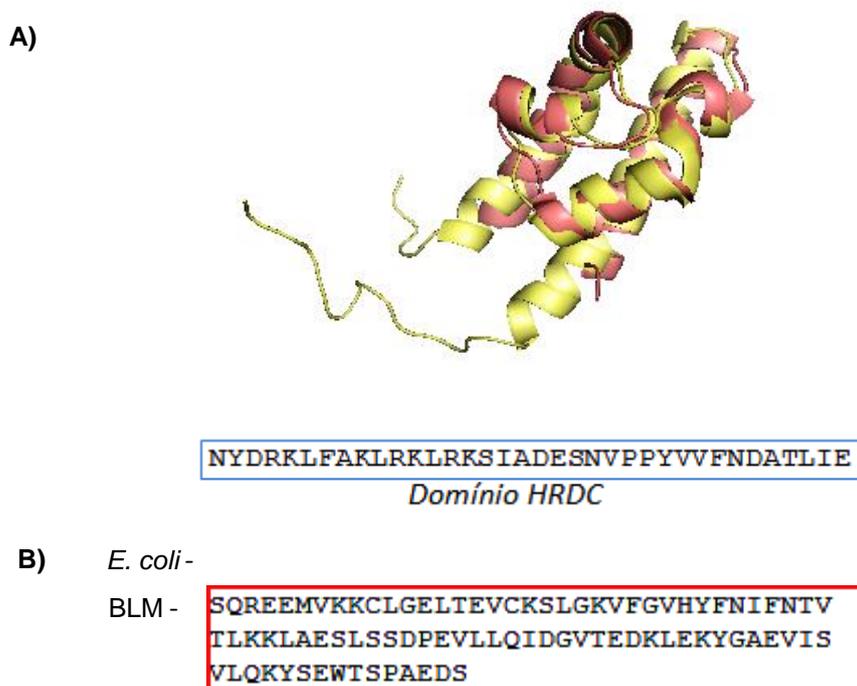


**Figura 05- A)** Estrutura da proteína RecQ de *E. coli* mostrando os principais domínios da proteína. Box azul – Domínio Helicase/DEAH-box; Regiões evidenciadas: azul claro – Motivo “0”. Verde petróleo – *P-loop* (Walker A); Azul marinho – Walker B. Box vermelho – domínio RQC contendo região ligadora de Zinco. Box verde – Domínio HRDC. **B)** Domínio RQC que é um domínio acessório com capacidade de ligação ao Zinco. As cisteínas coordenando o átomo de Zinco (esfera cinza) foram marcadas em verde. A sequência de aminoácidos é apresentada abaixo com as cisteínas conservadas marcadas em verde. (PDB code: 1OYW)

Marino *et. al.*, (2013) propôs que a proteína RecQ4 também possui o domínio RQC. Através de análises com ferramentas bioinformáticas mostraram que RecQ4 possui tanto o domínio Winged-helix tanto quando o domínio ligante de Zinco. (Marino *et. al.* 2013). Além do domínio RQC, Marino *et. al.*, (2013) observou a presença de outro domínio, o domínio chamado HRDC (Helicase-and-RNaseD-like-C-terminal) (Figura 06). O domínio HRDC está presente nas proteínas da família RecQ e, em conjunto com o domínio RQC, possuem um importante papel na especificidade a diferentes substratos tanto quanto na

atividade enzimática da proteína. Além disso, HRDC parece ter um papel central na interação proteína:proteína. (Killoran & Keck, 2006;Vindigni *et. al.*, 2010).

O domínio HRDC está envolvido no reconhecimento de diferentes substratos com diferentes especificidades. (Figura 06) Vindigni *et. al.*, (2010) e Killoran & Keck, (2006) mostraram que esse domínio é altamente variável dentre os membros da família RecQ. (Killoran & Keck, 2006) (Vindigni *et. al.*, (2010). Liu *et. al.*, (1999) mostraram que interações do domínio HRDC em *S. cerevisiae* (Sgs1) ocorrem por meio de interações eletrostáticas. (Liu *et. al.*, 1999).



**Figura 06 – A)** Domínio HRDC da helicase BLM (Salmão) e da RecQ helicase de *E. coli* (Amarelo). **B)** A sequência de aminoácidos do domínio HRDC de *E. coli* (box azul) e de BLM helicase (box vermelho) são mostrados. Apesar da baixa homologia e similaridade das sequencias de aminoácidos, os domínios possuem estruturas semelhantes.

Proteínas da família RecQ interagem com diversas proteínas com diferentes funções no metabolismo de DNA. (Opresko *et. al.* 2004). Além disso, foi observado que o estado de oligomerização das proteínas RecQ têm um importante papel na ativação e

funcionalização dessas proteínas. Muzzdini *et. al.*, (2007) mostrou que o nível de oligomerização é associado com atividade diferenciada da proteína: oligômeros grandes (tetrâmeros) estão associados com a atividade de ligação a fitas simples de DNA, como em funções desempenhadas na via de recombinação homóloga. Por outro lado, monômeros e dímeros estão associados com a função de helicase. (Muzzdini *et. al.*, 2007)

## 2.5 Funções desempenhadas pelas proteínas da família RecQ

### 2.5.1. Papel na replicação do DNA

Proteínas da família RecQ atuam durante o processo de replicação do DNA agindo sobre forquilhas de replicação bloqueadas (i.e. quebra na fita de DNA a qual a forquilha de replicação está atuando) ou forquilhas de replicação desfeitas. Recombinação homóloga está intimamente relacionada com ambos processos por permitir o reparo de lesões que bloqueiam a forquilha de replicação. Portanto, proteínas RecQ são capazes de resolver diversas estruturas intermediárias formadas durante o processo de recombinação. Além disso, proteínas RecQ previnem a troca de conteúdo gênico inapropriada entre as fitas de DNA durante o re-início da replicação. (Bachrati & Hickson, 2003) (Hickson, 2003). Apesar disso, é sugerido que proteínas RecQ atuem apenas como colaboradoras no processo de recuperação de forquilhas de replicação danificadas. (Opresko *et. al.*, 2004). Entretanto, RecQ4 e RecQ1 possuem um papel essencial no início da replicação. RecQ4 interage, por intermédio da proteína MCM10, com o complexo MCM 2-7 que possui função de helicase. (Xu *et. al.* 2009).

### 2.5.2. Papel na via de recombinação homóloga (HR),

Recombinação Homóloga (HR) é um processo essencial para a manutenção da integridade genômica reparando quebra na dupla-fita de DNA, troca de isotipo de imunoglobulinas (Ig), sucesso durante o processo de replicação do DNA (Harrigan *et. al.*

2006) e promove variabilidade durante a formação de células reprodutivas. Proteínas RecQ possuem uma alta afinidade por intermediários formados nesse processo, tais como triplo-DNA e quadruplo-DNA além de evitarem a troca de material genético descontrolada entre fitas de DNA. (Opresko *et. al.*, 2004). Além disso, estudos mostram que a helicase WRN está intimamente relacionada com outros tipos de vias de reparo, tais como o reparo de junção de pontas não homologas (NHEJ) e reparo por excisão de base (BER). (Ahn *et. al.*, 2004)(Wu *et. al.* 2006) (Svendsen & Harper, 2010) (Heyer *et. al.*, 2010) (Cheng *et. al.*, 2004).

### 2.5.3. Papel no metabolismo e manutenção de telômeros

Telômeros são regiões ao final de cada cromossomo e são compostas de repetições de sequências ricas em Guanina *in tandem* e em conjunto com proteínas acessórias formam uma estrutura característica de telômeros, denominada de *T-loop*. (O'Sullivan & Karlseder, 2010) (Griffith *et. al.*, 1999). Eles são importantes para a diferenciação das extremidades dos cromossomos das regiões onde houve quebra da dupla-fita e, além disso, facilitam a replicação das extremidades dos cromossomos. (Monnart Jr. 2010). Assim, o encurtamento das regiões teloméricas tem sido associado como um dos fatores responsáveis pelo envelhecimento e desenvolvimento de cânceres. Portanto, disfunções provocadas na estrutura dos telômeros podem iniciar respostas aos danos ao DNA e, além disso, pode provocar fusão e quebra dos cromossomos. (de Lange, 2002).

Proteínas WRN e BLM podem facilitar a replicação das regiões teloméricas promovendo a abertura das estruturas *T-loop* e, não obstante, a atividade enzimática da proteína WRN é importante para a replicação da fita descontínua. (O'Sullivan & Karlseder, 2010) (Opresko, 2008) (Crabbe *et. al.* 2007) Apesar dos esforços aplicados para caracterizar funcionalmente a proteína RecQ4 nos diversos aspectos citados acima, ainda

pouco se conhece sobre suas funções bioquímicas estruturais. Assim, outros organismos que não *H. sapiens* são propostos como modelos para a obtenção dessas informações.

### 2.6 Hrq1: uma homóloga a proteína RecQ4

Leveduras (*S. cerevisiae*) foram propostas por Ashton & Hickson 2010 como organismos modelos para o estudo de proteínas da família RecQ. Pensava-se que *S. cerevisiae* possuía apenas uma proteína da família RecQ em seu genoma, a proteína Sgs1. Entretanto, Barea *et. al.* (2008) encontrou, através de análises de bioinformática, uma nova proteína da família RecQ presente em plantas e fungos, a proteína Hrq1 (*Homologous to RecQ helicase 1*), que esta proteína seria ortóloga a proteína RecQ4 de metazoários. (Barea *et. al.* 2008). Recentemente, Grocock *et. al* (2012) propuseram a existência de Hrq1 em *Schizosaccharomyces pombe*. Além disso, o estudo propôs que HRQ1 estaria intimamente relacionada com a estabilidade genômica e reparo do DNA, assim como sua ortóloga RecQ4.

Através de análises filogenéticas com alinhamento de sequências, HRQ1 forma um clado único em conjunto com membros de helicases da família RecQ em eucariotos. Além disso, através de análises estruturais *in silico*, foram identificadas regiões conservadas tais como DEAH box e domínio similar ao RQC em RecQ4, chamado SRecQ-C. Um novo domínio, chamado RHCD presente apenas em RecQ4 de humanos e HRQ1 também foi proposto. O massa molecular dessa proteína foi estimado em 124kDa, tendo comprimento de 1077 aminoácidos. (Barea *et. al.* 2008). Entretanto, não foram realizados experimentos *in vivo* ou *in vitro* para confirmar os resultados encontrados pelas ferramentas bioinformáticas.

Com análises bioquímicas, Grocock *et. al.* (2012) sugeriram possíveis funções para proteína Hrq1 em *S. pombe*. Estudos *in vivo* com mutantes e *in vitro* sugeriram que Hrq1 estaria intimamente relacionada no reparo por excisão de nucleotídeo (NER) e no reparo

pós-replicativo (PRR). Além disso, a porção C-terminal não interfere na atividade helicase da proteína mas interfere na localização subcelular sugerindo que a porção C-terminal funcione como um “sinalizador” da proteína Hrq1. Hrq1 possui, de fato, atividade helicase com polaridade 3'->5' concordante com a atividade de sua suposta ortóloga RecQ4. (Grocock *et. al.* 2012).

Estudos com Hrq1 de *S. cerevisiae* mostraram que esta proteína possui afinidade por substratos em forma de forquilha de replicação. (Kwon *et. al.* 2012) Além disso, estes autores confirmaram a polaridade 3'->5' do domínio helicase como proposto por Grocock *et. al.* (2012). Curiosamente, eles mostraram que a atividade helicase de Hrq1 é aumentada significativamente pela presença de uma longa porção de 3'OH-DNA simples fita. E, contrariando outras helicases da família RecQ, atividade da proteína Hrq1 é inibida pela presença de SBP (single-strand binding proteins) tais como RPA. Kwon *et. al* (2012) ainda mostrou que Hrq1 possui atividade de DNA *strand annealing* dependente de ATP, assim como a proteína RecQ4.

Embora sejam necessários estudos bioquímicos e estruturais, a existência de uma proteína ortóloga a RecQ4 em leveduras abre um novo caminho para o melhor entendimento dos processos relacionados a síndromes provocadas por mutações desse gene. (Barea *et. al.* 2008). .

### **3. OBJETIVOS**

#### *3.1. Objetivo geral*

Este trabalho visa obter informações estruturais da proteína Hrq1 para inferir sobre a sua função no metabolismo do DNA em *S.cerevisiae*. Como a proteína Hrq1 é uma helicase semelhante a RecQ4, as informações estruturais obtidas poderão ajudar na elucidação dos processos envolvidos nas síndromes Rothmund-Thompson, Baller-Gerold e RAPADILINO relacionadas a proteína RecQ4.

#### *3.2. Objetivo específico*

i) clonar e expressar a proteína Hrq1 com o sistema *E. coli*; ii) estabelecer um protocolo de indução em larga escala; iii) estabelecer e otimizar um protocolo de purificação da proteína Hrq1; iv) verificar condições oligoméricas desta proteína e de seus domínios em solução; e v) estabelecer e otimizar as condições de cristalização da proteína Hrq1

## 4. MATERIAIS E MÉTODOS

4.1. *Análises estruturais in silico da proteína Hrq1 de S.cerevisiae* – A predição de estruturas secundárias da proteína Hrq1 foi realizada utilizando os programas FFAS03 (disponível no site: <http://ffas.burnham.org/ffas-cgi/cgi/ffas.pl>) e PsiPred (Disponível no site: <http://bioinf.cs.ucl.ac.uk/psipred/>). O alinhamento de sequências foi realizado pelos programas BLAST e ClustalW (disponível nos sites: <http://blast.ncbi.nlm.nih.gov/> e <http://www.ebi.ac.uk/Tools/msa/clustalw2/>, respectivamente).

4.2. *Clonagem do gene Hrq1 no vetor pET SUMO/CAT* – O desenho dos clones foi feito manualmente. O método utilizado para a clonagem foi método descrito por Unger et. al. 2010. Os *primers* (Sigma) constituem um *primer* misto contendo metade de sua sequência complementar ao vetor e metade complementar ao gene de interesse. O vetor utilizado foi pET SUMO/CAT (Invitrogen®). Os *primers* utilizados são mostrados no Apêndice 02. As condições das reações de amplificação do gene *hrq1* são descritas na tabela 01. O produto de PCR foi purificado utilizando o kit PCR purifier Qiagen® e a concentração de DNA foi medida utilizando o sistema NanoDrop® com leitura a 260nm. O produto purificado foi usado no segundo passo de amplificação. Para visualizar os resultados da amplificação por PCR foram realizadas eletroforeses em gel de agarose 1% em Tampão 1x TBE à 90V. O corante utilizado é SYBR Safe Invitrogen e os géis foram revelados utilizando luz UV proveniente do sistema BioRad®. A proteína inteira e seus domínios individualmente e/ou combinados foram clonados (Tabela 02). A tabela completa dos clones, tamanhos, pl teórico, coeficiente de extensão, nome dos fragmentos e seus respectivos símbolos encontra-se no Apêndice 01. O produto de PCR foi digerido da fita parental com a enzima *DpnI* (Invitrogen) por 1 hora a 37°C. Em seguida, 90µL de células DH5α foram incubados com 1µL de produto de PCR por 30 minutos à 4°C. Choque-térmico de 45 segundos foi realizado (42°C), 500µL de meio LB foi adicionado ao tubo e

as células foram incubadas em agitação por 1 hora. 200µL da solução foi plaqueado em meio contendo LB agar suplementado com canamicina (50µg/mL) e incubadas overnight à 30°C. As colônias foram coletadas e incubadas em tubo de ensaio contendo 5mL de meio LB suplementado com canamicina (50µg/mL) em agitação a 37°C overnight. A cultura de células foi centrifugada para a extração e purificação de plasmídeos utilizando Qiagen® Spin MiniPrep Kit.

	<b>Temperatura</b>	<b>Tempo</b>
1.	95°C	3'
2.	95°C	30''
3.	68°C	1'
4.	68°C	3'
5.	Go To 2	Rep 4
6.	95°C	30''
7.	60°C	1'
8.	68°C	3'
9.	Go To 6	Rep 4
10.	95°C	30''
11.	65°C	1'
12.	68°C	3'
13.	Go To 10	Rep 4
14.	95°C	30''
15.	50°C	1'
16.	68°C	3'
17.	Go To 14	Rep 4
18.	72°C	10'

19.	Hold 15°C	-
-----	-----------	---

**Tabela 01-** Programa Touchdown de PCR utilizado para amplificação do gene *hrq1* de *S. cerevisiae*.

Domínio (resíduos)	Símbolo	Tamanho (pb)
Full length 1 - 1031	$\delta$	3093
Full length 1 - 1077	$\Delta$	3231
Cdt1 1-191	$\alpha$	573
Cdt1 1-277	$\beta$	831
Helicase 268-645	$\gamma$	1119
Helicase 268-663	$\varepsilon$	1188
Helicase 277-645	$\eta$	1092
Helicase 277-663	$\theta$	1161
Helicase+Cterminal 268-1031	$\lambda$	2292
Helicase+Cterminal 268-1077	$\mu$	2430
Helicase+Cterminal 277-1031	$\pi$	2265
Helicase+Cterminal 277-1077	$\sigma$	2403
Cterminal 644-1031	$\phi$	1164
Cterminal 644-1077	$\chi$	1302
Cterminal 668-1031	$\varphi$	1092
Cterminal 668-1077	$\omega$	1230

**Tabela 02** Fragmentos do gene *hqr1* clonados para obtenção dos diferentes domínios. O nome das construções e o seu tamanho molecular (em pb) são descritos.

4.3. *Purificação de plasmídeo em pequena escala (MiniPrep)* – O protocolo de MiniPrep utilizado neste trabalho é descrito por Qiagen® Handbook. Resumidamente, 1) Resuspender o pellet em tampão P1 e transferir o conteúdo para um tubo de microcentrífuga. (Tabela 3); 2) Adicionar tampão P2 e agitar o tubo; 3) Adicionar tampão N3 e agitar imediatamente; 4) Centrifugação por 10 minutos à 14300 rpm; 5) Transferir o conteúdo para uma coluna QIAprep spin, centrifugar à 14300 rpm por 2 minutos e lavar

utilizando o tampão PE por três vezes; 6) Eluir com tampão EB; Plasmídeos foram armazenados à -20°C e, além disso, uma alíquota foi enviada para sequenciamento (BMR Genomics).

Tampão	Composição
P1	50mM Tris pH 8.0, 10mM EDTA, 100µg/mL RNase A
P2	200mM NaOH, 1% SDS
P3	3.0M acetato de potássio, pH 5.5
N3	4.2 M Guanidina, 0.9 M acetato de potássio, pH 4.8
PE	10 mM Tris pH 7.5, 80% Etanol
EB	10mM Tris pH 8.5

**Tabela 03-** Composição dos tampões utilizados no Qiagen Spin MiniPrep Kit

*4.4. Transformação bacteriana* – As estirpes BL21RP, Rosetta-2, C41 e BL21Star foram utilizadas. BL21RP e Rosetta-2 são estirpes padrões para super-expressão de proteínas por possuírem códons raros. C41 é uma estirpe para a expressão de genes que codificam proteínas tóxicas para a célula e, portanto, tem expressão reduzida. BL21Star possui uma mutação no gene RNaseE que maior estabilidade ao mRNA e, teoricamente, garante maior expressão da proteína de interesse. Em todas as estirpes a transformação foi realizada utilizando diferença de temperatura (células termocompetentes). O plasmídeo (0,4µL) foi incubado com 30-40µL de célula a 4°C por 30 minutos. Em seguida, um choque-térmico de 45 segundos a 40°C foi realizado e após 2 minutos a 4°C, 500µL de meio LB foram adicionados e as células foram incubadas a 37°C por 60 minutos sob

agitação. Placas contendo LB agar e canamicina (50µg/mL) foram utilizadas para o crescimento das colônias. As placas foram incubadas a 30°C overnight.

*4.5. Teste de expressão* – Pré-inóculo foi realizado utilizando 5 colônias das placas para obter uma média de expressão da proteína de interesse. O pré-inóculo foi realizado com 1L meio LB suplementado com canamicina (50µg/mL) e incubadas à 37°C graus em agitação. Uma alíquota de 22,5mL de pré-inóculo foram adicionadas a 500mL de meio de cultura para de auto-indução (Item 4.6). Para a indução com 0,5mM de IPTG, foram utilizados 200µL de pré-inóculo para 1000mL de meio de cultura. Além disso, os testes foram realizados em três diferentes temperaturas (18°C, 25°C e 37°C) com as 4 estirpes de células citadas acima.

*4.6. Condições de auto-indução* – Protocolo utilizado foi descrito por Studier. F. W. (2005). O preparo do meio de auto-indução requer alguns tampões adicionais (Tabela 04). Com este protocolo, é necessário que a D.O da cultura no início do crescimento para indução seja alta. Por esse motivo, foram utilizados 22,5mL de pré-inóculo para cada 500mL de cultura. Para a indução, foram utilizados 20 frascos.

Culturas foram crescidas sob agitação (240rpm) por 220 minutos à 37°C. Em seguida, a temperatura foi reduzida para 17°C e a cultura foi incubada 18-24 horas sob agitação. Esse protocolo é baseado na saturação do meio de cultura. Glucose é utilizada como primeira fonte de açúcar na cultura (aproximadamente as primeiras 4 horas). Na ausência de glucose, a cultura utiliza lactose e a indução tem início uma vez que o vector pET SUMO/CAT (Invitrogen®) possui promotor T7 e funciona sob regulação do sistema do operon *lac*. Esse período de indução corresponde ao período o qual a cultura é mantida à 17°C. Após auto-indução, a cultura foi então centrifugada por 30 minutos 8000rpm e o pellet armazenado à -80°C.

<b>Tampão</b>	<b>Composição</b>	<b>Volume</b>
50x5052	25g Glicerol, 2.5g Glucose, 10g $\alpha$ -lactose	20mL
20xNPS	0.5M $(\text{NH}_3)_2\text{SO}_4$ , 1M $\text{KH}_2\text{PO}_4$ , 1M $\text{Na}_2\text{HPO}_4$	50mL
1M $\text{MgCl}_2$	-	1mL
Meio LB	-	929mL
		1000mL

**Tabela 04** - Composição dos tampões utilizados no protocolo de auto-indução.

4.7. *Obtenção do extrato bruto de células* – O pellet obtido como descrito no item 4.6 foi re-suspenso em tampão de lise (1mg/mL DNase, 10mM  $\text{MgCl}_2$ , 1mM Inibidor de Protease AEBSF, 1mg/mL Lisozima, 250mM NaCl, 5mM imidazol, 5mM  $\beta$ -Mercaptoetanol, 20mM Tris pH 8,0, 5% Glicerol.) e incubados à 4°C em agitação. Em seguida, as células foram sonicadas em uma série de 30 minutos de sonicação (45” repouso – 15” sonicação). A solução foi centrifugada em tubos de policarbonato à 17000rpm por 20minutos a 4°C. O pellet foi descartado e o sobrenadante foi reservado para purificação.

4.8. *Purificação dos domínios da proteína Hrq1* – Nesse trabalho são apresentados o protocolo de purificação para os domínio  $\alpha$  e  $\omega$  (ver Tabela 04). O domínio  $\alpha$  foi primeiramente purificado em uma coluna de Ni-HisTrap (5 mL, GE HealthCare®). Eluição foi realizada em passo único com tampão com mesma composição do tampão de lise, entretanto, com 400mM de imidazol.

Em seguida, foi realizado a clivagem do cauda SUMO utilizando SUMO protease na proporção 1:500 (mg:mg) em tampão 25mM Tris pH 8.0, 10% Glicerol. A SUMO

protease é inativa em altas concentrações de sal ou imidazol. A clivagem foi realizada por 3 horas a 4°C sob agitação branda. Em seguida, a solução foi injetada novamente em coluna HisTrap (5mL, GE HealthCare®). O volume de injeção foi reservado e concentrado utilizando sistema MiliPore® com *cut-off* molecular de 10kDa. Em seguida, a amostra foi injetada por duas vezes em uma coluna de Gel Filtração Superdex 75/100 HiLoad (24mL, GE HealthCare®) equilibrada com tampão contendo: 200mM NaCl, 5%Glicerol, 50mM Tris pH 7,5 e 5mM  $\beta$ -Mercaptoetanol.

Para o domínio  $\omega$  foi utilizado primeiramente uma coluna de Ni-HisTrap (5mL, GE HealthCare®) seguido de clivagem com Sumo Protease e novamente uma coluna de Ni-HisTrap (5mL, GE HealthCare®). Em seguida, a amostra foi injetada em uma coluna de Heparina HiTrap (5mL, GE HealthCare®) equilibrada com. tampão composto de 50mM Tris pH 7,5, 50mM NaCl. A eluição foi realizada manualmente, com aumento progressivo de 100mM em cada fração até a concentração de 2M de NaCl. As frações foram analisadas em SDS-PAGE – o gel foi corado com InstantBlue®. As frações contendo a proteína de interesse foram concentradas em sistema MiliPore® *cut-off* de 30kDa e injetadas em uma coluna de Gel Filtração Superdex 75/100 HiLoad (24mL, GE HealthCare®). As frações foram analisadas em SDS-PAGE.

*4.9. Determinação das condições de cristalização do domínio  $\omega$*  – O sistema Mosquito® LCP foi utilizado para o preparo das placas de cristalização. Os kits utilizados foram: PACT suit, Classic II suit e JCSG suit (QIAGEN®) e PEG Ion 1, pH Clear II e Index (Hampton®). As gotas continham a proporção 0,1 $\mu$ L:0,1 $\mu$ L (Proteína : Solução de reservatório) e foram realizadas pelo método de gota sentada. As placas foram incubadas à 25°C. A análise das gotas foi realizada manualmente em um microscópio estereoscópio sob luz fria polarizada.

## 5. RESULTADOS E DISCUSSÃO

### 5.1. Análise *in silico* da proteína Hrq1 de *S. cerevisiae*

Análise realizada com o programa BLAST (<http://blast.ncbi.nlm.nih.gov/Blast>; para Hrq1p *S. cerevisiae* S288c NP\_010577.3) revelou que esta proteína é composta por três domínios: o domínio Cdt1 (resíduos 1 – 238), o domínio DEAD/DExH Helicase box (resíduos 292-633) e o domínio na porção C-terminal (resíduos 899-989) com função desconhecida DUF (*domain of unknown function*) (Figura 7). A presença do domínio DEAD/DExD box e o domínio Helicase indicam que Hrq1 é de fato uma helicase da superfamília II assim como a proteína RecQ4. A similaridade de sequências entre a proteína Hrq1 de *S.cerevisiae* com proteínas da família RecQ é superior a 40% (Barea *et. al* 2008)

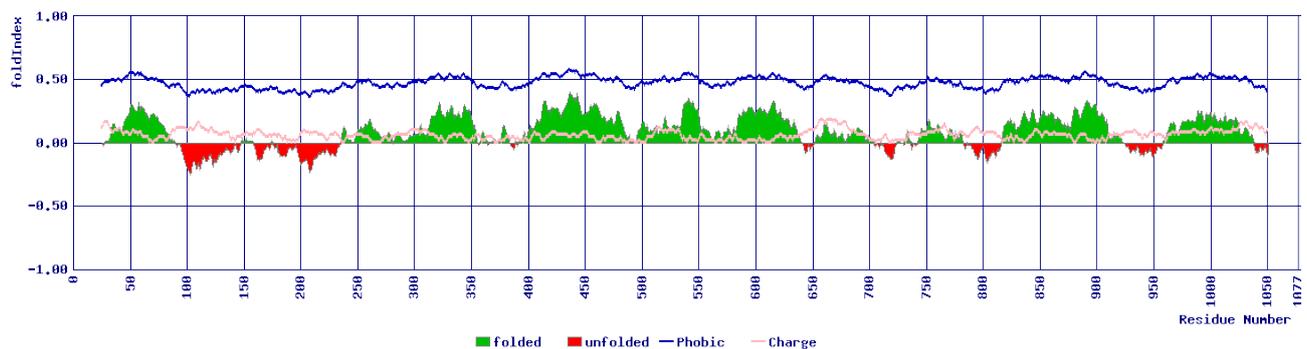
A mesma sequência de aminoácidos foi utilizada no programa FFAS03. Neste programa, além dos domínios já identificados pelo programa BLAST, foi identificado um domínio Cdt1, na porção N-terminal.. O domínio Cdt1 é encontrado em proteínas envolvidas na replicação do DNA e compreende os resíduos 20-207 da proteína Hqr1. De uma forma geral, a análise *in silico* sugere que a proteína Hrq1 é composta por três domínios: Cdt1, DEAD/DExD box Helicase e o domínio DUF. (Figura 07)



**Figura 07**– Arquitetura da proteína Hrq1 de *S. cerevisiae*.

HHPred foi utilizado para prever o enovelamento da proteína (Figura 08). As porções em verde indicam a probabilidade de enovelamento de uma região e as regiões em vermelho indicam que possivelmente esta região não possui estruturas secundárias e

que, portanto, não é enovelada. Como observado, a maior parte da proteína apresenta regiões com alta probabilidade de enovelamento. A região menos estruturada corresponde aos resíduos de aproximadamente 45 a 240, que se sobrepõe ao domínio Ctd1. Apesar disso, o domínio Cdt-1 foi incluído na estratégia de clonagem devido a sua alta similaridade com a proteína Cdt-1 de *S. cerevisiae*. Esta proteína é um importante fator na formação do complexo Pré-replicativo (Watson *et. al* 2006) e o domínio Cdt-1 presente em Hrql poderia indicar um papel na replicação do DNA.



**Figura 08** – Previsão do enovelamento da proteína Hrql pelo programa HHPred.

As regiões em verde indicam alta probabilidade de enovelamento, por outro lado, as regiões em vermelho indicam prováveis regiões sem qualquer estrutura secundária.

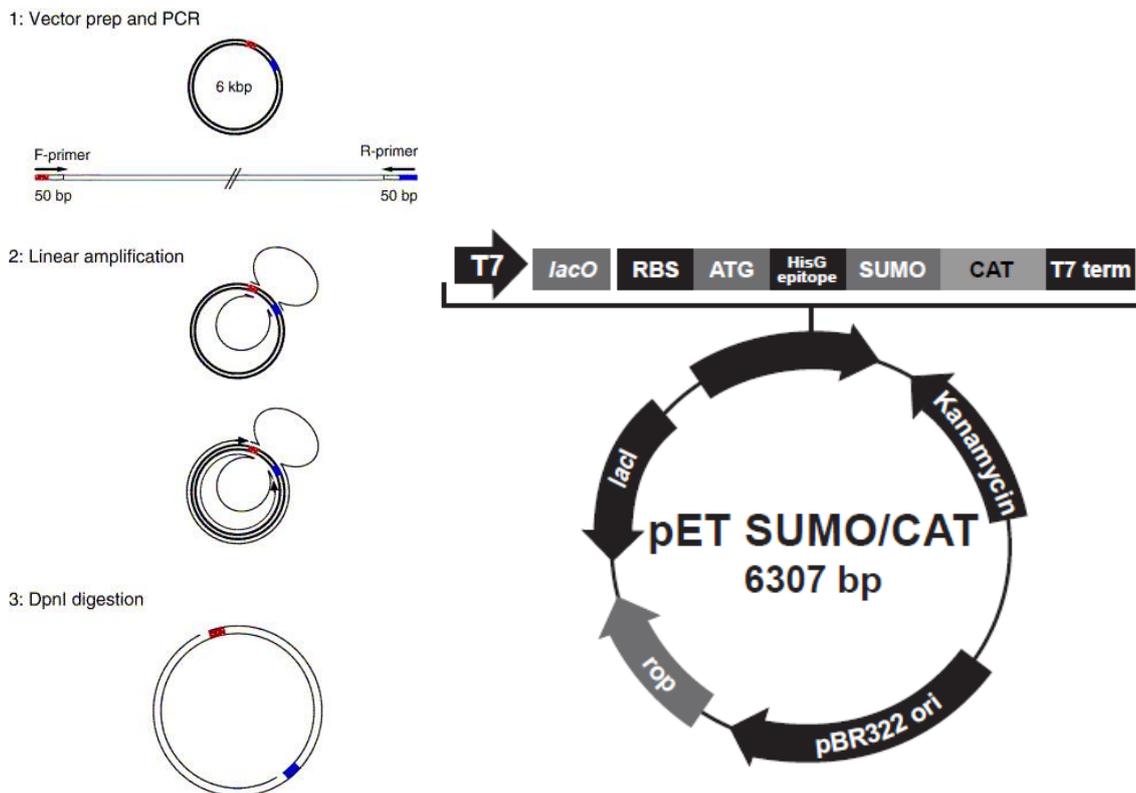
O programa PsiPred foi utilizado para prever as estruturas secundárias da proteína. Por este programa, a proteína é composta por uma região rica em folhas  $\beta$  na porção C-terminal, coincidente com o domínio DUF.

Múltiplos alinhamentos de sequências utilizando o programa ClustalW foram realizados com o intuito de determinar domínios e verificar as características da proteína. A partir dos dados bioinformáticos obtidos, as regiões a serem clonadas foram determinadas manualmente (Materiais e Métodos – Item 4.2).

## 5.2. Clonagem do gene *hrq1* *S. cerevisiae*

A técnica de clonagem utilizada foi desenvolvida por Unger *et. al.* (2010). O método de *RF cloning* utiliza *primers* mistos para amplificação, ou seja, uma região de 15-25pb presente nos *primers* é complementar ao gene de interesse enquanto que os outros 15-20pb são complementares ao vetor o qual o gene será inserido. No total, os *primers* têm em média 40-50pb, sendo metade complementar ao gene de interesse e metade, ao vetor de clonagem. A lista dos *primers* utilizados é apresentada no Apêndice 02. Essa técnica consiste em dois *rounds* de amplificação por PCR. No primeiro *round*, utiliza-se os *primers* mistos e o DNA molde contendo o gene de interesse. O produto de PCR criado nessa etapa consiste no gene de interesse flanqueado pela região complementar ao vetor. A eficiência da reação de PCR foi verificada em gel de Agarose 1% e o produto de PCR foi purificado. Nessa etapa, a amplificação é exponencial, como em uma reação de PCR comum (Figura 09).

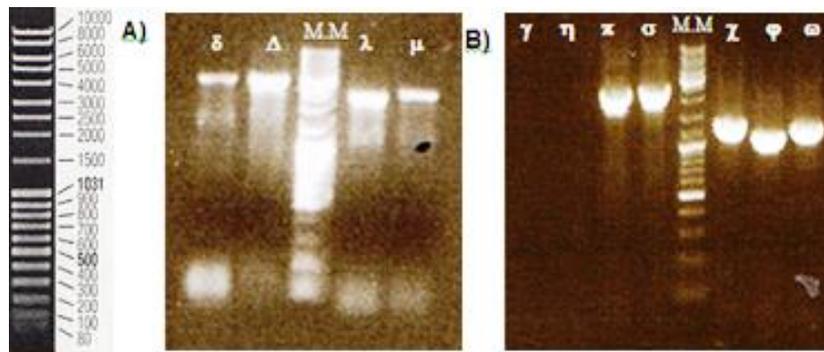
Na segunda etapa, o produto de PCR gerado na primeira etapa é utilizado como *primer* e o DNA molde é o próprio vetor de expressão. O produto de PCR possui uma região complementar ao vetor de clonagem e, portanto, funciona como *primer* para a amplificação completa do vetor. Quando o vetor é amplificado, o inserto é amplificado como parte do mesmo e o gene é inserido. Contudo, essa etapa caracteriza-se pela formação linear do produto de PCR. Isso porque os *primers* da segunda etapa reconhecem o vetor partindo do princípio da recombinação. Portanto, o vetor para ser amplificado, deve ser reconhecido por dois *primers* complementares. Além disso, o produto do primeiro ciclo é utilizado como *primer* para o segundo ciclo. (Unger *et.al.* 2010) (Figura 09).



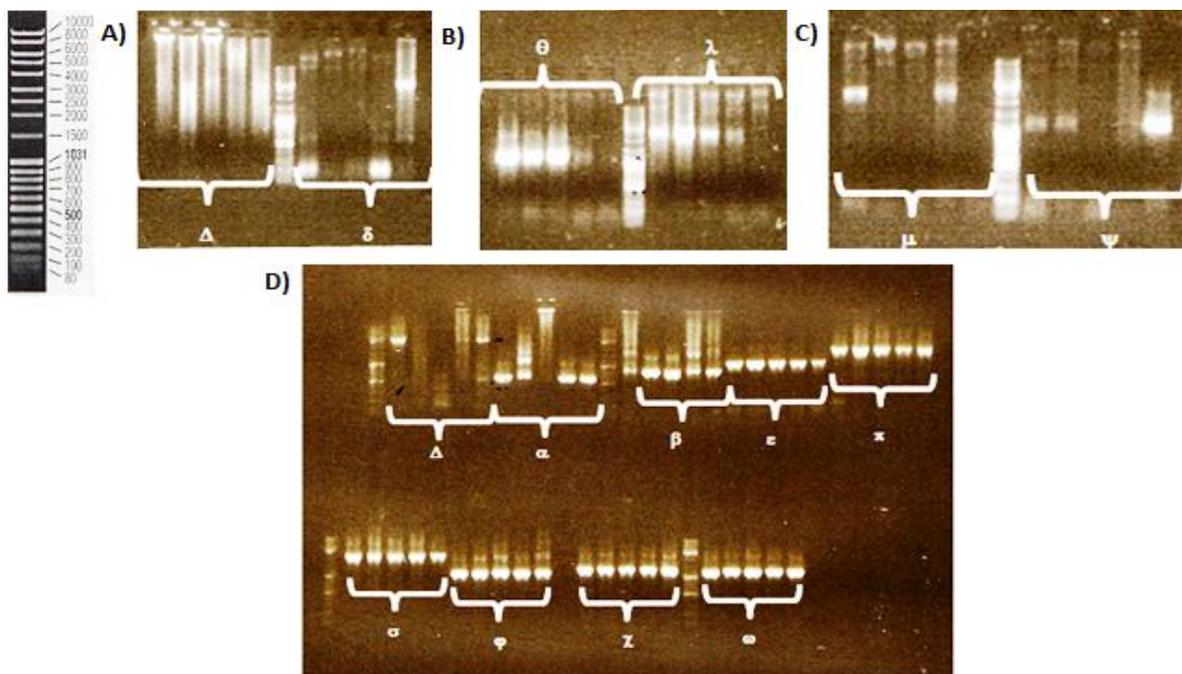
**Figura 09** – Esquema de *RF-cloning* no vetor pET-SUMO/CAT. À esquerda – visão geral do processo de *RF cloning*. 1) indica a primeira reação de PCR amplificando o gene de interesse; 2) amplificação linear do vetor + inserto; 3) digestão com enzima *DpnI*. À direita – Vetor utilizado nesse trabalho. pET SUMO/CAT possui um sitio de clonagem ajustado ao SUMO e a cauda poli-His. Além disso, existe um sítio de clivagem reconhecido pela proteína SUMO protease que remove a cauda na porção N-terminal.

A clonagem do gene *hrq1* foi realizada conforme o protocolo de Unger *et.al.* (2010). pET-SUMO/CAT (Invitrogen) foi utilizado como vetor. A expressão de proteínas por esse vetor está associada à fusão de uma cauda na porção N-terminal, contendo a proteína SUMO e uma cauda poli-His, além do sítio de clivagem. Como apresentado na tabela 04, diferentes construções dos diferentes domínios da proteína Hrq1, além da proteína completa foram inseridos no vetor de expressão e transformados em diversas estirpes de *E. coli* (Materiais e Métodos - Tabela 02).

A clonagem foi realizada com alto rendimento, sendo obtidas quase todas as construções desejadas com exceção das construções  $\gamma$  e  $\eta$  (Figura 10 e 11)



**Figura 10** - Perfil de eletroforese em gel de agarose 1% dos produtos de PCR obtidos a partir das diferentes construções de Hrq1 para a clonagem no vetor pET-SUMO/CAT. No canto esquerdo – Marcador de tamanho molecular GenRuler (SIGMA). **A)** e **B)** Produtos de PCR das construções  $\delta$ ,  $\Delta$ ,  $\lambda$ ,  $\mu$ ,  $\gamma$ ,  $\eta$ ,  $\sigma$ ,  $\chi$ ,  $\pi$ ,  $\varphi$  e  $\omega$ .



**Figura 11** - Perfil de eletroforese em gel de agarose 1% dos produtos de PCR obtidos a partir das colônias recombinantes para *screening* dos clones. No canto esquerdo superior

– Marcador de tamanho molecular GenRuler (SIGMA). **A) B) C) e D)** Produtos de PCR no *screening* das colônias obtidas para as construções  $\Delta, \delta, \alpha, \beta, \epsilon, \pi, \theta, \mu, \lambda, \sigma, \varphi, \chi$  e  $\omega$ .

A construção  $\omega$  representa a porção C-terminal da proteína Hrq1. Especificamente, esta construção foi desenhada para compreender o domínio de função desconhecida (DUF) abrangendo os resíduos 643-1077. Este domínio é rico em folhas  $\beta$  segundo análises do programa PsiPred. Além disso, como as demais proteínas da família RecQ, a porção C-terminal é fundamental para conferir a afinidade ao substrato, o foco deste trabalho foi este domínio.

A construção  $\omega$  apresenta cisteínas conservadas – CxxxCxxCxxxxC<sub>18</sub>xC - que possivelmente podem estar envolvidas em coordenação de algum metal. Como observado para a proteína RecQ4 que contém um domínio ligante de Zn na porção C-terminal. (Marino *et. al.* 2013) Além disso, as cisteínas são conservadas em diferentes organismos e diferentes helicases (DNA ou RNA helicases) sugerindo que este domínio desempenhe uma importante função para a proteína.

A construção  $\alpha$  é composto pelo domínio Cdt-1 que compreende os resíduos 1-191.

### 5.3. Otimização do protocolo de expressão

Os testes de expressão das construções foram realizadas utilizando diferentes estratégias e diferentes estirpes de *E.coli*. O teste-padrão foi realizado com células BL21RP (Invitrogen) em meio LB, na temperatura de 37°C, indução por um período de 4 horas com IPTG 0,5mM. Os domínios  $\omega$  e  $\alpha$  não foram expressos na forma solúvel, o que torna as proteínas inviáveis para experimentos de cristalização, análises bioquímicas e análises biofísicas. As temperaturas de 25°C e 18°C foram testadas na mesma estirpe

celular, entretanto com período de indução de 18 horas com IPTG 0,5mM. A redução na temperatura de indução está associada a uma expressão mais lenta. Por esse motivo, algumas proteínas têm a possibilidade de se enovelarem de maneira apropriada aumentando a sua solubilidade. Apesar de apresentarem diferentes níveis de expressão, as construções aqui testadas não sofreram nenhuma alteração na solubilidade da proteína pela redução da temperatura.

O próximo passo foi testar se diferentes estirpes podem influenciar na solubilidade da proteína. Cada estirpe apresenta diferentes sistemas de expressão, compostos por diferentes chaperonas bem como diferentes ambientes celulares. Células C41 (Invitrogen) foram utilizadas e foram obtidos os piores níveis de expressão dentre todas as estirpes testadas. Uma possível explicação é que C41 é utilizada para a expressão de proteínas tóxicas para a célula, assim, níveis reduzidos de expressão são esperados.

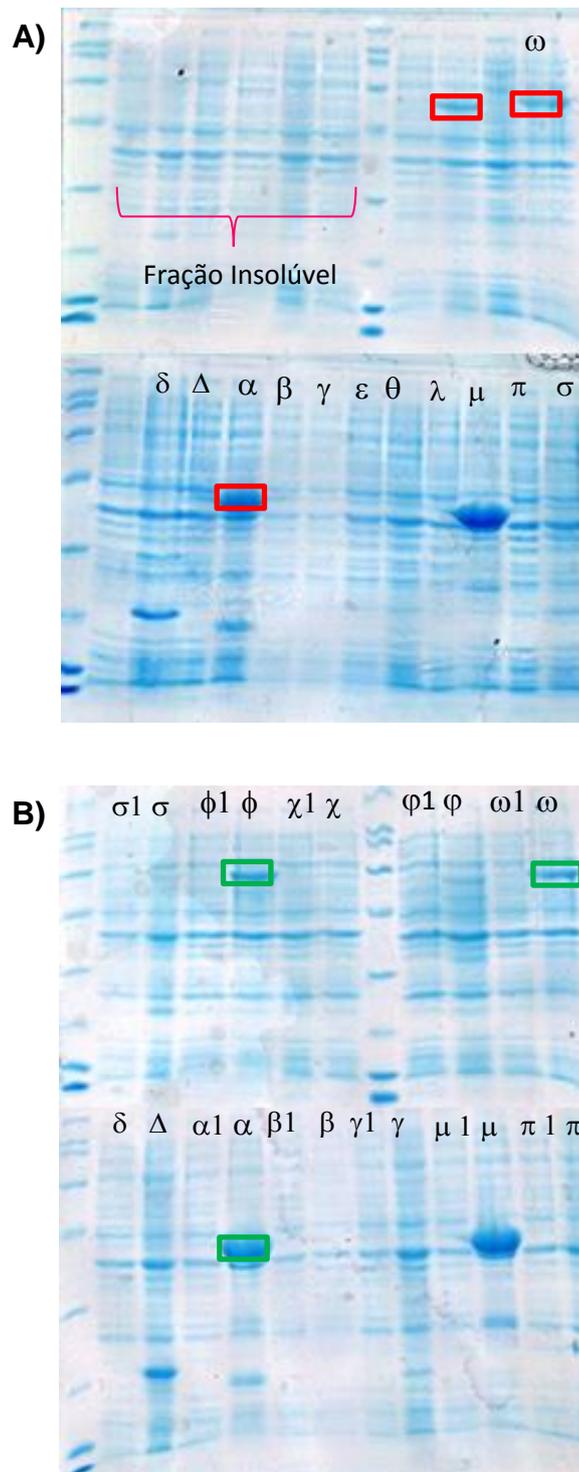
Domínio		BL21RP	C41 cells	Rosetta - 2				BL21Star
		18°C	18°C	Autoindução	18°C	25°C	37°C	Autoindução
Full length (1-1031)	$\delta$	x	x	x	x	x	x	x
Full length (1-1077)	$\Delta$	x	x	x	x	x	x	x
Cdt-1 (1-191)	$\alpha$	xx	xx	xxx	xx	xx	xx	xxx
Cdt-1 (1-277)	$\beta$							
Helicase (268-645)	$\gamma$							
Helicase (268-663)	$\epsilon$	xx	xx	x	x	x	x	x
Helicase (277-645)	$\eta$							
Helicase (277 -663)	$\theta$	x	x	x	x	x	x	X
Helicase + DUF (268-1031)	$\lambda$	x	x	x	x	x	x	X
Helicase + DUF (268-1077)	$\mu$							
Helicase + DUF (277-1031)	$\pi$	x	x	x	x	x	x	X
Helicase + DUF (277-1077)	$\sigma$	x	x	x	x	x	x	X
DUF (644-1031)	$\phi$	xx	x	xx	xx	xx	xx	xxx
DUF (644-1077)	$\chi$	x	x	x	x	x	x	x
DUF (668-1031)	$\varphi$	xx	x	xxx	xx	xx	xx	xx

DUF (668-1077)	ω	xx	x	xxx	xx	xx	xx	xxx

**Tabela 05** – Estirpes de *E.coli* e condições de indução testadas. A tabela representa as construções obtidas (x) que apresentaram expressão do vetor (xx) e as construções no qual as proteínas expressadas são solúveis (xxx).

Duas outras estirpes foram utilizadas, a Rosetta-2 e a BL21Star. Rosetta-2 foi testada nas temperaturas de 37°C, 25°C e 18°C com indução com IPTG 0,5mM e com protocolo de auto-indução (37°C – 17°C). Como não foi observada nenhuma expressão utilizando IPTG, decidido manter o protocolo de auto-indução em todas as expressões realizadas (Figura 12) (Materiais e Métodos – Item 4.6). As estirpes Rosetta-2 e BL21Star apresentaram as proteínas das construções α, φ, φ e ω na porção solúvel, resultado que não foi observado para a estirpe BL21RP. As construções α e ω foram escolhidas por serem expressas de maneira solúvel em ambas as estirpes e por serem os fragmentos completos dos domínios Cdt-1 e DUF, respectivamente. Um resumo do teste de expressão é apresentado na Tabela 05

α1 μ1 φ1 χ1 φ1 ω1    φ    χ    φ



**Figura 12** – Gel de SDS-PAGE 12% das frações solúveis e insolúveis obtidas a partir do teste de expressão utilizando o protocolo de auto-indução. Sobre as linhas estão os fragmentos indicados. O número “1” ao lado do fragmento (ex:  $\alpha_1, \omega_1, \dots$ ) indicam a fração insolúvel. **A)** BL21Star (caixas vermelhas indicam expressão do domínio de interesse

com os fragmentos  $\alpha$ ,  $\phi$  e  $\omega$ ). O colchete rosa indica as frações insolúveis; e **B)** Rosetta-2 (caixas verdes indicam expressão dos fragmentos de interesse  $\alpha$ ,  $\phi$  e  $\omega$ ).

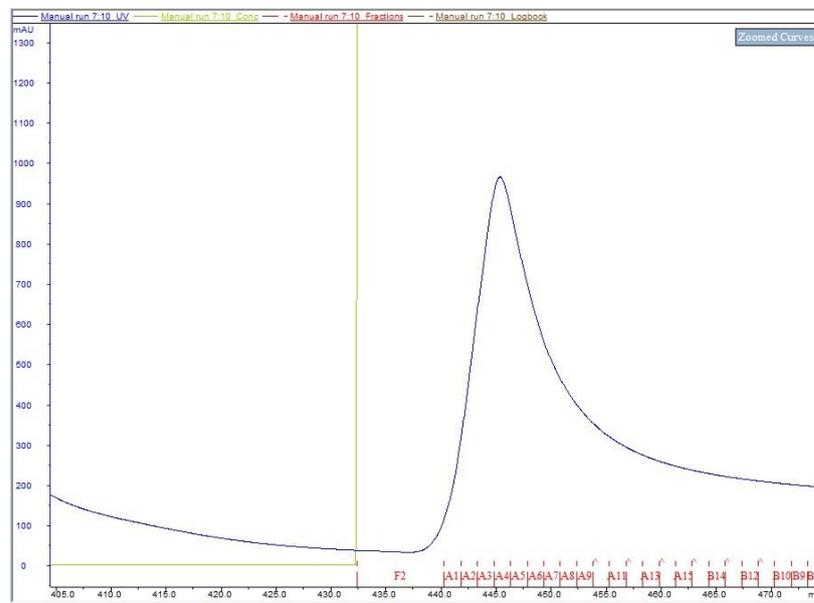
#### 5.4. Purificação dos domínios $\alpha$ e $\omega$

As etapas de purificação variaram entre as construções tendo em vista que ambas apresentam diferentes domínios e, portanto, possuem diferentes propriedades.

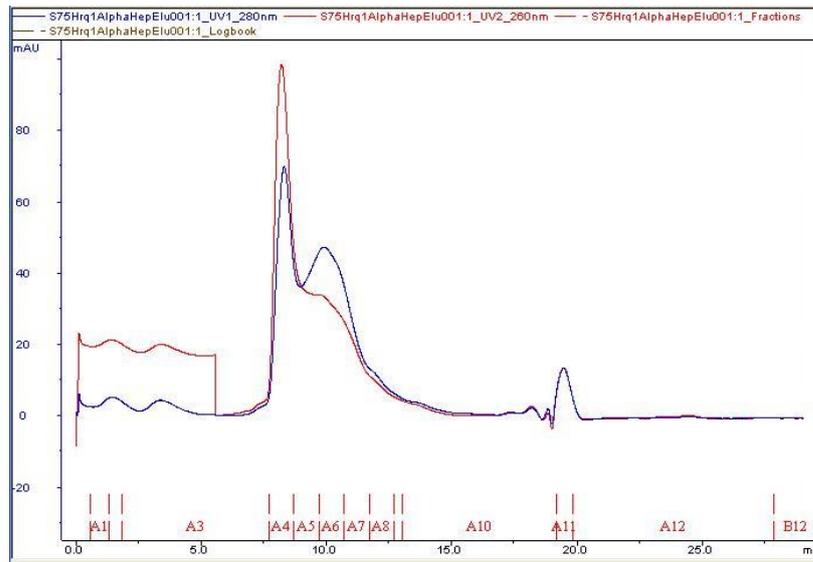
**5.4.1 Purificação do domínio  $\alpha$**  - Para o domínio  $\alpha$  foi utilizada uma primeira etapa em coluna de afinidade (Ni-HisTrap, 5mL, Ge Healthcare®) que apresenta afinidade com 6-His-tag. A eluição foi realizada em tampão contendo 500mM Imidazol (Figura 14). A fração foi então diluída 10x pois a atividade da proteína SUMO protease é inibida em presença de altas concentrações de sal e imidazol. SUMO protease reconhece um sítio de clivagem logo acima do início da proteína de interesse. A clivagem libera SUMO + 6-His. Como a SUMO protease possui a cauda poli histidina, ela interage com a coluna de Níquel e a proteína  $\alpha$  é liberada na eluição.

A amostra de proteína foi incubada a 4°C *overnight*. Em seguida, cromatografia de afinidade em coluna de Ni-HisTrap foi realizada e a proteína de interesse encontrava-se no volume de injeção. Por outro lado, SUMO protease e SUMO e cauda poli-His encontram-se ligados à coluna. Logo após, HiTrap Heparin (5mL, Ge Healthcare®) foi utilizada como o passo seguinte para a purificação. (Figura 15) e a proteína foi eluída com 400mM NaCl. Em seguida, eletroforese SDS-PAGE das frações eluídas confirmou a presença da proteína de interesse. Em seguida, essas frações foram concentradas para o volume de 500 $\mu$ L utilizando o sistema Amicon (MW *cut-off* 30kDa) e submetidas a cromatografia de exclusão de tamanho ou gel filtração utilizando coluna Superdex 75/100 HiLoad (24mL, GE HealthCare®). Na técnica de gel filtração é possível inferir o estado

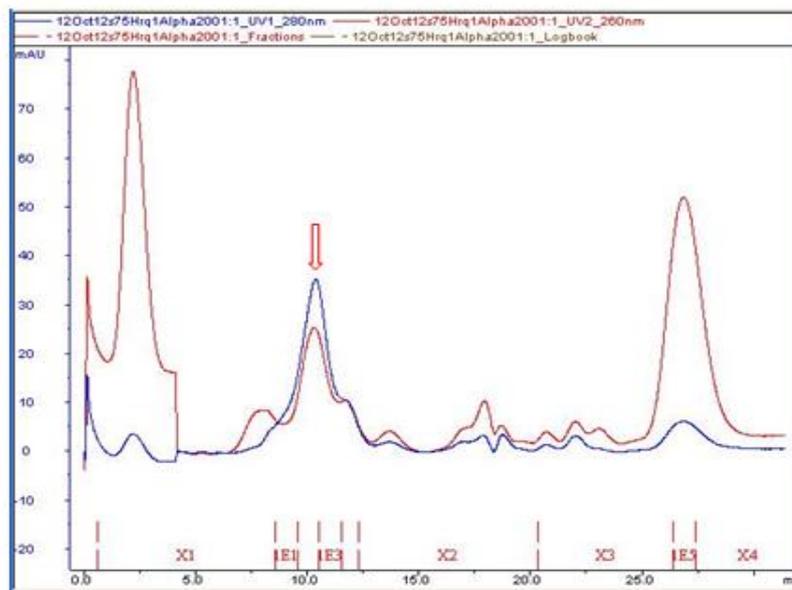
oligomérico da proteína a partir do volume de eluição a partir de um padrão fornecido pelo fabricante. Para o domínio  $\alpha$ , o volume de eluição é correspondente a uma proteína de aproximadamente 50kDa, sugerindo que este domínio é um dímero em solução (Figura 16). Entretanto, ao concentrar a proteína para os testes de *screening* das condições de cristalização, houve proteólise do fragmento e precipitação. Provavelmente, a ausência de estrutura secundária, como predito pelo programa HHPred (Figura 08), é responsável pela instabilidade do fragmento e sua susceptibilidade à degradação. Portanto, o domínio  $\alpha$  não foi cristalizado.



**Figura 14–** Perfil de eluição do domínio  $\alpha$  em coluna Ni-HisTrap (5mL, GEHealthCare®)



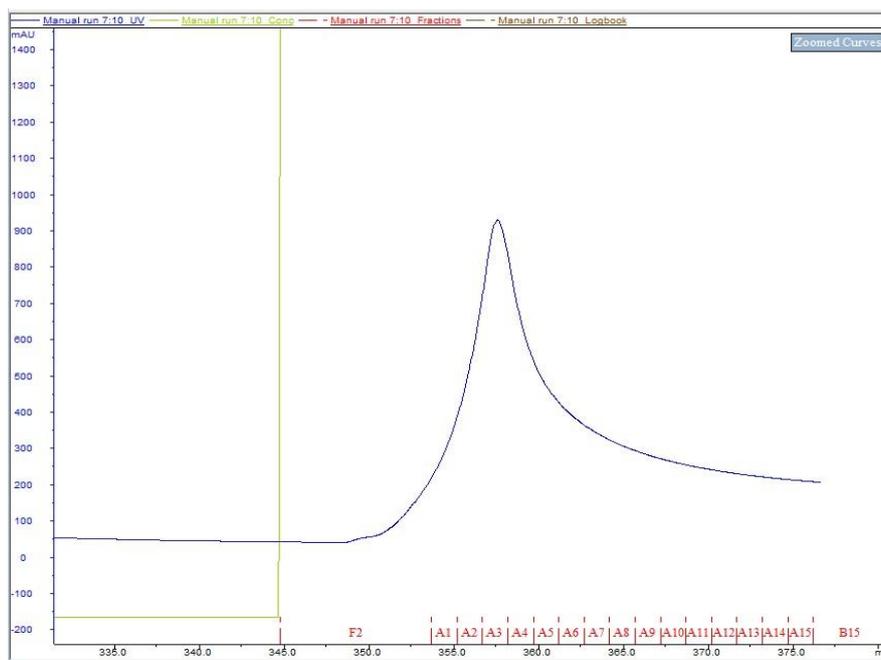
**Figura 15** – Perfil de eluição do domínio  $\alpha$  na primeira cromatografia de gel filtração com coluna Superdex 75/100 HiLoad (24mL, GEHealthcare®).



**Figura 16** – Perfil de eluição do domínio  $\alpha$  na segunda em cromatografia de gel filtração com coluna Superdex 75/100 HiLoad (24mL, GEHealthcare®).

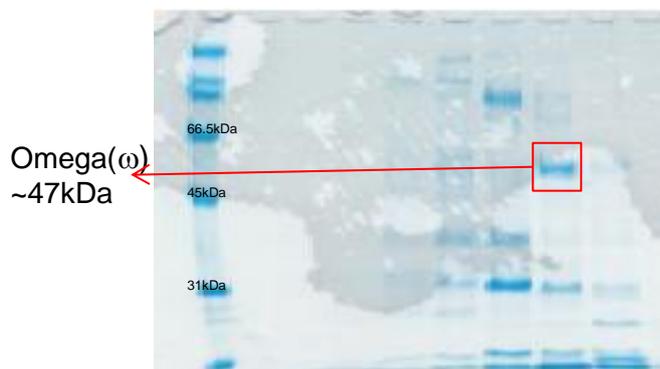
5.4.2. *Purificação do domínio  $\omega$*  - As primeiras etapas de purificação para o domínio  $\omega$ , foram semelhantes as utilizadas para o domínio  $\alpha$ , sendo 1) Coluna de Ni-HisTrap (Figura 17); 2) Clivagem de SUMO/His tag; 3) Coluna Ni-HisTrap; tendo em vista

que este domínio também possui SUMO e cauda poli-His. A diferença é que para o domínio  $\omega$  a etapa de clivagem que foi realizada por 3 horas à 4°C ao invés de *overnight* com o intuito de reduzir proteólise. Além disso, uma etapa de cromatografia de afinidade em HiTrap Heparin (5mLGE Helathcare®) foi realizada para este domínio. A proteína apresentou alta afinidade pela coluna, sendo eluída na concentração de 700mM NaCl. A eletroforese em SDS-PAGE confirmou a presença da proteína. (Figura 18). A amostra foi concentrada no sistema Amicon (MW *cutoff* 30kDa) e injetada em uma coluna de gel filtração Superdex 75/100 HiLoad (24mL, GE Helathcare®) (Figura 19). O volume de eluição sugere que a proteína encontra-se em estado monomérico (~47kDa) em solução.

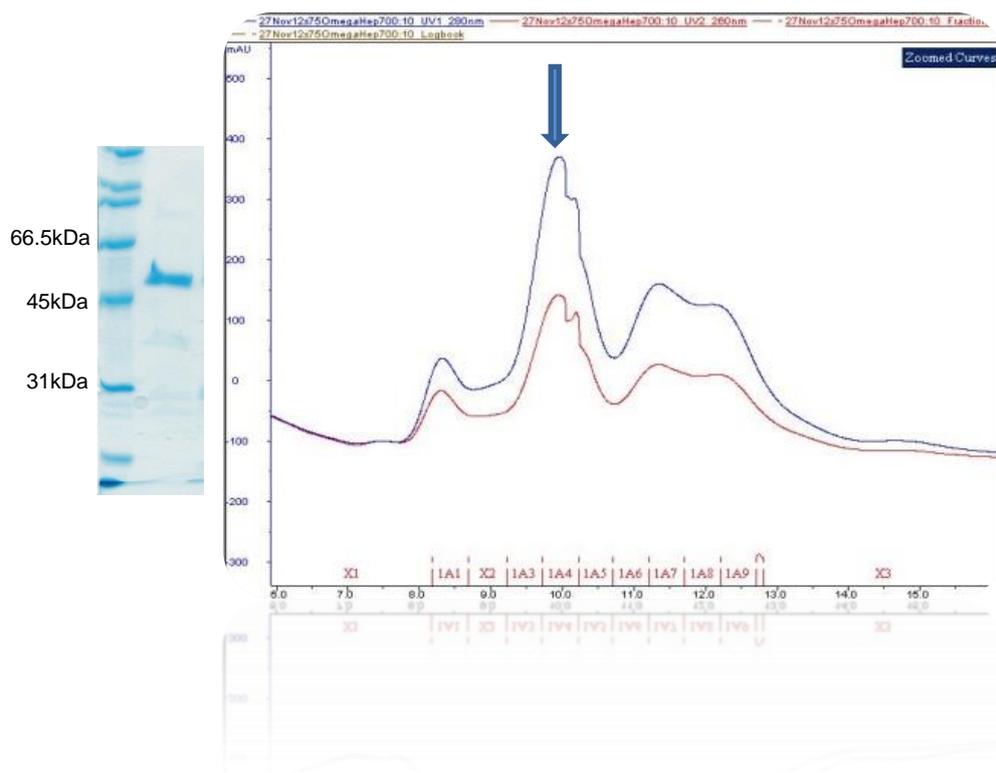


**Figura 17**– Perfil de eluição do domínio  $\omega$  em coluna Ni-HisTrap (5mL, GEHelathCare®)

100 200 300 400 500 600 700 800



**Figura 18** – Frações de eluição obtidas a partir da etapa de purificação do domínio  $\omega$  em coluna de Heparina HiTrap Heparin (5mL, GEHealthCare®) com gradiente manual (100mM-800mM) de NaCl. A concentração de NaCl de cada fração é apresentada sobre cada linha (em mM). A caixa vermelha indica a proteína  $\omega$  de interesse, a fração corresponde a uma concentração de 700mM de NaCl;



**Figura 19** - Perfil de eluição da construção  $\omega$  em cromatografia de exclusão de tamanho ou gel filtração com coluna Superdex 75/100 HiLoad (24mL, GEHealthcare®). À esquerda, um gel SDS-PAGE da fração indicada no cromatograma mostra o grau de pureza.

### 5.5.Determinação das condições de cristalização do domínio $\omega$

O sistema Mosquito® LCP foi utilizado para fazer as placas de *screening* de condições de cristalização a partir de uma solução de 6.5mg/mL de proteína pura do domínio  $\omega$ . (Figura 20) As placas continham 200  $\mu$ L de solução do reservatório as gotas foram feitas com 0,1  $\mu$ L de solução proteína e 0,1  $\mu$ L de solução de reservatório. O método utilizado para o *screening* foi o método gota sentada. Todas as placas foram incubadas a 20°C. A tabela 06 mostra os kits de condições de cristalização utilizadas nesse experimento.

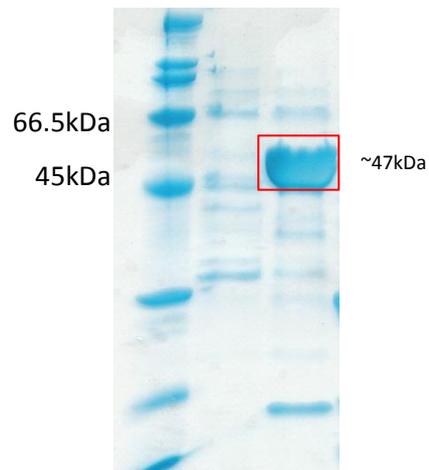
<b>Placa de Screening</b>	<b>Companhia</b>
PACT suíte	QIAGEN
Classic II suíte	QIAGEN
JSCG suíte	QIAGEN
Index	Hampton
PEG Ion 1	Hampton
pH Clear II	Hampton

**Tabela 06** – Condições utilizadas para o *screening* das condições de cristalização para o domínio  $\omega$ .

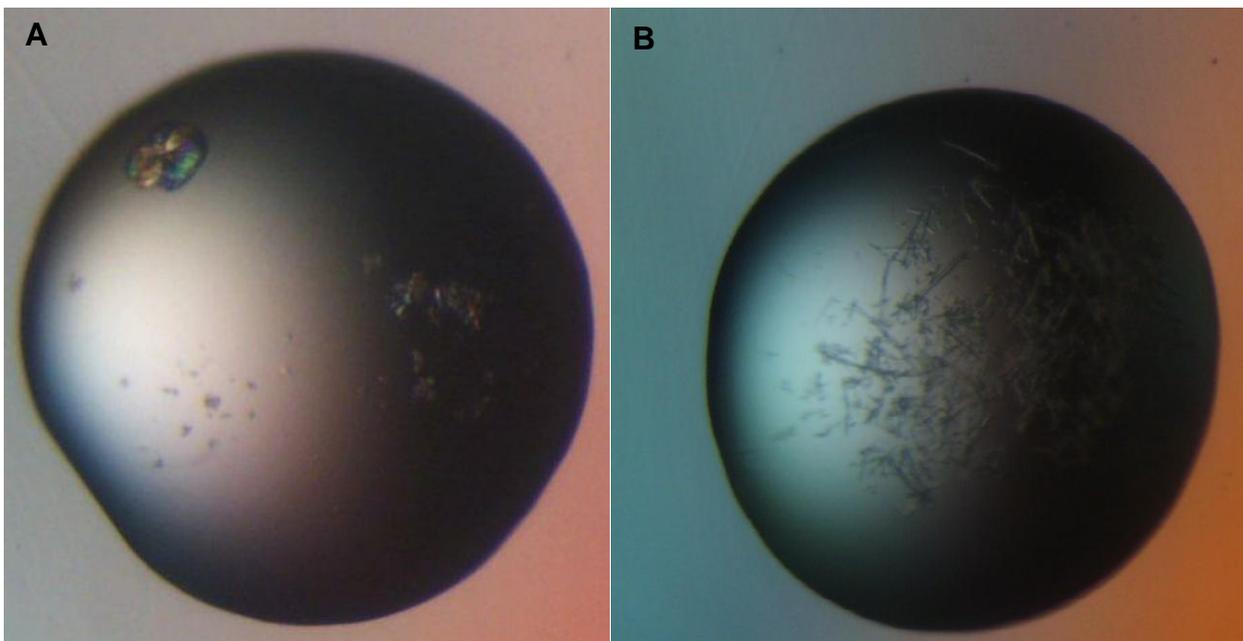
Cristais foram obtidos nas condições: **I)** 0,2M Acetato de Calcio, 20% w/v PEG 3350 e **II)** 0,2M Cloreto de Calcio dihidratado, 0,02M Cloreto de Cádmio hidratado, 0,02 Cloreto de Cobalto (II) hexahidratado, 20% w/v PEG 3350. Como observado na figura 21,

os cristais formados na condição **I**) não apresentavam forma definida enquanto que os cristais formados na condição **II**) apresentam forma de agulha ou *needle* (Figura 21).

O domínio  $\alpha$  , por sua vez, não foi cristalizado pelo fato



**Figura 20** – Perfil de eletroforese em gel SDS-PAGE do domínio  $\omega$  após a purificação e concentração da solução. A banda delimitada pela caixa vermelha corresponde ao domínio de interesse.



**Figura 21** – Condições de cristalização do domínio  $\omega$ . **A)** Cristais formados na condição **I)** 0,2M Acetato de Calcio, 20% w/v PEG 3350. Como é possível observar, os cristais são amorfos; **B)** Cristais formados na condição **II)** 0,2M Cloreto de Cálcio dihidratado, 0.02M Cloreto de Cádmiio hidratado, 0,02 Cloreto de Cobalto (II) hexahidratado, 20% w/v PEG 3350 – *needles*.

## 6. CONCLUSÕES

Os protocolos de indução e purificação dos domínios  $\alpha$  e  $\omega$  na forma solúvel foram estabelecidos com sucesso. Além disso, algumas condições preliminares de cristalização foram obtidas para o domínio  $\omega$  e estão sendo otimizadas.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- Ahn, B., Harrigan, J. A., Indig, F. E., Wilson, D. M., Bohr, V. A., (2004) **J.Biol.Chem.** v.279p.53465-53474
- Bachrati, C. Z., & Hickson, I. D., (2003) **Biochem. J.** v.374p.577-606
- Bachrati, C. Z., & Hickson, I. D., (2008) **Chromosoma** v.117p.219-233
- Bahr, A., de Graeve, F., Kedinger, C., Chatton, B., (1998) **Oncogene** v.17p.2565-2571
- Barea, F., Tessaro, S., Bonatto, D., (2008) **Comp. Biol. Chem.** v.32p.349-358
- Bernstein, D. A., Zittel, M. C., Keck, J. L., (2003) **EMBO Journal** v.22p.4910-4921
- Bohr, V. A., (2008) **Trends Bioch. Science** v.33p.609-620
- Cheng, W. H., von Kobbe, C., Opresko, P. L., et. al. (2004) **J.Biol.Chem.** v.279p.21169-21176
- Chu, W. K., & Hickson, I. D., (2009) **Nature Rev. Cancer** v.9p.644-654
- Crabbe, L., Jauch A., Naeger, C. M., Holtgreve-Grez, H., Karlseder, J., (2007) **PNAS** v.24p.8437-8446
- De Lange, T., (2002) **Oncogene** v.21p.532-540
- Ellis, N. A., Groden, J., Ye, T. Z., Straughen, J. Lennon, D. J., Ciocci, S., Proytcheva, M., German, J., (1995) **Cell** v.83p.655-666
- Frei, C., & Gasser, S. M., (2000) **J. Cell Science** v.113p.2641-2646
- Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L., Rothstein, R., (1991). **Mol. Cell. Biol.** v.14p.8391-8398
- Garcia, P. L., Liu, Y., Jiricny, J., West, S. C., Janscak, P., Hickson, I. D., (2005) **EMBO Journal** v.23p.2882-2891
- German, J., (1993). **Medicine.** v.72p.393-406
- Goto, M., Miller R. W., Ishikawa, Y., Sugano, H., (1996). **Cancer Epidemiol. Biomarkers Prev.** v.5p.239-246
- Grant, S., Wenger, S., Latimer, J., Thull D., Burke, L., (2000) **Genetics** v.58p.209-215
- Griffith, J. D., Comeau, L., Rosenfield, S., et. al. (1999). **Cell** v.97p.503-514
- Grocock, L. M., Prudden, J., Perry, J. J. P., Boddy, M. N., (2012) **J. Mol. Cell Biol.** *in press*
- Guo, R. B., Rigolet, P., Zargarian, L., Fermandjian, S., Xi, X. G., (2005) **Nucleic Acids Research** v.33p.3109-3124

- Harrigan, J. A., Wilson, D. M., Prasad, R., et. al. (2006) **Nucleic Acids Research** v.34p.745-754
- Heyer, W-D, Ehmsen K. T., Liu, J., (2010) **Annu. Rev. Genet.** V.44p.113-139
- Hickson, I. D., (2003) **Nat. Rev. Cancer** v.3p.169-178
- Imbert, A., Chaffanet, M., Essioux, L., Noguchi, T., Adelaide, J., Kerangueven, F., Le Paslier, D., Bonaiti-Pellie, C., Sobol, H., Birnbau, D., Pebusque, M. J., (1996). **Genomics.** v.32p.29-38
- Kerr, B., Aschcroft, G. S., Scott, D., Horan, M. A., Ferguson, M. W., Donnai, D., (1996) **J. Med. Genet.** V.33-p.928-934
- Killoran, M. P., & Keck, J. L., (2006) **J.Biol.Chem.** v.281p.12849-12857
- Kitao, S., Shimamoto, A., Goto, M., Miller, R. W., Smithson, W. A., Lindor, N. M., Furuichi, Y. (1999). **Nature Genetics.** v.22p82-84
- Lee, Y. J., & Yang, W., (2006) **Cell** v.127p.1349-1360
- Liu, J. L., Rigolet, S. X., Dou, P. Y., Wang, P. Y., Xi, X. G., (2004) **J.Biol.Chem.** v.279p.42794-42802
- Liu, Z., Macias, M. J., Bottomley, M. J., Stier, G., Linge, J.P., Nilges, M., Bork P., Sattler, M., (1999) **Structure** v.7p.1557-1566
- Lohman, T. M., Tomko, E. J., Wu, C. G., (2008) **Nature. Rev. Mol. Cell Biol.** v.9p.391-401
- Maire, G., Yoshimoto, M., Chilton-MacNeill, S., et. al (2009) **Neoplasia** v.11p.280-286
- Matson, S. W., Bean, D. W., George, J. W., (1994) **Bioessays** v16p.13-22
- Monnat Jr., (2001) **Scientific Soc. Press.** p.83-94
- Monnat Jr., (2002) in Fletcher, C., Unni, K., Mertens, F., editors. **Monograph** p.273-274
- O'Sullivan, R. J., & Karlseder, J., (2010) **Nature Rev. Mol. Cell Biol.** v.11p.171-181
- Opresko, P. L., (2008) **Mech. Ageing Dev.** V.129p.79-90
- Pike *et. al.* (2009) **PNAS.** V.106: p.1039-1044
- Prince P. R., Emond, M. J., Monnat Jr, R. J., (2001) **Genes Dev.** v.15p.933-938
- Saintigny Y., Makienko, K., Swanson, C., Emond, M. J., Monnat Jr, R. J., (2002) **Mol. Cell Biol.** v.22p.6971-6978
- Salk, D., Au, K., Hoehn, H., Martin, G. M., (1985) **Adv. Exp. Med. Biol.** v.190p.541-546

- Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmuller, L., Lautwein, A., et al. (1997). **Science** v.277p.333–38
- Sharma, S., Stumpo, D. J., Balajee, A. S., et. al . (2007) **Mol. Cell Biol.** v.27p.1784-1794
- Siitonen H. A., Sotkasiira, J., Biervliet, M., et. al. (2009). **European Journal Hum. Genet.** V.17p.151-160
- Singleton M.R., Wigley D.B., (2002). **J. Bacteriol.** v.184:p.1819–26
- Singleton, M. R., Dillingham, M. S., Wigley, D. B. (2007). **Annu. Rev. Biochem.** v.76: p.23–50
- Studier F. W (2005) **Protein Expr Purif.** v.41(1)p.207-234
- Svendsen, J. M. & Harper, J. W., (2010) **Genes Dev.** V.24p.521-536
- Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. (2010) **J Struct Biol.** v.172(1) p.33:44
- Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., Wigley, D. B., (1999) **Cell** v.97p.75-84
- Von Hippel, P. H., (2004) **Nature Struct. Mol. Biol.** v.11p.494-496
- Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., (1982) **EMBO J.** v.1p.945–51
- Wang, L. L., Levy, M. L., Lewis, R. A. et.al. (2001) **Am. Journal Hum. Genet.** V.102p.11-16
- Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., Losick, R., (2006) *Molecular biology of the gene.* 5a ed. Editor Artmed.
- Watt, P. M., Louis, E. J., Borts, R. H., Hickson, I.D., (1995) **Cell.** v.81p.253-260
- Wu L., Bachrati, C. Z., Ou, J., et. al (2006) **PNAS** v.103p.4068-4073
- Xu X., Rochette P. J., Feyissa E. A., Su, T. V., Liu, Y., (2009) **EMBO Journal** v.28.p.3005-3014
- Ye. J., Osborne A. R., Groll, M., Rapoport T. A., (2004). **Biochim. Biophys. Acta** v.1659p1–18
- Yu, C. E., Oshima, J., Fu, Y. H., Wijisman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan J., Schellenberg, G. D., (1996). **Science.** v.272p.258-262

## 8. APÊNDICE

Domain (residues)	Symbol		Bp length	Aa length	MW	Theoretical pl	Ext.Coeff.
Full length 1 - 1031	$\delta$	<b>delta</b>	3093	1031	118.20/ <b>132.61</b>	6,33	83660/ <b>85150</b>
Full length 1 - 1077	$\Delta$	<b>DELTA</b>	3231	1077	123.54/ <b>136.60</b>	6,22	83660/ <b>85150</b>
Cdt1 1-191	$\alpha$	<b>alpha</b>	573	191	22.4358/ <b>35.6345</b>	8.66	11460/ <b>12950</b>
Cdt1 1-277	$\beta$	<b>beta</b>	831	277	32.4332/ <b>45.632</b>	8.43	18910/ <b>20400</b>
Helicase 268-645	$\gamma$	<b>gamma</b>	1119	373	42.4/ <b>56.1251</b>	8.09	30370/ <b>31860</b>
Helicase 268-663	$\epsilon$	<b>epsilon</b>	1188	396	44.96/ <b>58.3155</b>	6.79	31860/ <b>33350</b>
Helicase 277-645	$\eta$	<b>eta</b>	1092	364	41.45/ <b>55.1129</b>	7.43	28880/ <b>30375</b>
Helicase 277-663	$\theta$	<b>theta</b>	1161	387	43.95/ <b>57303.3</b>	6.55	30370/ <b>31860</b>
Helicase+Cterminal 268-1031	$\lambda$	<b>lambda</b>	2292	764	86.92/ <b>101.33</b>	6.17	66240/ <b>67730</b>
Helicase+Cterminal 268-1077	$\mu$	<b>mi</b>	2430	810	92.27/ <b>106.68</b>	6.07	66240/ <b>67730</b>
Helicase+Cterminal 277-1031	$\pi$	<b>pi</b>	2265	755	85.91/ <b>100.32</b>	6.08	64750/ <b>66240</b>
Helicase+Cterminal 277-1077	$\sigma$	<b>sigma</b>	2403	801	91.2/ <b>105.67</b>	5.99	64750/ <b>66240</b>
Cterminal 644-1031	$\phi$	<b>phi</b>	1164	388	44.22/ <b>58.63</b>	5.43	35870/ <b>37360</b>
Cterminal 644-1077	$\chi$	<b>chi</b>	1302	434	49.56/ <b>63.97</b>	5.38	35870/ <b>37360</b>
Cterminal 668-1031	$\varphi$	<b>psi</b>	1092	364	41.44/ <b>55.85</b>	5.69	32890/ <b>34380</b>
Cterminal 668-1077	$\omega$	<b>omega</b>	1230	410	46.79/ <b>61.21</b>	5.59	32890/ <b>34380</b>
HRQ1+SUMOHis			3579	1193	136,5911	6,16	85150
SUMOHis				110	14,4188	5,86	1490

**Apêndice 01** – A tabela mostra as clonagens da proteína Hrql realizadas. Os símbolos gregos representam os nomes dados às construções. Bp length – Tamanho dos fragmentos clonados. Aa length – Tamanho da sequência de aminoácidos. MW – peso molecular das proteínas de cada fragmento. Theoretical pl – pl teórico das proteínas de cada clone. Ext. Coeff. – Coeficiente de extensão das proteínas. Os dados em negrito representam dados provenientes de cada proteína ligados à cauda SUMOHis.

### Primers para clonagem em pET-SUMO/CAT – RF cloning

**F1** : GAGGCTCACAGAGAACAGATTGGTGGTATGGAGGAAGGACCTATC

**Tamanho:** 45  
**GC%:** 51.1 %  
**Temperatura MELTING:** 67.2 °C

**F268** : GAGGCTCACAGAGAACAGATTGGTGGTACAGCGAAGTATAAAGG

**Tamanho:** 44

**GC%:** 47.7 %  
**Temperatura MELTING:** 66.0 °C

F277 : GAGGCTCACAGAGAACAGATTGGTGGT GAGCTTGCACCGGAAG

**Tamanho:** 43  
**GC%:** 55.8 %  
**Temperatura MELTING:** 69.7 °C

F644 : GAGGCTCACAGAGAACAGATTGGTGGT GACTCGCCTGTTGATCAAC

**Tamanho:** 46  
**GC%:** 52.2 %  
**Temperatura MELTING:** 68.7 °C

F668 : GAGGCTCACAGAGAACAGATTGGTGGT CAAGATTTGGTTCTTG

**Tamanho:** 43  
**GC%:** 46.5 %  
**Temperatura MELTING:** 65.9 °C

R191 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA GTCAAGGTTAGCCTTTTCC

**Tamanho:** 52  
**GC%:** 42.3 %  
**Temperatura MELTING:** 67.1 °C

R277 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA CTCAAACAGAGACC

**Tamanho:** 48  
**GC%:** 41.7 %  
**Temperatura MELTING:** 66.0 °C

R645 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA CGAGTCACTTGCAACCACG

**Tamanho:** 52  
**GC%:** 46.2 %  
**Temperatura MELTING:** 68.6 °C

R663 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA ATTATTAACCTCCAGCAAGG

**Tamanho:** 53

<b>GC%:</b>	39.6 %
<b>Temperatura MELTING:</b>	66.3 °C

R1031 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA TGGCTCTGGACCATCTTTG

<b>Tamanho:</b>	52
<b>GC%:</b>	44.2 %
<b>Temperatura MELTING:</b>	67.8 °C

R1077 : TTTGCGCCGAATAAATACCTAAGCTTGTCT TCA TATCTCTTTTTTAATG

<b>Tamanho:</b>	49
<b>GC%:</b>	32.7 %
<b>Temperatura MELTING:</b>	62.8 °C

**APÊNDICE 02** – Lista dos *primers* utilizados neste trabalho para a clonagem dos domínios da proteína Hrq1 de *S.cerevisiae* no vetor pET-SUMO/CAT. Os *primers* com a letra *F* representam *Forward* e *R* representam *Reverse*. Os números representam o resíduo de início (no caso do *primer F*) e do final (no caso do *primer R*). A porção vermelha representa a porção complementar ao vetor pET-SUMO/CAT. Os três nucleotídeos em azul representam a o códon de parada e a porção em preta representa a porção complementar ao gene *hrq1*.