

Universidade Estadual de Ponta Grossa
Programa de Pós-Graduação em Ciências
Área de Concentração: Física

Wellington Claiton Leite

Caracterização Funcional e Determinação da Estrutura
Tridimensional por Cristalografia de Raios X da
Proteína RecA de *Herbaspirillum seropedicae*

Ponta Grossa
2016

Wellington Claiton Leite

Caracterização Funcional e Determinação da Estrutura Tridimensional por Cristalografia de Raios X da Proteína RecA de *Herbaspirillum seropedicae*

Tese submetida ao programa de Pós Graduação em Ciências, Área de Concentração: Física, da Universidade Estadual de Ponta Grossa, como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências/Física.

Orientador: Prof. Dr. Sérgio da Costa Saab

Ponta Grossa

2016

Ficha Catalográfica
Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG

L533 Leite, Wellington Claiton
 Caracterização funcional e determinação
 da estrutura tridimensional por
 cristalografia de raios X da proteína recA
 de *Herbaspirillum seropedicae*/ Wellington
 Claiton Leite. Ponta Grossa, 2016.
 131f.

 Tese (Doutorado em Ciências - Área de
 Concentração: Física), Universidade
 Estadual de Ponta Grossa.

 Orientador: Prof. Dr. Sérgio da Costa
 Saab.

 1.RecA. 2.Cristalografia. 3.Filamento
 de nucleoproteína RecA. 4.Atividade
 ATPásica. 5.Reação de troca de fitas.
 I.Saab, Sérgio da Costa. II. Universidade
 Estadual de Ponta Grossa. Doutorado em
 Ciências. III. T.

CDD: 548.8


TERMO DE APROVAÇÃO


WELLINGTON CLAITON LEITE

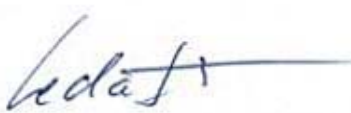
"Caracterização Funcional e Determinação da Estrutura Tridimensional por Cristalografia de Raios X da Proteína RecA de *Herbaspirillum seropedicae*"

Tese aprovada como requisito parcial para obtenção do grau de Doutor no Programa de Pós-Graduação em Ciências - Física da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora.


Orientador:


Prof. Dr. Sérgio da Costa Saab
Departamento de Física – UEPG/PR


Profa. Dra. Elaine Benelli
Universidade Federal do Paraná - UFPR


Profa. Dra. Leda Satie Chubatsu
Universidade Federal do Paraná - UFPR


Prof. Dr. André Maurício Brinatti
Departamento de Física – UEPG/PR


Prof. Dr. André Vitor Chaves de Andrade
Departamento de Física – UEPG/PR

Ponta Grossa, 06 de setembro de 2016

Dedico Para toda minha família, em especial minha mãe Neide Piedade Venâncio Martins, cujos esforços foram fundamentais para a realização de mais este sonho.

Agradecimentos

- Aos meus orientadores Prof. Dr. Sérgio da Costa Saab e Prof(a). Dr(a). Carolina Weigert Galvão, por todo incentivo, apoio, confiança e dedicação. Por sempre terem me possibilitado fazer minhas próprias escolhas, possibilitando meu crescimento pessoal e profissional. Com certeza em muitos momentos foram um pai e uma mãe. Reitero meu grande agradecimento à professora Carolina, mesmo não sendo orientadora oficial, abriu as portas do seu laboratório para minha pesquisa, sempre me incentivou a continuar, me apresentou um mundo novo na área da pesquisa, fica aqui minha admiração por toda sua inteligência e humildade em dividir seus conhecimentos.
- Ao professor Rafael Mazzer Etto, seus ensinamentos com certeza foram de grande valia. Obrigado por toda a sua paciência, muitas das idéias do projeto amadureceram após nossas conversas. Obrigado também pelos conhecimentos transmitidos, principalmente de Bioquímica.
- Ao professor Dr. Jorge Iulek, pela pronta disposição em ajudar, disponibilizando seu laboratório e facilidades. Obrigado também pelas discussões e conhecimentos transmitidos, principalmente de Cristalografia de proteínas.
- Aos Professores Dr. André Vitor Chavez de Andrade, Dr. André Maurício Brinatti, Dr. Luiz Fernando Pires e Dr. Fábio Augusto de Meira Cássaro, pelas discussões e ideias para o trabalho.
- Aos amigos do Lab. de Física Aplicada a Solos e Ciências Ambientais-UEPG, pelos bons momentos vividos! Todos foram ótimas companhias durante todo esse tempo.
- Aos membros e ao Lab. de Purificação e Estruturas de Proteínas-UEPG.
- Ao membros e ao Lab. de Biologia Molecular Microbiana (LabMOL)-UEPG.

- À duas grandes amizades que nasceu nesse tempo no LabMOL, Precila Zambotto e Gabrielle Silveira de Campos, obrigado por serem verdadeiros anjos que Deus colocou na minha vida.
- Aos supervisores do doutorado sanduíche no exterior Prof. Dr. Michael Mike Cox e Prof. e Dr. James L. Keck, ambos da Universidade De Wisconsin-Madison.
- Agradeço a todos que de alguma forma contribuíram para esse trabalho, espero que Deus possa lhes dar infinitamente mais do que me foi dado.
- Agradeço ao grande amigo Christiano Wetzel por todo apoio ao longo desse trabalho, considero como se fosse um irmão de sangue, que Deus lhe retribua infinitamente.
- Ao Laboratório Nacional de Luz Síncrotron, por disponibilizar suas facilidades para medidas de difração de raios X.
- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado e de doutorado sanduíche (Bolsista da CAPES - Proc. n BEX 12380 13-0).

Resumo

A proteína RecA bacteriana desempenha um importante papel no complexo sistema de reparo de danos ao DNA. Na presença de ATP, a proteína RecA se auto-polimeriza sobre o DNA simples fita (ssDNA) (do inglês *single-strand* DNA (ssDNA)) como um filamento de nucleoproteína helicoidal, cataliza a reação de troca de fitas entre as moléculas ssDNA e a fita de DNA dupla fita homóloga (dsDNA) (do inglês *double-strand DNA* (dsDNA)). Estas atividades são suportadas ou estimuladas por proteínas acessórias, como a proteína ligadora de ssDNA SSB (do inglês *single-stranded binding protein* (SSB)). Neste trabalho é apresentado a caracterização estrutural e funcional da proteína RecA da bactéria *Herbaspirillum seropedicae*. A estrutura tridimensional do complexo HsRecA-ADP/ATP foi resolvida numa resolução 1,7 Å. A estrutura monomérica da proteína HsRecA consiste em um pequeno domínio N-terminal, um domínio central contendo um sítio ATPásico e e um grande domínio C-terminal, similar com proteínas RecAs homólogas. Análises estruturais comparativas mostraram que o motivo de polimerização da região N-terminal de proteínas da família RecA que incluem archaea e eucariotos, também está presente na proteína RecA bacteriana. O motivo de polimerização da região N-terminal de bactérias contém a sequência de resíduos (Serina, Valina ou Isoleucina, Metionina, Arginina ou Lisina, Leucina, Glicina) que interage com a sequência de resíduos do core ATPásico (Aspartato, Asparagina, Leucina, Leucina, Leucina, Valina, Cisteína, Serina). Na proteína RecA inativa esta interação é do tipo β loop - β strand, respectivamente, enquanto na proteína RecA ativa essa interação se torna uma dupla β -strand. Em ambas formas da RecA, o motivo de polimerização parece estabilizar a interface subunidade-subunidade por interações hidrofóbicas. No motivo N-terminal a presença de uma Metionina altamente conservada talvez desempenha um importante papel na estabilidade e formação do filamento de nucleoproteína. A atividade ATPásica e a estrutura do filamento de nucleoproteína da proteína HsRecA e da *Escherichia coli* RecA (EcRecA) foram analisadas na presença e ausência da proteína SSB. Quando a SSB foi adicionada após RecA+ssDNA, as proteínas HsRecA e EcRecA mostraram similar atividade ATPásica e estrutura de nucleofilamento. Entretanto, quando a SSB não estava incluída ou quando adicionada anteriormente a adição RecA+ssDNA, a proteína HsRecA mostrou maior atividade ATPásica e formou maiores filamentos de nucleoproteína que a proteína EcRecA. Ainda, a proteína HsRecA é mais eficiente em deslocar a SSB do ssDNA que a proteína EcRecA. A proteína HsRecA também promove a reação de troca de fitas mais eficientemente: uma maior quantidade de produtos duplex substrato convertido em duplex circular foram obtidos em um curto intervalo de tempo. A reconstrução do potencial eletrostático da estrutura hexamérica da proteína HsRecA revelou uma maior densidade de cargas positivas no seu interior, que é consistente com o fato que o ssDNA ligar-se internamente ao filamento hexamérico. Isto talvez possa explicar capacidade melhorada da proteína HsRecA ligar-se ao ssDNA, formando um contínuo filamento de nucleoproteína, deslocando a SSB e ainda promovendo de forma eficiente a reação de trocas de fitas.

Palavras-chave: RecA, Cristalografia, filamento de nucleoproteína RecA, atividade ATPásica, reação de troca de fitas, estrutura cristalina, análise estrutural.

Abstract

The bacterial RecA protein plays a role in the complex system of DNA damage repair. In the presence of ATP, RecA proteins polymerize onto single-strand DNA (ssDNA) as right-handed helical nucleoprotein filaments, and catalyze strand exchange reaction between the ssDNA and homologous double-strand DNA (dsDNA) molecules. These activities are supported or stimulated by accessory proteins, as the single-stranded binding protein (SSB). Here, we report a functional and structural characterization of the *Herbaspirillum seropedicae* RecA protein (HsRecA). We report the crystal structure of HsRecA-ADP/ATP complex to 1.7 Å of atomic resolution. HsRecA protein contains a small N-terminal domain, a central core ATPase domain and a large C-terminal domain, similarly to homologous RecA proteins. Comparative structural analysis showed that the N-terminal polymerization motif of archaeal and eukaryotic RecA family proteins are also present in bacterial RecAs. The bacterial polymerization motif contains the sequence SV/IMR/KLG which interacts with the core ATPase domain residues DNLLLV/CS. In the inactive RecA, it is a β loop - β strand interaction, respectively, while in the active RecA it becomes a dyad β -strand. In both RecA forms, the polymerization motif seems to stabilize the subunit-subunit interface by hydrophobic interactions. The methionine of this motif may play an important role in the stability and formation of a right-handed helical nucleoprotein filament. The ATPase activity and the structure of the nucleoprotein filament of HsRecA and *Escherichia coli* RecA (EcRecA) were analyzed in the presence and absence of SSB. When SSB was added after RecA+ssDNA, HsRecA and EcRecA showed similar ATPase activity and nucleofilament structure. However, when SSB was either not included or it was added before RecA+ssDNA, the HsRecA showed higher ATPase activity and formed longer nucleoprotein filaments than EcRecA. Thus, HsRecA protein is more efficient at displacing SSB from ssDNA than EcRecA protein. HsRecA promoted DNA exchange more efficiently: a greater yield of nicked circular products were obtained in a shorter time. Reconstruction of electrostatic potential from the hexameric structure of HsRecA-ADP/ATP revealed a high positive charge along the inner side, which is consistent with the fact that ssDNA binds inside the filament. It may explain the enhance capacity of HsRecA protein to bind ssDNA, forming a contiguous nucleoprotein filament, displace SSB and promote efficiently the DNA strand exchange reaction.

Keywords: RecA, Crystallography, RecA nucleoprotein filament, ATPase activity, DNA strand exchange, crystal structure, structural analysis.

Lista de Figuras

- 2.1 A ilustração mostra as etapas do modelo de reparo via recombinação homóloga da DBS. Os principais processos biológicos estão indicados nas etapas de 1 - 4, de acordo com o texto. Neste modelo destaca-se a proteína RecA, responsável por catalisar a troca de fitas de DNA e assim efetuar o reparo da fita danificada. 32
- 2.2 Modelo estrutural do hexâmero da proteína EcRecA gerado por reconstruções de imagens de microscopia eletrônica. Na figura, é mostrado um filamento contendo três subsequentes hexâmeros. A posição de ligação do ADP está indicada na interface monômero-monômero. Cada monômero possui uma cor para diferenciação. 34
- 2.3 Diagrama de fitas do modelo cristalográfico da proteína RecA de *E. coli*, esse consiste de três domínios: uma região N-terminal, um domínio central (região sombreada) e um domínio C-terminal. Também está indicado a região do sítio de ligação do nucleotídeo ADP. 35
- 2.4 Alinhamento sequencial das proteínas de RecA de *H. seropedicae* (HsRecA), *M. tuberculosis* (MtRecA), *M. smegmatis* (MsRecA), *D. radiodurans* (DrRecA), (EcRecA) *E. coli*. 36
- 2.5 Diagrama de fitas do modelo cristalográfico do filamento pré-sináptico, representando o estado do filamento antes da reação de troca de fitas, da proteína RecA de *E. coli*. O modelo contém 6 monômeros ligados por um *loop* flexível. No filamento o ssDNA está ligado internamente ao filamento. Cada monômero está colorido de uma cor para diferenciação. 38

2.6	Diagrama de fitas do modelo cristalográfico do filamento pós-sináptico, produzido após a reação de trocas de fitas, da proteína RecA de <i>E. coli</i> . No filamento o dsDNA está ligado internamente ao filamento. O modelo contém 5 monômeros ligados por um <i>loop</i> flexível.	39
2.7	Polimerização da proteína RecA sobre o ssDNA. A RecA nucleia-se sobre o DNA como oligômeros de 4-6 monômeros, posteriormente a polimerização pode ocorrer a taxas maiores que 1000 monômeros/min a 37 °C. E por fim, ocorre a dissociação de um único monômero por vez. Ambas polimerização e dissociação são unidirecionais, ocorrendo na direção 5' → 3' do ssDNA. .	41
2.8	Sobreposição da topologia de seis proteínas tipo-RecA. Essas seis proteínas pertencem a diversas famílias de proteínas, incluindo a superfamília AAA protease FtsH (vermelho), ABC transporte MalK (azul), PilT família conjugativa de plasmídeo TrwB (verde), Superfamília da helicase I PcrA (amarelo), tipo-DnaB T7 helicase (azul ciano), e F1 ATPase (magenta). .	43
2.9	Modelo de facilitação de rotação do DNA. Em (a) é mostrado um esquema de rotação do DNA ao redor do filamento, no modelo de facilitação de rotação a proposição é que a reação de troca de fitas pode estar acoplada a hidrólise de ATP. Em (b) gráfico de Arrhenius para a taxa na qual ocorre a reação de troca de fitas (em unidades de $pbmin^{-1}$) e a taxa de hidrólise de ATP (em unidades de min^{-1}) durante a reação de troca de fitas. . . .	46
2.10	Diagrama de fitas da estrutura tetramérica da proteína SSB de <i>E. coli</i> . .	48
3.1	Procedimento realizado no método de substituição molecular para determinação das fases iniciais. Um modelo estrutural conhecido é transladado e rotacionado para a posição da proteína na qual a estrutura deve ser determinada.	55

4.1	Superexpressão da HsRecA a temperatura de 37 °C por 4 horas. MW - Padrão de massa molecular (Bio-Rad), (1) Fração não induzida e (2) fração induzida. A banda mais intensa do gel corresponde à proteína HsRecA, localizada próxima ao peso molecular de 37 kDa.	66
4.2	Lise celular. MW - Padrão de massa molecular (Bio-Rad), (1) Fração induzida, (2) precipitado da lise celular e (3) sobrenadante da lise celular . A banda mais intensa do gel corresponde à proteína HsRecA.	67
4.3	Análise das frações coletadas durante a cromatografia na coluna DEAE. MW - Padrão de massa molecular (Bio-Rad), (1-13) frações eluídas.	69
4.4	Análise das frações coletadas durante a cromatografia na coluna HAP. MW - Padrão de massa molecular (Bio-Rad), (1-12) frações eluídas.	71
4.5	Análise das frações coletadas durante a cromatografia nas colunas Source-15S e Source-15Q. MW - Padrão de massa molecular (Bio-Rad), as frações eluídas de cada coluna estão indicadas abaixo da identificação de cada coluna no gel.	72
4.6	Análise das frações coletadas durante a cromatografia na coluna Butil. MW - Padrão de massa molecular (Bio-Rad), (1-9) frações eluídas.	73
5.1	Cristal da proteína HsRecA obtidos pelo método da difusão de vapor em gota suspensa. Condição de cristalização: (a) 0,25 M $CaCl_2$ e 14% (w/v) polietileno glicol 3500 (PEG 3500), (b) 0,25 M $CaCl_2$ e 14% (w/v) polietileno glicol 3500 (PEG 3500) e 5% (w/v) de polipropileno glicol 400 (PPG 400).	83
5.2	Padrão de difração do cristal da HsRecA. A linha verde indica a região de mais alta resolução, que foi de 1,7 Å.	84

5.3	Gráfico de Ramachandran para o modelo estrutural refinado da proteína HsRecA. Em vermelho são as regiões permitidas, em amarelo adicionalmente permitidas, azul claro generosamente permitidas e azul escuro as proibidas. Os pontos brancos sobre o mapa referem-se a combinação de ângulos torcionais $\phi(^{\circ})$ e $\psi(^{\circ})$ dos 308 resíduos do modelo. Gráfico gerado pela validação no programa Molprobity.	90
5.4	Diagrama de fitas da estrutura cristalográfica do monômero da proteína HsRecA. Regiões que não puderam ser modeladas estão indicadas pelo último resíduo-número. O monômero é composto por um domínio N-terminal (NTD), um domínio central contendo um sítio ATPásico e um grande domínio C-terminal (CTD). No domínio central está presente um íon de Ca^{2+} (esfera magenta). O sítio ATPásico está parcialmente ocupado pelos nucleotídeos ATP e ADP.	91
5.5	Alinhamento da sequência de aminácidos presentes estrutura cristalográfica contra a sequência de aminoácidos das proteínas HsRecA e EcRecA. A assinatura da estrutura secundária está indicada acima e abaixo de cada sequência.	92
5.6	<i>Stereoview</i> do sítio de ligação do íon Ca^{2+} e dos resíduos que coordenam este íon. Os valores mostrados em verde se referem as distâncias interatômicas em Å do íon. As águas 141, 222 e 304 fazem a coordenação íon juntamente com os resíduos Asparagina 119 (Ans119) e Aspartato 120 (Asp120).	93
5.7	Mapas de densidade eletrônica (mFo-DFc) do sítio ATPásico: (a)Refinamento com ATP possuindo 100% de ocupação; a densidade eletrônica negativa sobre o fosfato- γ indica que esta posição não pode estar completamente ocupada. (b) Refinamento com o ATP possuindo 39% de ocupação; a densidade eletrônica positiva sobre o correspondente ADP indica que este deve ocupar a posição sobre a densidade mostrada. Os mapas estão contornados em $+3\sigma$ (verde) e -3σ (vermelho).	95

5.8	Interações dos nucleotídeos com aminoácidos do sítio ATPásico da HsRecA. Em (a) interações encontradas para o ADP, em (b) interações encontradas para o ATP.	96
5.9	Distribuição de potencial eletrostático de superfície na região do sítio ATPásico. Em (a) é mostrado o sítio ATPásico da proteína EcRecA, em (b) o sítio ATPásico da proteína HsRecA. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).	98
5.10	Reação de troca de fitas de DNA promovida pelas proteínas HsRecA e EcRecA. (A) Eletroforese em gel de agarose 0,8% do decurso temporal da reação de troca de três fitas, realizada como descrito no capítulo de Material e Método e abaixo, a ilustração dos substratos e produtos da reação de troca de fitas. A reação contendo $10 \mu\text{M}$ nt cssDNA M13mp18 e $3,5 \mu\text{M}$ HsRecA ou EcRecA, a reação foi incubada por 20 min a 37°C , e então $3 \mu\text{M}$ ATP e $1 \mu\text{M}$ SSB foram adicionados e incubados por outros adicionais 10 min. Os decurso temporal mostrado representa o tempo da reação após adição de $20 \mu\text{M}$ nt ldsDNA M13mp18. (B) A percentagem de duplex substrato convertido em duplex circular (Produto) em função do decurso temporal. .	100
5.11	Perfil da atividade ATPásica das proteínas HsRecA e EcRecA na presença e ausência da proteína SSB. Reação 1: $5 \mu\text{M}$ nt M13mp18 cssDNA e $3 \mu\text{M}$ HsRecA ou EcRecA foram previamente incubados por 20 min a 37°C , seguindo da adição de $3 \mu\text{M}$ ATP e $0,5 \mu\text{M}$ SSB ($t=0, \textit{start}$). Reação 2: Os mesmos procedimentos da Reação 1 foram realizados na ausência da proteína SSB. O tempo $t = 0$ indica a adição de somente ATP. O valor aparente do K_{cat} foi determinado após o estado estacionário da RecA ligada ao cssDNA.	103

5.12	Velocidade máxima de atividade ATPásica da proteína HsRecA em diferentes concentrações da proteína SSB. As reações contendo 5 μM nt M13mp18 cssDNA e 3 μM HsRecA, foram previamente incubadas por 20 min a 37 °C, seguido da adição de 3 μM ATP e diferentes concentrações da proteína SSB. Foi utilizado um polinômio de segunda ordem para indicar a tendência dos dados, essa representada pela linha tracejada.	104
5.13	Perfil de ATPase quando a SSB é adicionada antes da HsRecA ou EcRecA. Na reação 5 μM nt M13mp18 cssDNA, 3 μM ATP e 0,35 μM SSB foram previamente incubados por 10 min a 37 °C, seguindo da adição de EcRecA ou HsRecA (tempo $t = 0$)	105
5.14	Imagens de microscopia eletrônica dos filamentos das proteínas EcRecA e HsRecA na ausência e presença da proteína SSB. Em (a) filamentos de EcRecA-cssDNA na ausência da proteína SSB, (b) filamentos de HsRecA-cssDNA na ausência da proteína SSB, (C) filamentos de EcRecA-cssDNA na presença da proteína EcSSB adicionada após 20 min, (d) filamentos de HsRecA-cssDNA na presença da proteína EcSSB adicionada após 20 min. O sistema de reação contendo 6,7 μM HsRecA ou EcRecA e 20 μM cssDNA M13mp18 foi incubado a 37°C por 20 min, e então 2 μM SSB foi adicionada ou não. Após 10 min, os filamentos foram estabilizados com a adição de 3 μM ATP γ S, seguido de incubação por 3 min e depositados no grind de Alcian.	108
5.15	Comprimento médio de 10 filamentos das proteínas EcRecA e HsRecA sobre o cssDNA M13mp18, na presença e ausência da proteína SSB. As letras acima de cada barra se referem aos painéis A-D da figura (5.14).	109

5.16	Alinhamento sequencial e estrutural das proteínas de RecA de <i>H. seropedicae</i> (HsRecA), <i>E. coli</i> (EcRecA), <i>M. tuberculosis</i> (MtRecA), <i>M. smegmatis</i> (MsRecA), <i>D. radiodurans</i> (DrRecA), <i>N. gonorrhoeae</i> (NgRecA), <i>P. aeruginosa</i> (PaRecA), e da família RecA de <i>M. gonorrhoeae</i> (MvRadA), <i>P. furiosus</i> (RadA), <i>H. sapiens</i> (HsRad51), <i>P. furiosus</i> (Rad51), <i>S. cerevisiae</i> (ScRad51 e Dmc1).	112
5.17	Superposição da estrutura monomérica da proteína EcRecA inativa e pré-sináptica sobre a inativa da proteína HsRecA. As proteínas EcRecA e HsRecA estão coloridas em tons de cinza e ciano, respectivamente. (a) Superposição da estrutura monomérica inativa da proteína EcRecA (pdb:1xmv) sobre a proteína HsRecA. Na estrutura da EcRecA o íon Mg^{2+} interage com o N-MP, e na estrutura da proteína HsRecA o íon Ca^{2+} interage com Core-MP. (b) Superposição da estrutura monomérica pré-sináptica da proteína EcRecA (pdb:3cmu:A) sobre a proteína HsRecA. O ADP, ATP e ADP- AlF_4 e os íons Mg^{2+} e Ca^{2+} , o ssDNA ligado nas estruturas foram removidos para melhor clareza.	114
5.18	Distribuição de potencial eletrostático de superfície da proteína HsRecA. (a) visão lateral de dois subsequentes hexâmeros. (b) e (c) observação nas direções 5' e 3', relativo a direção axial do filamento, respectivamente. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).	116
5.19	Distribuição de potencial eletrostático na região dos motivos de polimerização N-MP e core-MP para a proteína HsRecA. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).	117
5.20	Dois consecutivos hexâmeros do cristal da proteína HsRecA. Na figura é mostrado o passo helicoidal do hexâmero.	118

Lista de Tabelas

5.1	Estatísticas de processamento e coleta de dados de raios X para o cristal da proteína HsRecA. Os valores entre parênteses se referem a faixa de maior resolução de 1,73 - 1,70 Å.	85
5.2	Valores de rotação, translação e os scores para os modelos utilizados pela substituição molecular na resolução da estrutura da proteína HsRecA. A estrutura da proteína EsRecA (pdb 1xmv), separada em dois modelos, foi utilizado como estrutura de referência para os cálculos das fases iniciais. Na tabela α° , β° e γ° , são os ângulos euclidianos da operação de rotação. T_x , T_y e T_z correspondem as respectivas operações de translações nos eixos x; y; z. RFZ e TFZ representam os scores da função de rotação e translação, respectivamente.	87
5.3	Estatísticas de processamento da coleta de dados da proteína HsRecA. Os valores entre parênteses se referem a faixa de resolução de 1,73 - 1,70 Å. . .	88
5.4	Estatísticas de validação geométrica do modelo estrutural da proteína Hs-RecA via Molprobit.	90

Lista de Abreviaturas e Símbolos

ssDNA	DNA simples fita (do inglês <i>single-strand</i> DNA) ;
dsDNA	DNA dupla fita (do inglês <i>double-strand</i> DNA);
cssDNA	DNA circular simples fita (do inglês <i>circular single-strand</i> DNA);
ldsDNA	DNA linear dupla fita (do inglês <i>linear double-strand</i> DNA);
cdsDNA	DNA circular dupla fita (do inglês <i>circular double-strand</i> DNA);
DBS	Dupla quebra de fita (do inglês <i>double-strand break</i>);
ATP γ S	Adenosina 5' - [γ -tio] trifosfato
ME	Microscopia eletrônica
ADP	Difosfato de adenosina
ATP	Trifosfato de adenosina
dATP	Trifosfato de desoxiadenosina
IPTG	Isopropilo β -D-1-tiogalactopiranosido
PEG	Polietileno glicol
PPG	Polipropileno glicol
DO	Densidade óptica
DTT	Ditiotreitol
pb	Pares de base
nt	Nucleotídeos
EcRecA e EcSSB	Proteínas RecA e SSB de <i>E. coli</i>
HsRecA	Proteína RecA de <i>Herbaspirillum seropedicae</i>
SDS-PAGE	Eletroforese em gel de poliacrilamida em condições desnaturantes
k_{cat}	Número de renovação

Sumário

Lista de Figuras	10
Lista de Tabelas	17
Lista de Abreviaturas e Símbolos	18
Sumário	19
1 Introdução	21
2 Estrutura e Função da Proteína RecA	29
2.1 Fontes de Lesões e Mecanismos de Reparo de DNA	29
2.2 A Proteína RecA	33
2.3 Proteínas Regulatórias da Atividade da Proteína RecA	46
3 Cristalografia de Proteínas	49
3.1 As Bases da Cristalografia de Macromoléculas	49
3.2 Determinação Experimental do Módulo do Fator de Estrutura $ F(h, k, l) $.	51
3.3 O Problema das Fases - $\alpha(h, k, l)$	52
3.4 Superfície de Potencial Eletrostático	60
4 Material e Métodos	62
4.1 Métodos Gerais para DNA	62
4.2 Purificação de Proteínas	65
4.3 Microscopia Eletrônica	75

4.4	Determinação da Estrutura Tridimensional da proteína HsRecA	77
5	Resultados e Discussões	82
5.1	Estrutura Tridimensional da Proteína HsRecA	82
5.2	Descrição Geral da Estrutura da Proteína HsRecA	91
5.3	O Sítio ATPásico	93
5.4	Reação de Troca de Fitas de DNA <i>in vitro</i>	98
5.5	A Dinâmica de Formação dos Filamentos sobre o ssDNA	101
5.6	Visualização dos filamentos por Microscopia Eletrônica	107
5.7	A Polimerização da Proteína RecA no Cristal	110
5.8	Passo helicoidal Hexamérico	118
6	Conclusão e Trabalhos Futuros	120
	Referências Bibliográficas	122

Capítulo 1

Introdução

A cristalografia de macromoléculas é utilizada para estudar sistemas biológicos, por meio da determinação da estrutura tridimensional em nível atômico de importantes classes de moléculas, como proteínas e ácidos nucleicos. A cristalografia de macromoléculas deu-se início com o trabalho de Kendrew e Perutz com a resolução da estrutura a baixa resolução da proteína mioglobina da baleia (KENDREW et al., 1958). Desde então, mais de cem mil novas estruturas foram resolvidas por meio da cristalografia de raios X e depositadas no banco de dados Protein Data Bank (<http://www.rcsb.org>) (BERMAN, 2000).

Informações sobre a estrutura tridimensional de uma macromolécula podem ser obtidas utilizando-se de várias técnicas: difração de raios X, difração de nêutrons, ressonância magnética nuclear (RMN), espalhamento de raios X a baixo ângulo (SAXS), microscopia eletrônica (ME), modelagem por homologia, modelagem teórica, entre outras. Dentre essas, a difração de raios X devido à alta resolução com que se pode descrever a densidade eletrônica dos átomos envolvidos têm sido umas das técnicas mais utilizadas.

Basicamente a cristalografia de macromoléculas consiste em construir um modelo estrutural no qual as posições dos átomos do modelo representam uma média espacial das posições dos átomos da macromolécula estudada. Uma vez que a estrutura tridimensional da macromolécula possui informação essencial para o entendimento detalhado de vários processos biológicos como catálise enzimática, infecção viral, divisão celular e metabolismo

de DNA, por exemplo, além de ter aplicação no planejamento racional de novos fármacos e vacinas (BRANDEN; TOOZE, 1999).

As proteínas são macromoléculas extremamente importantes. São responsáveis por diversas funções biológicas como catálise, transporte, regulação, proteção, entre outras, sendo portanto vitais para os seres vivos. Estruturalmente as proteínas são polímeros formados por aminoácidos, ligam-se por meio de ligação peptídica e apresentam uma estrutura tridimensional. As proteínas organizam-se em quatro níveis estruturais: estrutura primária (ligação covalente de aminoácidos formando uma cadeia polipeptídica), estrutura secundária (interação de aminoácidos vizinhos formando α -hélices, fitas- β e loops), estrutura terciária (arranjo tridimensional da cadeia polipeptídica devido interação de aminoácidos distantes na cadeia) e estrutura quaternária (interações entre subunidades, arranjo tridimensional de diferentes cadeias polipeptídicas). A demanda do conhecimento da estrutura tridimensional de uma proteína provém intimamente da relação entre estrutura e função (LEHNINGER; NELSON; COX, 2008).

Uma proteína que desperta interesse há mais de 3 décadas é a proteína RecA, essa é um protótipo de uma classe de proteínas encontradas em todos os organismos, desde bactérias até humanos. A proteína RecA possui importante função em um complexo sistema de reparo a danos ao DNA. Na presença de danos no DNA a proteína RecA polimeriza-se sobre o DNA simples-fita (do inglês *single-strand* DNA - (ssDNA)) como um filamento helicoidal e ao identificar o DNA dupla-fita (do inglês *double-strand* DNA - (dsDNA)) homólogo cataliza a reação de trocas de fitas e, assim promove o reparo do DNA pela via de recombinação homóloga (STASIAK; CAPUA, 1982; PUGH; COX, 1988a; KOWALCZYKOWSKI; EGGLESTON, 1994; COX, 2003; JANION, 2008).

Diversos estudos foram realizados para elucidar os mecanismos de formação dos filamentos da proteína RecA sobre o DNA (REGISTER; GRIFFITH, 1985; COX, 1997; GALLETTO et al., 2006; STASIAK; CAPUA, 1982; STASIAK; EGELMAN, 1986; EGELMAN; STASIAK, 1986; PUGH; COX, 1988a; STASIAK; EGELMAN; HOWARD-FLANDERS, 1988; CHANG et al., 1988; YU et al., 2004). Inicialmente, o filamento da proteína RecA-DNA foi caracterizado por imagens de microscopia de eletrônica (ME).

Observou-se que a proteína RecA liga-se ao DNA na proporção de 3 nucleotídeos (ssDNA) ou 3 pares de base (pb) (dsDNA) por monômero de RecA, formando um filamento hexamérico com passo helicoidal de 95 Å, possuindo 6,18 subunidades de RecA por volta helicoidal. Posteriormente, utilizando-se da mesma técnica observou-se que as subunidades se polimerizam do terminal 5' para o 3' no ssDNA, e também que o filamento apresenta um sulco profundo onde se encontra ligado o DNA. A ligação RecA-DNA é dependente da ligação conjunta de ATP. Estudos com ATP γ S revelaram um acréscimo na quantidade de monômeros de RecA ao DNA, e portanto, mostraram que a ligação RecA-DNA é dependente (ou otimizada) pela ligação conjunta de ATP (STASIAK; CAPUA, 1982; STASIAK; EGELMAN, 1986; EGELMAN; STASIAK, 1986; PUGH; COX, 1988a; STASIAK; EGELMAN; HOWARD-FLANDERS, 1988; CHANG et al., 1988; YU et al., 2004). Um estudo de espalhamento de nêutrons a baixo e alto ângulo dos complexos RecA- ATP γ S, RecA-ssDNA e RecA-dsDNA, corrobora com resultados anteriores de ME na estruturação dos filamentos da proteína RecA sobre o DNA (TIMMINS; RUIGROK; CAPUA, 1991; DICAPUA et al., 1990).

Story, Weber e Steitz (1992) obtiveram o primeiro cristal da proteína RecA da bactéria *Escherichia coli* (EcRecA). Os dados de cristalografia de raios X mostram uma organização helicoidal como aquela previamente predito por ME. O cristal pertence ao grupo espacial $P6_1$, e a simetria cristalográfica 6_1 forma um hexâmero no filamento do cristal que possui um passo helicoidal de 82,7 Å (correspondente ao *eixo-c* da cela unitária). Posteriormente, para essa mesma proteína foi determinada sua estrutura complexada a molécula de ADP (STORY; STEITZ, 1992), ambos os cristais foram obtidos na ausência de DNA.

A partir do modelo cristalográfico gerado constatou-se que a estrutura tridimensional da proteína EcRecA consiste de um grande domínio central, próximo a dois pequenos subdomínios carboxílico e amino terminal. O domínio central (aa 34 - 268) contém 8 fitas β e 6 α -hélices, intercaladas entre si. Este domínio possui o sítio de ligação do ATP, os motivos Walker A (aa 66 - 73) e B (aa 195 - 209), existem duas regiões de ligação ao DNA, *loops* L_1 (aa 157 - 164) e L_2 (aa 195 - 209) essas são desordenadas no cristal. O

domínio C-terminal (aa 270 - 328) é a região menos conservada, essa possui 3 α -hélices e 2 fitas- β e é projetado para fora do filamento de RecA. O domínio N-terminal (aa 1 - 30) é a região de interação entre os monômeros da RecA na formação do filamento. Na estrutura ainda existe 25 resíduos do domínio C-terminal que não são visíveis no mapa de densidade eletrônica e fazem parte de uma região desordenada da estrutura. A interação entre subunidades no cristal de EcRecA ocorre por meio de diversas pontes de hidrogênio e também interações iônicas (STORY; WEBER; STEITZ, 1992; STORY; STEITZ, 1992).

A estrutura das proteínas RecA de *Mycobacterium smegmatis* (MsRecA), *Mycobacterium tuberculosis* (MtRecA), também foram determinadas por cristalografia de raios X (DATTA et al., 2003; CHANDRAN et al., 2015). Essas se mostraram estruturalmente similares a estrutura da proteína EcRecA relatada acima. Entretanto, o passo helicoidal do filamento no cristal dessas proteínas é menor, 74,41 Å e 72,89 Å para a MsRecA e MtRecA, respectivamente. As proteínas MsRecA e MtRecA dividem 89% de identidade nas suas sequências, e possuem pequenas diferenças estruturais. Foram realizadas co-cristalizações das proteínas MsRecA e MtRecA com ADP, ATP γ S e dATP, constatou-se que a região da ligação do nucleotídeo dessas proteínas possuem uma considerável variação nas posições atômicas desses nucleotídeos. Essa variações foram explicadas pela presença de aminoácidos polares na região de ligação, essa região possui baixa identidade com a mesma região da proteína EcRecA (DATTA et al., 2003). Na MsRecA e MtRecA foram observadas variações estruturais responsáveis pela formação dos filamentos. A superfície do filamento da MsRecA é mais carregada negativamente, quando comparada com MtRecA, os autores atribuem essa diferença devido a sequências de aminoácidos dessas proteínas. As pontes de hidrogênio e interações iônicas observadas entre as subunidades da EcRecA, não foram observadas em MsRecA e MtRecA (DATTA et al., 2003; CHANDRAN et al., 2015).

O organismo *Deinococcus radiodurans* descoberto em 1956 por Arthur W. Anderson durante testes de esterilização de alimentos, mostrou-se altamente resistente a altas doses radioativas, essa resistência decorre em parte da sua alta capacidade de reparo às lesões no seu DNA (STORY; WEBER; STEITZ, 1992; STORY; STEITZ, 1992). A pro-

teína RecA de *Deinococcus radiodurans* (DrRecA) possui 56% de identidade com proteína EcRecA, entretanto a proteína DrRecA é extremamente tóxica para *E. coli*. O passo helicoidal do filamento da DrRecA possui 67,5 Å, é o menor passo determinado para o filamento das proteínas RecAs homólogas. Embora a estrutura da proteína DrRecA seja similar a da proteína EcRecA, existe uma grande reorientação da região C-terminal no cristal. A conformação monomérica da proteína DrRecA difere das demais estruturas já resolvidas, o domínio C-Terminal é deslocado quando comparado com EcRecA, desviando-se de 14 em relação ao eixo de empacotamento do cristal. Este domínio é importante pois possui o sítio de ligação para dsDNA e também parte dos resíduos menos conservados. Na região interna do filamento da proteína DrRecA foi constatado um aumento de potencial eletrostático positivo em relação a proteína EcRecA, o que supostamente aumentaria a afinidade ao DNA simples-fita (RAJAN; BELL, 2004).

Outras três estruturas da proteína EcRecA ligadas ao ADP foram analisadas (XING; BELL, 2004). O domínio central dessas novas estruturas são essencialmente idênticos, comparada com a estrutura de EcRecA resolvida anteriormente (). No entanto, um pequeno deslocamento é verificado na região N-terminal, que é reorientado para preservar a interação monômero-monômero no empacotamento do cristal. Também foi constatado uma significativa mudança na orientação e conformação do domínio C-terminal. Em um dos cristais, chamado de "forma 2", um íon de Ca^{2+} está ligado próximo a fita- β do domínio C-terminal e também próximo ao resíduo Asp38 do centro do filamento. Nos cristais denominados de "forma 3" e "forma 4", os íons SO_4^{-2} ou PO_4^{-2} , estão ligados no mesmo sítio que proteína EcRecA liga o nucleotídeo ADP. Nestes novos cristais o passo helicoidal do filamento é de aproximadamente de 84 Å. Estes são relativamente mais comprimidos que o da estrutura original da proteína EcRecA, que possui passo helicoidal de 82,7 Å (XING; BELL, 2004).

Também foram reportadas as estruturas cristalográficas dos complexos EcRecA-ssDNA e EcRecA-dsDNA (CHEN; YANG; PAVLETICH, 2008), devido a importância dessas estruturas ligadas ao DNA uma descrição detalhada será realizada no Capítulo 2 deste trabalho.

A proteína RecA apresenta atividade de hidrólise de ATP dependente de DNA (atividade ATPásica) (ROCA; COX, 1997; KIM et al., 2002). A atividade ATPásica vem sendo utilizada para estudar a dinâmica de formação dos filamentos (polimerização) dessa proteína sobre o DNA. Se a estrutura do DNA (linear ou circular) utilizada é conhecida, parâmetros cinéticos podem ser aferidos para estabelecer a completa formação de filamentos contínuos da proteína RecA sobre o DNA, uma vez que, a taxa de hidrólise de ATP é diretamente proporcional à quantidade de monômeros de RecA ligados ao DNA. Em condições padrões a polimerização ocorre em equilíbrio dinâmico, qualquer agente externo que perturbe o equilíbrio irá afetar diretamente os parâmetros cinéticos. Uma condição padrão comumente utilizada é aquela em que o DNA circular simples-fita (do inglês circular *single-strand* DNA (cssDNA)) é incubado com a proteína RecA, a fim de permitir que essa faça a nucleação no DNA, seguido da adição de ATP e da proteína SSB (agente externo). A proteína SSB remove a estrutura secundária do DNA, e promove a formação de filamentos contínuos da proteína RecA. Nesta situação filamentos contínuos são formados, refletindo a completa ou quase completa ligação da proteína RecA ao DNA disponível. A dinâmica de polimerização foi caracterizada pela atividade ATPásica para as proteínas RecA de *E. coli*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae* e *Deinococcus radiodurans*, na presença ou ausência da proteína SSB (KOWALCZYKOWSKI et al., 1987; ROY et al., 2009; STOHL et al., 2011; BAITIN et al., 2006; KIM et al., 2015), e mostrou-se variável de organismo para organismo.

A polimerização dos monômeros da proteína RecA sobre o DNA, conjuntamente com o passo helicoidal dos filamentos, foram identificadas como um dos aspectos importantes para as atividades gerais dessa proteína. O tamanho do passo helicoidal mostrou-se ser um parâmetro importante para identificar duas conformações distintas na qual os filamentos colapsam, sendo essas:

- Conformação comprimida (inativa): A forma comprimida do filamento é funcionalmente inativa, formada pela proteína RecA sozinha ou ligada ao ADP, tanto na presença ou ausência de DNA, nessa conformação os hexâmeros do filamento possuem um passo helicoidal que variam de 65-85 Å, possuindo 6,2 monômeros de

RecA por volta (STORY; WEBER; STEITZ, 1992; GALLETTO et al., 2006; RAJAN; BELL, 2004; DATTA et al., 2003; CHANDRAN et al., 2015; ELLOUZE et al., 1995);

- Conformação estendida (inativa): A forma estendida e funcionalmente ativa, é formada pela RecA ligada ao ssDNA ou dsDNA, na presença de ATP ou análogos de ATP, mantêm-se 6,2 monômeros de RecA por volta, e assume um passo de 90-100 Å, determinado por ME e técnicas de espalhamento (EGELMAN; STASIAK, 1986; TIMMINS; RUIGROK; CAPUA, 1991; DICAPUA et al., 1990).

A importância da proteína RecA dá-se pelo fato de que ela faz parte um conjunto de proteínas cruciais para recombinação genética, que por sua vez, faz parte de uma família de proteínas RecAs, utilizadas para a recombinação em ambos, procariotos e eucariotos (CHANET et al., 1996).

Embora algumas estruturas tridimensionais da proteína RecA tenham sido resolvidas, muitas questões ainda permanecem em aberto. Em particular, no que diz respeito da polimerização da RecA sobre o DNA. Informações estruturais sobre a proteína HsRecA e o filamento RecA-DNA, juntamente com dados fisiológicos podem contribuir para o entendimento das funções dessa proteína.

A proposta deste trabalho consistiu em explorar a partir de uma nova estrutura da proteína RecA os mecanismos moleculares que promovem a polimerização da RecA como filamentos. Para tal propósito foi realizada a determinação da estrutura tridimensional da proteína RecA da bactéria *Herbaspirillum seropedicae* SmR1 (HsRecA) a alta resolução atômica, por meio da técnica de difração de raios X. Utilizando ferramentas de bioinformática foram gerados alinhamentos estruturais com diversas RecAs homólogas, sobreposições estruturais e potenciais eletrostáticos de superfície, a fim de identificar se existe um mecanismo estrutural geral que governa a polimerização da RecA na forma de filamentos helicoidais.

E também, caracterizar a dinâmica de formação dos filamentos da proteína RecA de *H. seropedicae*. Para tal proposta foi utilizado o monitoramento da atividade ATPásica

da RecA por meio de espectroscopia de absorção no ultravioleta e imagens de microscopia eletrônica de transmissão.

A presente tese é uma complementação à apresentada pela Profa. Dra. Carolina Weigert Galvão ao Curso de Pós de Graduação em Ciências, Área Bioquímica, do Setor de Ciências Biológicas da Universidade Federal do Paraná, que teve como objetivo geral a caracterização bioquímica das proteínas RecA e RecX de *H. seropedicae*. A bactéria *H. seropedicae* é uma bactéria endofítica capaz de fixar de nitrogênio sob condições de limitação de amônio e oxigênio, sendo muito importante para sistemas agrícolas.

Assim, esta tese teve como objetivo geral:

- Realizar a caracterização estrutural por cristalografia de raios X, e funcional por técnicas espectroscópicas da proteína HsRecA. Direcionando os estudos para a dinâmica de formação e estabilidade dos filamentos da proteína HsRecA na ausência e presença de DNA.

No qual conduziu aos seguintes objetivos específicos:

- Realizar a expressão e purificação da proteína HsRecA na sua forma nativa por métodos cromatográficos de alta especificidade;
- Cristalizar e determinar a estrutura tridimensional da proteína HsRecA por meio da técnica de difração e cristalografia de raios X, adicionalmente realizar estudos estruturais comparativos com o modelo cristalografico obtido;
- Caracterizar a dinâmica de formação e estabilidade dos filamentos da proteína HsRecA, por meio da atividade ATPásica e microscopia eletrônica.
- Elucidar os mecanismos moleculares que promovem a polimerização da RecA como filamentos.

Capítulo 2

Estrutura e Função da Proteína RecA

2.1 Fontes de Lesões e Mecanismos de Reparo de DNA

Todo organismo vivo está sujeito a danos no seu material genético. O reparo desses danos são essenciais para preservar uma variedade de processos biológicos relacionados ao metabolismo de DNA. Neste capítulo será discutido algumas fontes de lesões ao DNA, e os mecanismos de reparo. Nos mecanismos de reparo será dado ênfase especial para a via por recombinação homóloga uma vez que essa é mediada pela proteína RecA. A proteína RecA é chave dos mecanismo de reparo e objeto de estudo deste trabalho.

O genoma de todo organismo contém importantes informações que direcionam seu desenvolvimento, reprodução e sobrevivência. O material genético está sujeito a vários danos, assim como qualquer outra macromolécula está sujeita a modificações espontâneas, como as decorrentes da instabilidade das ligações químicas, ou ainda lesões causadas por agentes físicos ou químicos. O reparo dessas lesões é essencial para a manutenção de toda informação genética, e alterações podem levar a modificação da estrutura primaria de proteínas, estrutura de RNN funcional e controle de expressão, uma vez que o próprio DNA é molde para sua duplicação, produção de diversos tipos de RNAs e proteínas (KUZMINOV, 1999; LINDAHL; NYBERG, 1972; TORNALETTI; HANAWALT, 1999).

As lesões no DNA podem ser de origem externa e/ou interna a célula. Internamente espécies reativas de oxigênio são os principais agentes deletérios, ocorrendo normal-

mente durante o metabolismo celular, entre eles pode-se citar o peróxido de hidrogênio, esse reage com íons metálicos como o Fe^{2+} próximo ao DNA, gerando radical OH que por sua vez reagem imediatamente com a molécula de DNA (HENLE; LINN, 1997).

As fontes externas de lesões ao DNA incluem luz ultravioleta, radiações ionizantes como raios X, raios γ e agentes químicos. Tratando das lesões por radiações ionizantes, essas podem causar simples e/ou duplas quebras no DNA (*double-strand break* - DSB). Na presença de lesões no DNA, eventos como a replicação podem ser interrompidos, e somente será restabelecido até que algo seja realizado para contornar o problema. No entanto, dependendo do tipo de lesão a replicação é mantida, isso ocorre adicionando um nucleotídeo aleatório frente à lesão, esse sistema é chamado de síntese translesão (*bypass*), é suscetível a erros e aumenta a frequência de mutações e, uma potencial origem de câncer (BAYNTON; FUCHS, 2000), essa é uma situação bastante peculiar. No entanto, a maioria das lesões são reparadas por meio de diversos mecanismos, os quais serão explicados posteriormente neste capítulo.

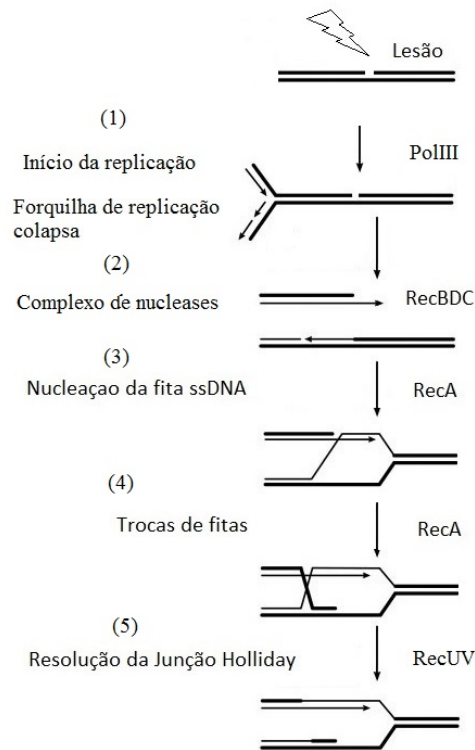
Do ponto de vista da replicação, essa quando é interrompida ocasiona pelo menos três danos à molécula de DNA: (i) indução de deleções e rearranjos genômicos, (ii) indução de mutações e (iii) indução de letalidade, o próprio DNA afim de suprimir ou reduzir a frequência de alterações desenvolveu uma rede enzimática composta de diversas proteínas, capaz de contornar tais danos (FRIEDBERG; WALKER; SIEDE, 1995).

A rede enzimática desenvolvida utiliza-se de diferentes mecanismos de atuação, como por exemplo reparo por reversão direta, reparo por excisão de bases, nucleotídeos e pareamentos incorretos e o reparo por recombinação homóloga (TORNALETTI; HANAWALT, 1999; FRIEDBERG; WALKER; SIEDE, 1995; COX, 2001). Sendo que a recombinação homóloga faz parte de um mecanismo biológico mais amplo e não é utilizada somente para fins de reparo e é mediada pela proteína RecA (COSTECHAREYRE; BERTOLLA; NESME, 2009).

Reparo por Recombinação Homóloga

O processo de reparo por recombinação homóloga promove o reparo de DNA de forma não mutagênica (ROBU; INMAN; COX, 2001). No reparo não mutagênico, a restauração da replicação ocorre sem comprometer a integridade do genoma. Isso pode ocorrer por duas vias, (I) reparo de dupla quebra na forquilha de replicação (do inglês *double strand break*, DSB) e (II) reparo de lacuna (do inglês *gap*). A figura (2.1) apresenta sistematicamente o modelo de reparo da DSB via recombinação homóloga. Devido a presença de uma lesão a forquilha de replicação composta pela DNA polimerase III (polIII) encontra uma quebra na estrutura do DNA em umas das fitas molde e colapsa, etapa (1), de acordo com a figura. O duplex de fitas contendo a lesão é processado pelo complexo de nucleases RecBDC, que digere a sequência de nucleotídeos da fita danificada na direção 5'–3', até encontrar a sequência chi (5'–GCTGGTGG–3'), resultando em uma região de DNA simples fita (ssDNA), etapa (2). A proteína RecA é recrutada e forma um filamento ao redor do ssDNA, identifica a fita homóloga no duplex inferior de dsDNA, etapa (3), e a própria proteína RecA realiza a trocas de fitas utilizando a fita complementar, etapa (4). Por fim, após a troca de fitas é gerado uma estrutura ramificada (em forma de cruz) denominada junção de Holliday, que é processada pelo complexo ReuABC, etapa (5), e a replicação é restabelecida.

Figura 2.1: A ilustração mostra as etapas do modelo de reparo via recombinação homologa da DBS. Os principais processos biológicos estão indicados nas etapas de 1 - 4, de acordo com o texto. Neste modelo destaca-se a proteína RecA, responsável por catalisar da troca de fitas de DNA e assim efetuar o reparo da fita danificada.



Fonte: Adaptado de Cox (2001).

Dentro dessa complexa maquinaria para conservar a integridade genética, destaca-se a proteína RecA, a função da proteína RecA no sistema de reparo por recombinação consiste em catalisar as trocas de fita de DNA entre as duas sequências homólogas, dessa forma promove o reparo por recombinação, preservando a fidelidade da sequência de nucleotídeos da fita filha.

2.2 A Proteína RecA

Funções Gerais da Proteína RecA

A proteína RecA exerce papel central no complexo sistema de reparo a lesões do DNA. Ela também exerce importante papel na indução de expressão de proteínas que co-fetueam o reparo de DNA atuando sobre uma rede regulatória conhecida como resposta SOS. A proteína RecA é encontrada em todos os organismos, incluindo eucariotos onde sua homóloga é denominada de Rad51. Basicamente, a RecA é a proteína central do sistema de reparo de DNA e suas funções incluem:

- (A) Ativação do sistema SOS: A proteína RecA ao clivar a proteína LexA aciona o sistema de reparo SOS, esse evento desencadeia a expressão de mais de 43 genes ligados a esse processo, incluindo a própria RecA (COURCELLE et al., 2001);
- (B) Ativação do mecanismo de reparo mutagênico: Ao clivar o complexo UmuDC ativa o sistema de reparo mutagênico, geralmente expresso numa situação de graves lesões no DNA (TANG et al., 1999; GRUBER et al., 2015);
- (C) Reação de trocas de fitas: Na presença de lesões no DNA a proteína RecA realiza a troca de fitas entre duas sequências homólogas, dessa forma promove o reparo pela via de recombinação (COX, 1997; MACFARLAND et al., 1997);
- (D) Atividade ATPásica: A proteína RecA forma filamentos helicoidais ao redor do DNA e liga-se cooperativamente ao ATP. As atividades acima citadas são suportadas pela hidrólise de ATP (LEE; COX, 1990).

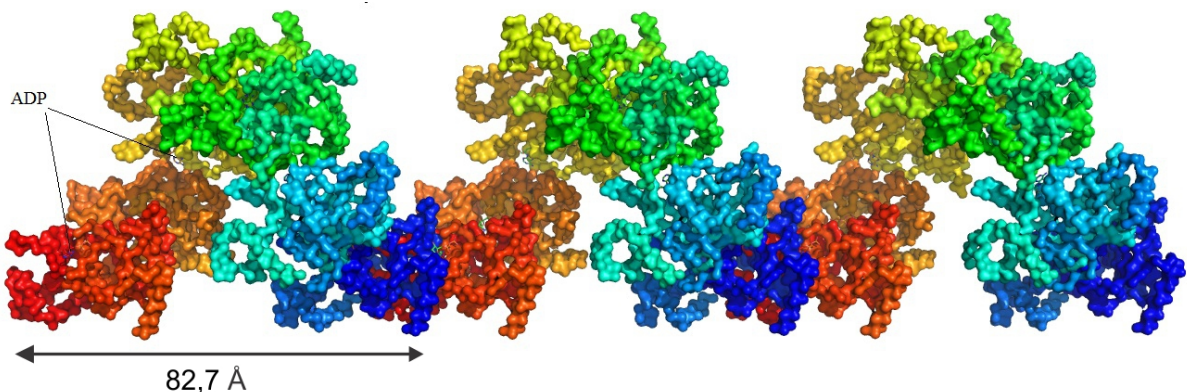
Estrutura Tridimensional da Proteína RecA

A proteína RecA na sua forma funcional polimeriza-se ao redor do DNA e forma filamentos helicoidais, possuindo hexâmeros como unidade biológica (YU; EGELMAN, 1997). Imagens por microscopia eletrônica (ME) têm sido utilizadas para caracterizar o filamento RecA-DNA. A figura (2.2) apresenta 3 hexâmeros da proteína RecA, formando

parte de um modelo de filamento, esse modelo foi proposto utilizando reconstruções tridimensionais das imagens geradas por ME e possui 20 Å de resolução atômica.

O modelo provém uma imagem pictórica da estrutura do filamento da proteína RecA. No modelo, significativa alteração é observada nos filamentos contendo RecA-dsDNA-ATP, onde mudanças na região C-terminal são dependentes do estado do nucleotídeo ligado, ou seja, ATP ou ADP. Em reconstruções de imagens de ME o sítio de ligação do ADP está na interface entre adjacentes monômeros. O modelo não possui resolução suficiente para identificar o DNA ligado (VANLOOCK et al., 2003).

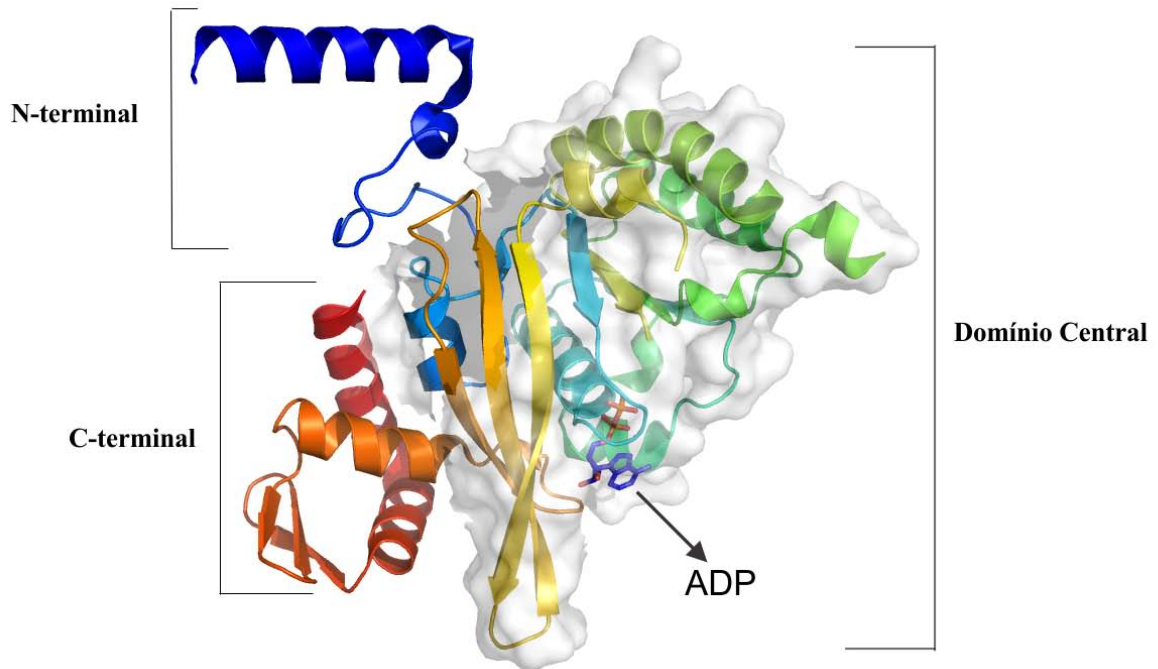
Figura 2.2: Modelo estrutural do hexâmero da proteína EcRecA gerado por reconstruções de imagens de microscopia eletrônica. Na figura, é mostrado um filamento contendo três subsequentes hexâmeros. A posição de ligação do ADP está indicada na interface monômero-monômero. Cada monômero possui uma cor para diferenciação.



Fonte: Adaptado de VanLoock et al. (2003), pdb:1N03.

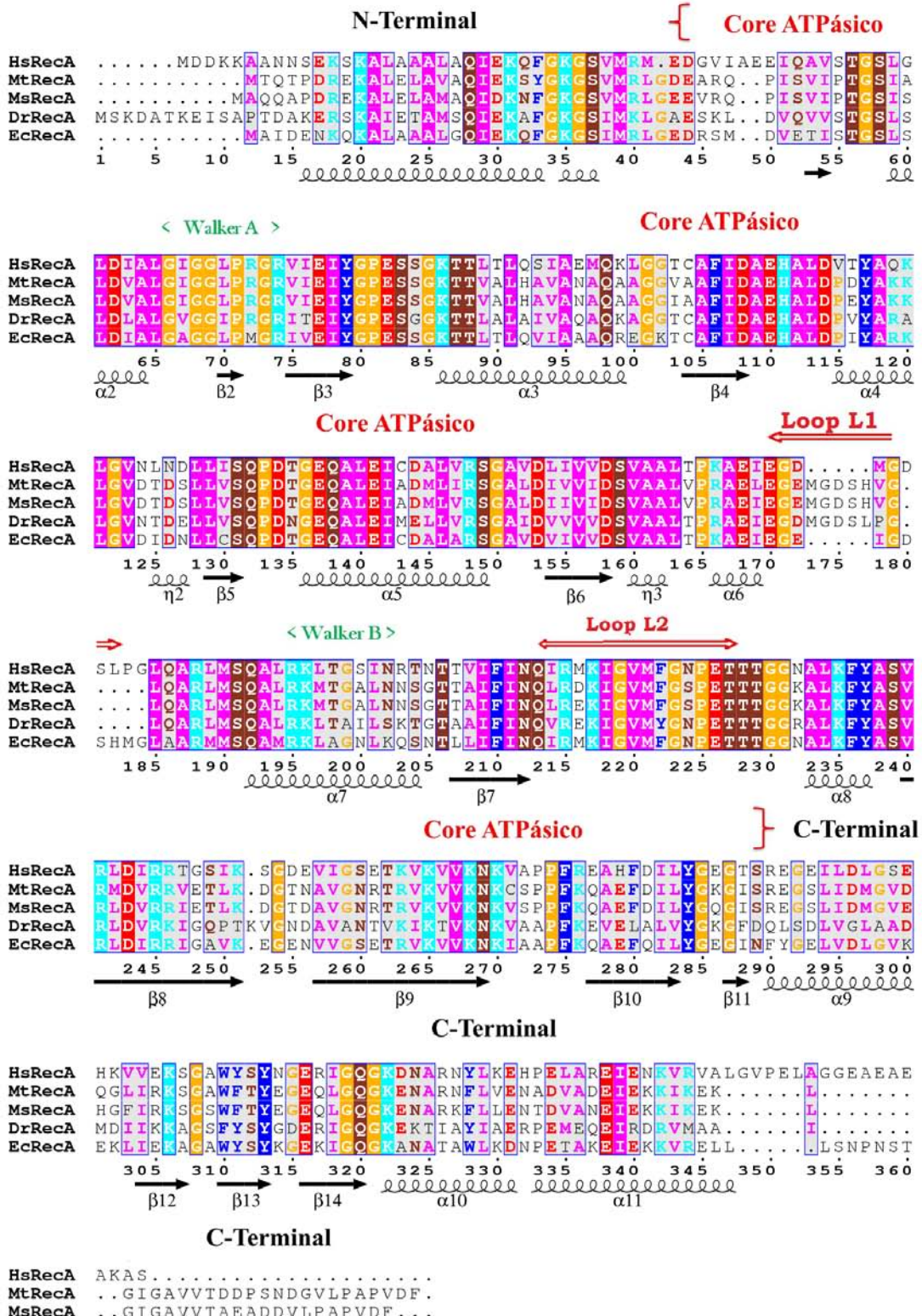
O primeiro modelo cristalográfico do monômero foi obtido para a proteína RecA da bactéria *E. coli* (EcRecA). A estrutura monomérica da EcRecA consiste de um grande domínio central, também chamado de core ATPásico, próximo a dois subdomínios N-terminal (NT) e C-terminal (CT). A representação em diagrama de fitas da estrutura monomérica da EcRecA é mostrada na figura (2.3) e um alinhamento sequencial, com os respectivos os motivos estruturais com outras sequências de bactérias é mostrado na figura (2.4) (STORY; WEBER; STEITZ, 1992; STORY; STEITZ, 1992).

Figura 2.3: Diagrama de fitas do modelo cristalográfico da proteína RecA de *E. coli*, esse consiste de três domínios: uma região N-terminal, um domínio central (região sombreada) e um domínio C-terminal. Também está indicado a região do sítio de ligação do nucleotídeo ADP.



Fonte: Adaptado de Story, Weber e Steitz (1992), Story e Steitz (1992), *merge* pdb:1REA e 2REB.

Figura 2.4: Alinhamento sequencial das proteínas de RecA de *H. seropedicae* (HsRecA), *M. tuberculosis* (MtRecA), *M. smegmatis* (MsRecA), *D. radiodurans* (DrRecA), (EcRecA) *E. coli*.



Fonte: O Autor.

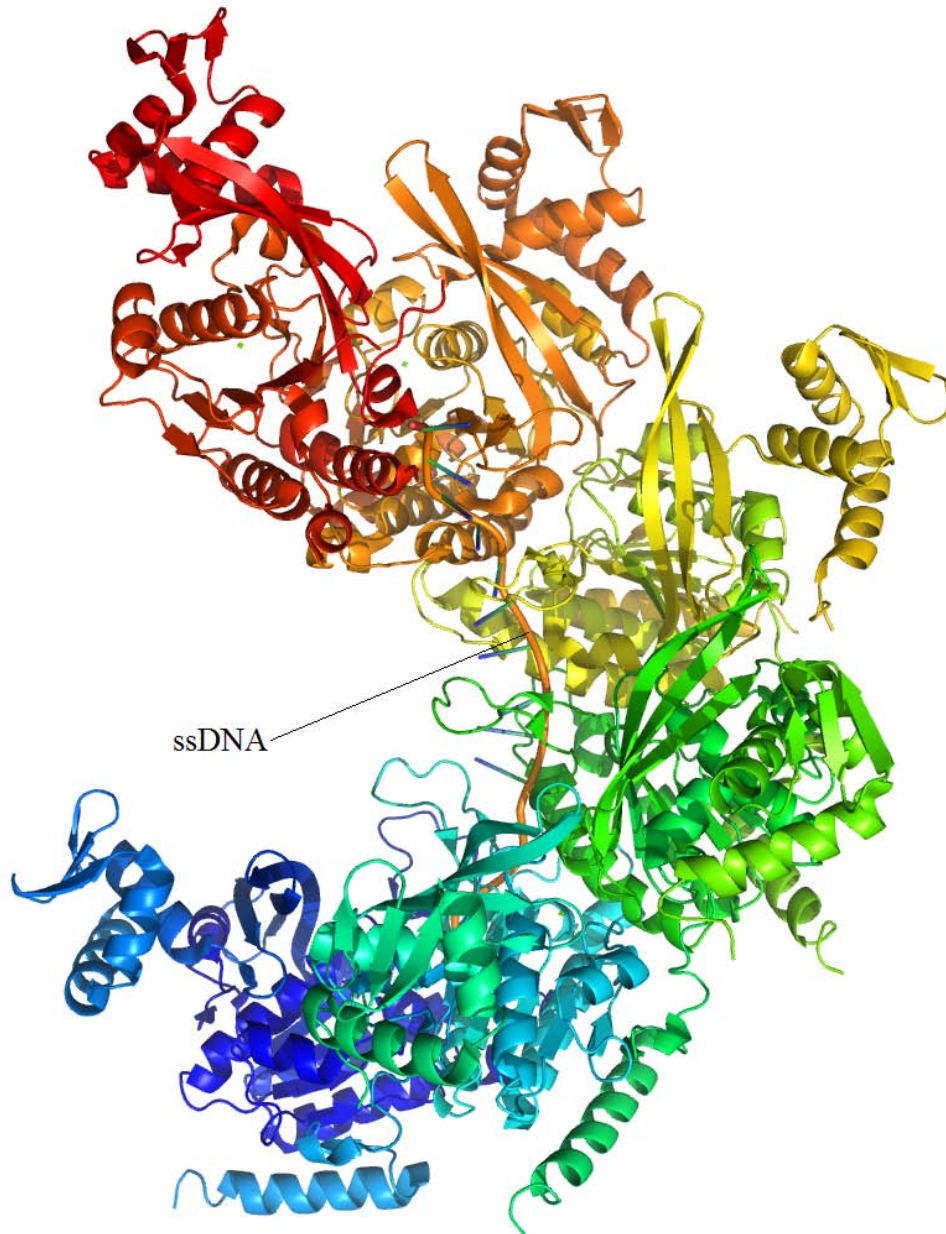
O cristal pertence ao grupo espacial ($P6_1$), que por simetria gera um filamento também helicoidal, em concordância com os dados de ME. O eixo c da cela unitária possui 82,7 Å, esse corresponde ao passo helicoidal do filamento na estrutura cristalográfica. O subdomínio N-terminal (aa 1 - 30) é a região de interação entre monômero-monômero, conjuntamente com o domínio central estabiliza a formação do filamento no cristal. O domínio central (aa 34 - 268) contém 8 fitas β e 6 α - hélices intercaladas entre si. Este domínio possui o sítio de ligação do ATP, os motivos Walker A (aa 65 - 73) e B (aa 194 - 209), existem ainda duas regiões de ligação ao DNA, designadas loops L_1 (aa 156 - 164) e L_2 (aa 194 - 209) esses que são desordenadas no cristal e não estão visíveis nos mapas de densidade eletrônica. O subdomínio C-terminal (aa 269 - 328), constitui-se de 3 α -hélices e 2 fitas β e é projetado para fora do filamento, o C-terminal também possui um sítio de ligação para o dsDNA. Ainda existem 25 resíduos na região C-terminal da estrutura que não são visíveis no mapa de densidade eletrônica e fazem parte de uma região desordenada da estrutura (STORY; WEBER; STEITZ, 1992; STORY; STEITZ, 1992).

Os monômeros da proteína EcRecA se ligam ao DNA produzindo um filamento helicoidal dextrogiro. Chen, Yang e Pavletich (2008) relataram a estrutura cristalográfica dos complexos EcRecA-ssDNA e EcRecA-dsDNA. A proteína EcRecA se liga a cada três nucleotídeos de ssDNA ou 3 pares de base de dsDNA. Devido a ligação da RecA ao DNA a conformação tipo B do DNA é alterada: na conformação B o DNA possui 3,4 Å e aloca 18,6 pares de base por volta helicoidal, porém quando ligado ao filamento, altera-se para 5,1 Å e aloca 10,5 pares de base por volta helicoidal. O ssDNA e ATP ligam-se na interface monômero-monômero cooperativamente, explicando o motivo pela qual ligação do ATP é dependente de ligação ao DNA.

A figura (2.5) mostra a estrutura da EcRecA-ssDNA, essa corresponde a molécula $EcRecA_6(dT)_{18}(ADP-AlF_4^- Mg)$, onde o $ADP-AlF_4^- Mg$ é um análogo de ATP não hidrolisável, que imita o estado de transição da hidrólise de ATP. Esse filamento representa o intermediário estrutural no pareamento para doador dsDNA na reação de trocas de fitas e é chamado de filamento pré-sináptico. A figura (2.6) mostra a estrutura da EcRecA-dsDNA, essa corresponde a molécula $RecA_5(dT)_{15}(dA)_{12}(ADP - AlF_4^- Mg)$, e

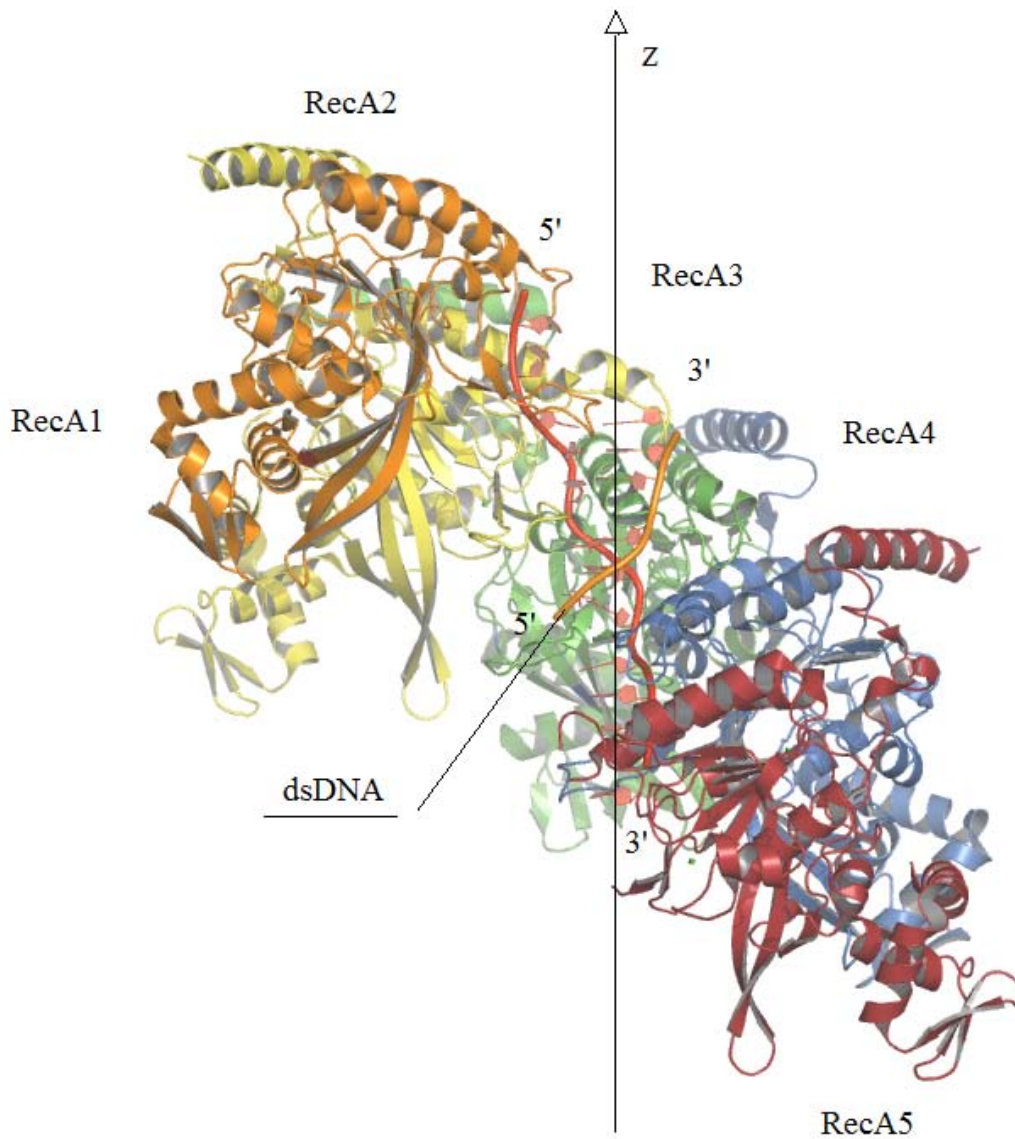
foi postulada como sendo o produto após a reação de trocas de fitas (CHEN; YANG; PAVLETICH, 2008).

Figura 2.5: Diagrama de fitas do modelo cristalográfico do filamento pré-sináptico, representando o estado do filamento antes da reação de troca de fitas, da proteína RecA de *E. coli*. O modelo contém 6 monômeros ligados por um *loop* flexível. No filamento o ssDNA está ligado internamente ao filamento. Cada monômero está colorido de uma cor para diferenciação.



Fonte: Adaptado de Chen, Yang e Pavletich (2008), pdb:3cmu.

Figura 2.6: Diagrama de fitas do modelo cristalográfico do filamento pós-sináptico, produto após a reação de trocas de fitas, da proteína RecA de *E. coli*. No filamento o dsDNA está ligado internamente ao filamento. O modelo contém 5 monômeros ligados por um *loop* flexível.



Fonte: Fonte: Adaptado de Chen, Yang e Pavletich (2008), pdb:3cmv.

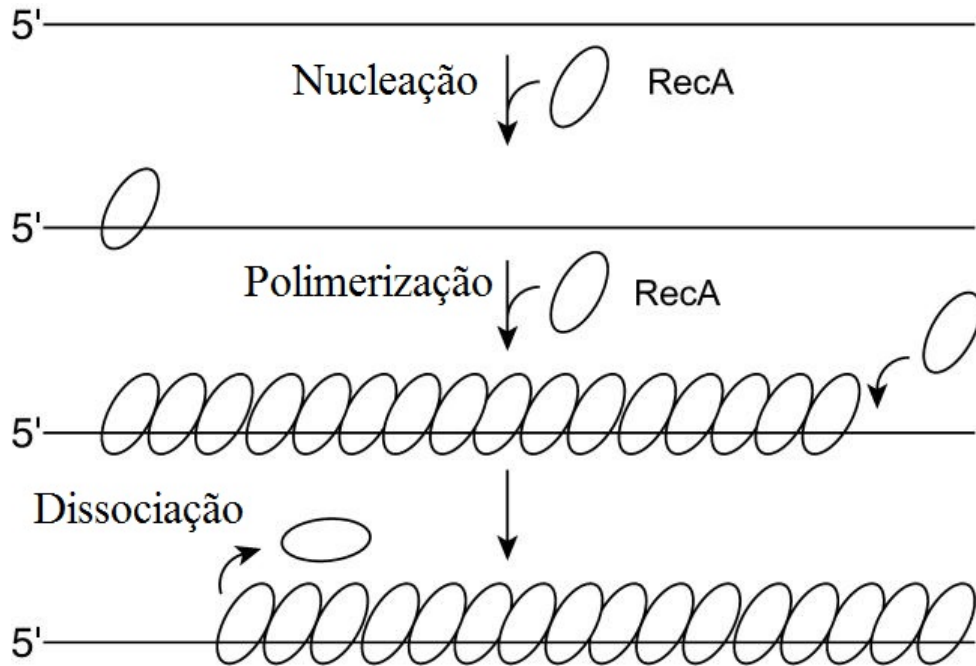
Nas estruturas contendo ssDNA e dsDNA os loops de ligação ao DNA L_1 e L_2 no cristal passa a ser ordenado, onde L_1 forma uma pequena hélice (αL_1), seguido por uma volta e L_2 forma duas fitas β antiparalelas $\beta L_1 - \beta L_2$.

Para obtenção das estruturas reportadas por Chen, Yang e Pavletich (2008), todos os monômeros de RecA de um hexâmero foram expressos ligados covalentemente através de um link de aminoácidos, sendo que o primeiro monômero apresentava o domínio N-terminal deletado e o último monômero apresentava o domínio C-terminal mutado nas seguintes posições Cys117Met, Ser118Val e Gln119Arg. Segundo os autores, tais mudanças foram realizadas afim de evitar a oligomerização entre hexâmeros. Na *RecA*₅ seu C-terminal foi mutado nas seguintes posições Cys117Met, Ser118Val e Gln119Arg, essas mutações também impedem o processo de oligomeração. Alterações na região C-terminal já demonstraram anteriormente que as atividades de ligação ao DNA e trocas de fitas são modulações que afetam negativamente as atividades da proteína RecA em *E. coli*. As *RecA*₂₋₄ permaneceram com suas sequências inalteradas. A presença do *link* (Thr-Gly-Ser-Thr-Gly-Ser-Gly-Thr-Thr-Gly-Ser-Thr-Gly-Ser) não foi abordado pelos autores sobre a possibilidade de induzir alterações na estrutura do filamento.

A Polimerização da Proteína RecA Sobre o DNA

A proteína RecA pode ligar-se em ambos ssDNA e dsDNA, mas a polimerização ocorre preferencialmente sobre o ssDNA (REGISTER; GRIFFITH, 1985; LUSETTI; COX, 2002). A formação do filamento de RecA sobre o ssDNA, na presença de ATP ou análogos de ATP, é um passo crucial no reparo da DSB durante o processo de reparo via recombinação (KOWALCZYKOWSKI; EGGLESTON, 1994). Esse processo ocorre em duas etapas distintas: na primeira etapa a proteína RecA nucleia-se sobre o DNA como oligômeros de 4-6 monômeros, na segunda, ocorre uma rápida polimerização que se estende sobre o DNA a taxas maiores que 1000 monômeros/min a 37 °C, essa é unidirecional, ocorrendo na direção 5' → 3' do ssDNA. Neste processo é formado um contínuo filamento da proteína RecA sobre o DNA. Por fim, ocorre o processo de dissociação do monômeros a partir da extremidade oposta ao terminal de associação, ou seja, a dissociação também é unidirecional ocorrendo na direção 5' → 3' do ssDNA (GALLETTO et al., 2006; SHAN; COX, 1997; FU et al., 2013). A figura (2.7) apresenta sistematicamente a formação dos filamentos de RecA sobre o DNA.

Figura 2.7: Polimerização da proteína RecA sobre o ssDNA. A RecA nucleia-se sobre o DNA como oligômeros de 4-6 monômeros, posteriormente a polimerização pode ocorrer a taxas maiores que 1000 monômeros/min a 37 °C. E por fim, ocorre a dissociação de um único monômero por vez. Ambas polimerização e dissociação são unidirecionais, ocorrendo na direção 5' → 3' do ssDNA.



Fonte: Adaptado de Lusetti e Cox (2002).

Nos filamentos de RecA formados sobre o DNA circular simples-fita (*circular single-strand DNA* (cssDNA)) o equilíbrio entre dissociação e associação dos monômeros mantém o cssDNA totalmente preenchido, consequentemente a taxa de hidrólise de ATP mantém-se constante e máxima, 30 μ mol/L por minuto, na presença de lacunas, essas são rapidamente preenchidas pela rápida extensão do filamento, não alterando significativamente a taxa de hidrólise. A manutenção da taxa de hidrólise de ATP é característica exclusiva do filamento de RecA formado sobre cssDNA, que permite verificar a influência de outras proteínas sobre a atividade de hidrólise de ATP da proteína RecA (DREES; LUSETTI; COX, 2004; DREES et al., 2004).

O filamento formado sobre o DNA colapsa em duas conformações: A primeira conformação é a forma comprimida do filamento e é funcionalmente inativa, é formada pela proteína RecA sozinha ou ligada ao ADP. Tanto na presença ou ausência de DNA, nessa

conformação os hexâmeros do filamento possuem um passo helicoidal de 65-85 Å (STORY; WEBER; STEITZ, 1992; GALLETTO et al., 2006; RAJAN; BELL, 2004; DATTA et al., 2003; CHANDRAN et al., 2015; ELLOUZE et al., 1995), e possui 6,2 monômeros de RecA por volta. A segunda conformação é a forma estendida e funcionalmente ativa, essa é formada pela RecA ligada ao ssDNA ou dsDNA, na presença de ATP ou análogos de ATP, mantêm 6,2 monômeros de RecA por volta, e assume um passo helicoidal de 90-100 Å, determinado por microscopia eletrônica (EGELMAN; STASIAK, 1986).

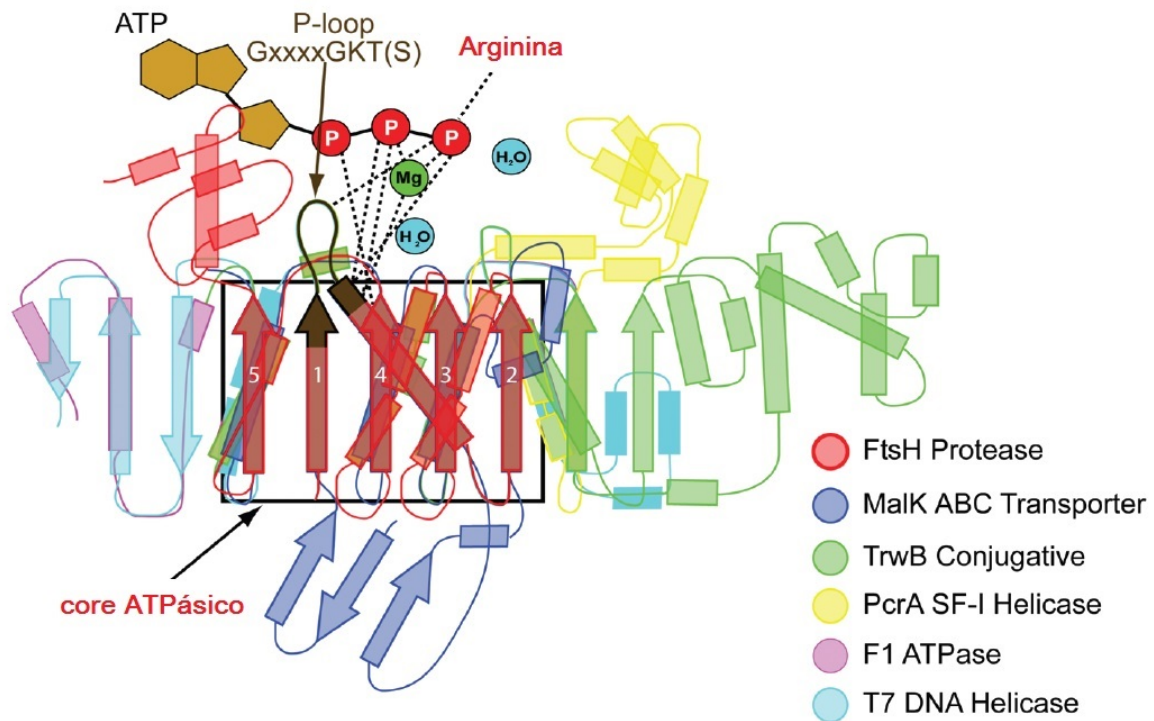
Atividade ATPásica da Proteína RecA

A ligação do ATP otimiza a polimerização da RecA ao DNA. Enquanto a hidrólise de ATP não é requerida para a associação dos monômeros de RecA no ssDNA, sem a hidrólise de ATP não ocorre a sua dissociação (SHAN; COX, 1997; LINDSLEYS, 1990; SHIVASHANKAR et al., 1999). A hidrólise de ATP ocorre uniformemente através do filamento (BRENNER et al., 1987), não foram observadas evidências de aumento dessa taxa nas suas extremidades. Na ausência de qualquer agente que altere a atividade ATPásica intrínseca de RecA, a taxa de hidrólise de ATP é diretamente proporcional à quantidade de monômeros de RecA ligados ao DNA, e essa pode ser utilizada para caracterizar de forma indireta a dinâmica de formação e estabilidades dos filamentos (SHAN; COX, 1997; DREES; LUSETTI; COX, 2004).

A atividade ATPásica na proteína RecA é dependente do DNA, apenas em condições de alta concentração de sal (1,5-2 mol/L) e pH 6,0 a hidrólise ocorre na ausência de DNA (PUGH; COX, 1988b). O valor do k_{cat} da RecA para a hidrólise de ATP é de aproximadamente de 30 min^{-1} quando ligada ao ssDNA e de 16 - 20 min^{-1} quando ligada ao dsDNA a 37°C.

Diversas proteínas que exibem atividade ATPásica usam a energia da hidrólise do ATP para realizar trabalho mecânico, elas possuem um domínio estrutural comum chamado core ATPásico, que foi primeiro identificado na proteína RecA. A figura (2.8) mostra a topologia do core ATPásico de diversas proteínas tipo-RecA.

Figura 2.8: Sobreposição da topologia de seis proteínas tipo-RecA. Essas seis proteínas pertencem a diversas famílias de proteínas, incluindo a superfamília AAA protease FtsH (vermelho), ABC transporte MalK (azul), PilT família conjugativa de plasmídeo TrwB (verde), Superfamília da helicase I PcrA (amarelo), tipo-DnaB T7 helicase (azul ciano), e F1 ATPase (magenta).



Fonte: Adaptado de Liao (2011).

Estruturalmente o core ATPásico se organiza de forma que suas fitas β constituintes seguem a ordem 51432, como mostrado na figura (2.8). As fitas β_1 e β_2 são sempre conectadas pelo motivo Walker-A (ou *P-loop*) e uma α -hélice, com resíduos conservados, tipicamente (G/A)xxxxGK(S/T) (LIAO, 2011).

A estrutura das proteínas EcRecA, MtRecA e DrRecA foram determinadas na presença de vários nucleotídeos, no quais incluem ADP, dATP, $ADP - AlF_4$, AMP-PNP e $ATP\gamma S$. Estes nucleotídeos fazem contato com seis regiões do sítio ATPásico no monômero da RecA: (i) o motivo Walker A ou *P-loop* (resíduos 68-73) que interage com o β e γ - fosfatos do nucleotídeo ligado, e também com íon Mg^{2+} ; (ii) o motivo Walker B (resíduos 139-144), que projeta o resíduo Asp-148 próximo ao íon Mg^{2+} ; (iii) uma região que assume

a conformação de loop e contém o resíduo Glu-96, esse resíduo ativa o ataque de água para efetuar a hidrólise; (iv) uma região precedente ao loop de ligação ao DNA L_2 , que contém o resíduo Gln-194, o qual funciona com uma chave de indução de efeito alostérico sobre o *loop*; (v) hélice D, que estabiliza o anel de adenina; e por fim (vi) um *loop* (resíduos 264-269) entre a domínio central e C-terminal, que está próximo a ribose (BELL, 2005).

A Hipótese da RecA como Motor Molecular

Diversos modelos foram elaborados a fim de elucidar como a proteína RecA utiliza a energia proveniente da hidrólise do ATP para realizar trabalho mecânico ao mover e/ou rotacionar segmentos de DNA ao longo do filamento, no entanto, todos falharam em explicar por exemplo como a RecA efetua a reação de troca de fitas utilizando-se dessa energia. Atualmente, dois modelos parecem satisfazer o acoplamento da hidrólise de ATP e a reação de trocas de fitas, o primeiro modelo é da redistribuição dos monômeros de RecA sobre o filamento, e o segundo que acopla a rotação do ssDNA e do dsDNA homólogo (MACFARLAND et al., 1997).

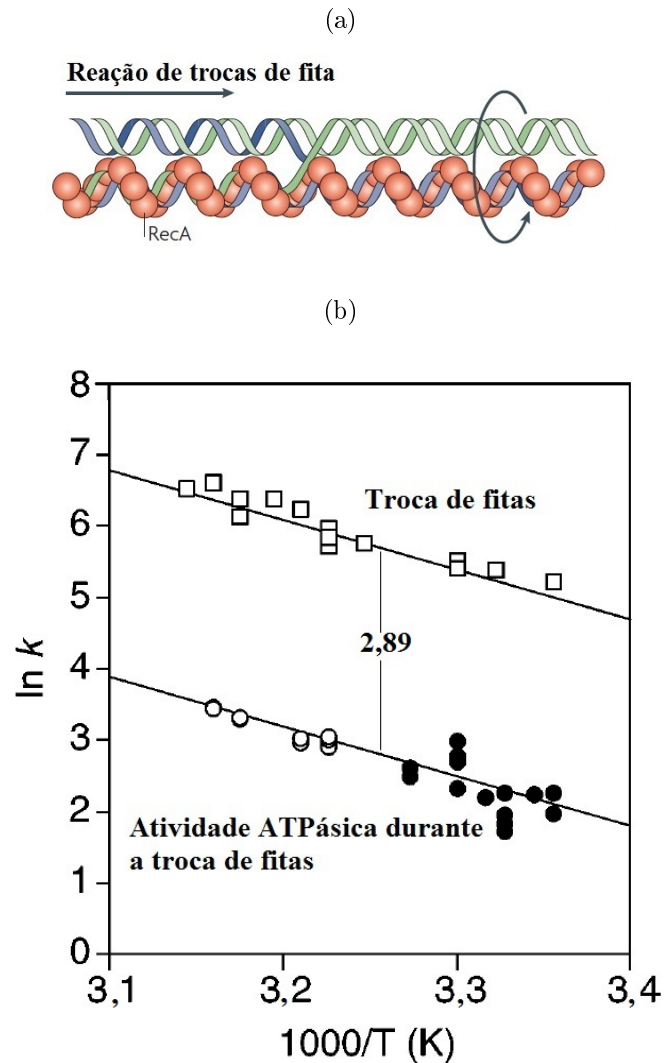
O modelo de redistribuição não suporta diversas evidências experimentais e nessa discussão será apresentado apenas o modelo de rotação. A reação de troca de fitas ocorre de maneira independente da hidrólise de ATP quando o DNA apresenta até 3 kpb, porém, nessa condição, a reação de troca de fitas ocorre de forma menos eficiente devido a formação de pareamentos improdutivos.

O modelo de rotação faz a proposição que a reação de troca de fitas ocorre de forma menos eficiente na ausência de hidrólise de ATP, devido a formação de pareamentos improdutivos ou em decorrência de outras interações entre RecA-ssDNA e o homólogo dsDNA (SIGURDSSON et al., 2001). Nesse contexto a RecA possuiria a função de um motor molecular, cuja fonte de energia seria proveniente da hidrólise de ATP. Na hipótese do modelo de rotação o DNA passa ao redor de um hexâmero do filamento para o subsequente, com seis passos cada rotação completa. Duas moléculas de DNA, apenas uma que está dentro do filamento de RecA é rotacionada sobre a outra, uma ilustração dessa rotação é mostrada na figura (2.9a). Para que a rotação ocorra os ciclos de hidró-

lise do ATP de monômeros da RecA teriam que ser coordenados, com cada sexagésimo monômero hidrolisando ATP num dado momento. Ainda, uma rotação completa requer que cada monômero do filamento hidrolise ATP. A nível de monômeros, uma rotação de DNA é igual a um evento de hidrólise de ATP e a cada evento de rotação 18 pb do DNA movem-se para o subsequente hexâmero.

O modelo de rotação também prediz que a taxa de hidrólise de ATP (k_{cat} em (min^{-1})) por cada monômero pode ser relatada com os 18 pb que se movimentam para o subsequente hexâmero em unidade de pb por minuto ($pbmin^{-1}$). Por exemplo, se cada monômero da proteína RecA está hidrolisando 20 ATPs por minuto, isso implicaria em 20 rotações, e no deslocamento de 360 pb ao redor do filamento. A taxa na qual ocorre a reação de trocas de fitas em $pbmin^{-1}$ possui correlação direta com a taxas de hidrólise de ATP (k_{cat}) durante a reação de troca de fitas. Um gráfico contendo as respectivas taxas em função do logaritmo da temperatura é mostrado na figura (2.9b)), onde T corresponde as diferentes temperaturas na qual a reação foi realizada. O gráfico mostra uma correlação linear para as duas taxas, e a separação entre as taxas é de exatamente um fator de 2,89 em unidades do logaritmo natural ($\ln 18 = 2,89$), que se correlaciona com os 18 pb que se movimentam para o subsequente hexâmero durante a reação de troca de fitas (para revisão detalhada (COX, 2003)).

Figura 2.9: Modelo de facilitação de rotação do DNA. Em (a) é mostrado um esquema de rotação do DNA ao redor do filamento, no modelo de facilitação de rotação a proposição é que a reação de troca de fitas pode estar acoplada a hidrólise de ATP. Em (b) gráfico de Arrhenius para a taxa na qual ocorre a reação de troca de fitas (em unidades de $pbmin^{-1}$) e a taxa de hidrólise de ATP (em unidades de min^{-1}) durante a reação de troca de fitas.



Fonte: Adaptado de Cox (2001).

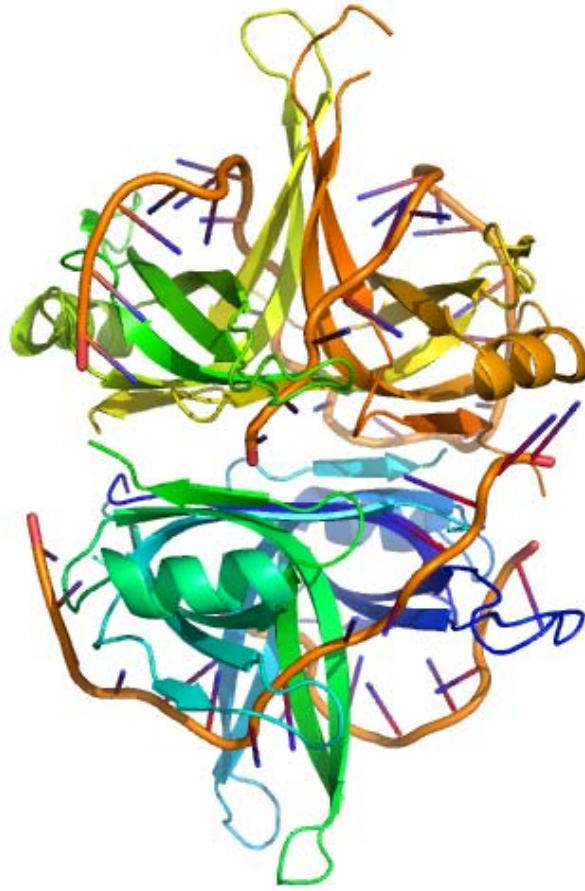
2.3 Proteínas Regulatórias da Atividade da Proteína RecA

Diversas proteínas atuam conjuntamente nos processos de reparo ao DNA, no total mais de 40 proteínas fazem parte desse processo, algumas delas são regulatórias das

atividades da proteína RecA. A modulação das atividades da proteína RecA durante os processos de recombinação homóloga, resposta SOS e/ou mutagênese SOS é efetuada pelas proteínas SSB, LexA, PolV, DinI e RecX e pelos complexos protéicos RecBCD, RecFOR, entre outras. A seguir será descrita a participação da proteína SSB, ou proteína ligadora de DNA simples-fita (SSB).

A presença do ssDNA na célula apresenta vários problemas, o primeiro é o que o ssDNA é termodinamicamente menos estável que o dsDNA, o que leva a formação espontânea de duplex de estrutura secundária de DNA e impede a manutenção dos processos genômicos, como o reparo por exemplo. A proteína SSB liga-se ao ssDNA com alta afinidade e independente da sequência do DNA. A proteína SSB contém ao menos 1 sítio de ligação ao DNA tipo *OB fold*, que consiste de cinco fitas- β arranjadas como um β barril capeada por uma α -hélice e como unidade biológica apresenta-se como um tetrâmero. O tetrâmero é responsável pela ligação ao ssDNA e oligomerização da proteína (MARCEAU, 2012). A proteína SSB mostrou-se importante nos processos de reparo de DNA por que estimula a primeira fase, que é a pré-sináptica polimerização da RecA sobre o ssDNA, e a última fase, que é a reação de trocas de fitas (EGNER et al., 1987; RADDING et al., 1983; EGGLER; LUSETTI; COX, 2003). A figura (2.10) mostra a estrutura da proteína SSB de *E. coli*.

Figura 2.10: Diagrama de fitas da estrutura tetramérica da proteína SSB de *E. coli*.



Fonte: Adaptado de Raghunathan et al. (2000), pdb:1eyg.

A proteína SSB é tipicamente utilizada em ensaios *in vitro* para remover a estrutura secundária do DNA e facilitar a formação dos contínuos filamentos da proteína RecA sobre o ssDNA (KOWALCZYKOWSKI et al., 1987; UMEZU; KOLODNER, 1994; ROY et al., 2009; SHEREDA et al., 2008).

Capítulo 3

Cristalografia de Proteínas

3.1 As Bases da Cristalografia de Macromoléculas

A compreensão de fenômenos biológicos sempre demandam do conhecimento dos processos a nível molecular. Para elucidar a função de uma biomolécula em determinado processo, muitas vezes necessita-se do conhecimento da sua estrutura tridimensional. Uma técnica frequentemente utilizada para obter a estrutura tridimensional de biomoléculas é a cristalografia de raios X. Nesse capítulo serão abordados alguns tópicos relevantes para o entendimento da cristalografia aplicada a macromoléculas, em especial para a sua utilização em proteínas. Para uma discussão detalhada recomenda-se Blow (2002), Giacovazzo (2011).

A cristalografia de macromoléculas é baseada na interação da radiação eletromagnética na faixa dos Raios X, com a matéria no estado cristalino. A interação produz um padrão de difração característico, por consequência, o padrão de difração pode ser interpretado e utilizado para reconstruir a densidade eletrônica da macromolécula.

No processo de difração consideramos os raios X como ondas, essas interagem apenas com os elétrons dos átomos envolvidos. E o espalhamento causado pelos núcleos é negligenciável. Se supormos que o espalhamento é coerente, isto é, o feixe incidente de raios X é espalhado por cada centro espalhador sem perdas de energia, então todas as ondas espalhadas possuem o mesmo comprimento de onda (λ), relativo a onda incidente.

Como consequência, uma interferência coerente é observada entre as ondas espalhadas na mesma direção. Obviamente espalhamentos incoerentes também são observados, porém podem ser ignorados, pois são detectados somente como uma radiação de fundo.

Pode-se definir matematicamente um cristal de macromolécula como uma repetição regular de celas unitárias, no espaço tridimensional. Cada cela unitária contém uma ou várias unidades da macromolécula de interesse. A estrutura tridimensional da macromolécula pode ser obtida por sua densidade eletrônica. A densidade eletrônica não pode ser determinada diretamente, por meio de um experimento de difração de raios X. Entretanto, pode ser obtida de forma indireta por meio do fator de estrutura $F(h, k, l)$, expresso na equação (3.1) (CULLITY; STOCK, 2001).

$$F(h, k, l) = \frac{1}{V} \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) \exp[2\pi i(h.x + k.y + l.z)] dx dy dz \quad (3.1)$$

onde, V é o volume da cela unitária, $\rho(x, y, z)$ a distribuição de densidade eletrônica expressa em elétrons por unidade de volume, $(h; k; l)$ são índices do espaço do padrão de difração, também conhecidos como índices de Miller e $(x; y; z)$ são coordenadas fracionais de um ponto da cela unitária. O padrão de difração somente será observado para os valores inteiros dos índices de Miller, e para aquelas reflexões que obedecem as condições impostas por Bragg e Laue.

O problema geral da determinação de estrutura consiste em determinar a densidade eletrônica, e obter da análise dos seus máximos, as posições atômicas na cela unitária.

A densidade eletrônica $\rho(x, y, z)$ pode ser recuperada por meio transformada de Fourier inversa de $F(h, k, l)$, expressa na equação (3.2) para o caso discreto.

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F(h, k, l) \exp[-2\pi i(h.x + k.y + l.z)] \quad (3.2)$$

A densidade eletrônica da equação (3.2) deve ser interpretada em termos de po-

sições atômicas dos átomos envolvidos. O seu cálculo em cada posição $(x; y; z)$ requer o conhecimento inteiramente de $F(h; k; l)$, pois a soma de $F(h; k; l)$ se estende sobre todos os índices de Miller possíveis de serem obtidos. O $F(h; k; l)$ não é completamente determinado a partir das medidas do padrão de difração, somente seu módulo ($|F(h; k; l)|$) é determinado a partir das intensidades observadas.

Se consideramos fator de estrutura no caso discreto $F(h, k, l)$ em termos do seu módulo, temos que:

$$F(h, k, l) = |F(h, k, l)| \exp[i\alpha(h, k, l)]$$

Podemos re-escrever a equação (3.2), como segue:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h, k, l)| \exp[(-2\pi i(h.x + k.y + l.z) + i\alpha(h, k, l))] \quad (3.3)$$

O cálculo da densidade eletrônica em cada posição (x, y, z) requer o conhecimento inteiramente de $F(h, k, l)$, no qual não é completamente determinado a partir das medidas do padrão de difração, somente seu módulo ($|F(h, k, l)|$) é determinado a partir das intensidades observadas. A fase de espalhamento $\alpha(h, k, l)$ é perdida durante o processo de medida. Desprovido da fase a densidade eletrônica não pode ser corretamente calculada Blow (2002), Giacovazzo (2011). Em cristalografia de macromoléculas este é chamado de problemas das fases, o qual será abordado posteriormente.

3.2 Determinação Experimental do Módulo do Fator de Estrutura $|F(h, k, l)|$

Para um cristal rotacionando com velocidade angular ω , a intensidade integrada $I(h, k, l)$ é obtida por meio da equação (3.4).

$$I(h, k, l) = \frac{\lambda^3}{\omega V_{cela}^3} \left(\frac{e^2}{mc^2} \right) [V_{cristal} I_0 p L A] |F(h, k, l)|^2 \quad (3.4)$$

onde, λ é o comprimento de onda dos raios X utilizado, V_{cela} e $V_{cristal}$ os volumes da cela unitária e do cristal, I_0 a intensidade do feixe incidente, L, p e A são fator de Lorentz, polarização e absorção, respectivamente. O termo $(\frac{e^2}{mc^2})[V_{cristal}I_0p]$ corresponde a intensidade integrada espalhada somente por um elétron, sendo que, c é a velocidade da luz no vácuo, e e m correspondem a carga e massa do elétron, respectivamente. Devido aos termos constantes da equação (3.4), essa pode ser reescrita na forma da equação (3.5).

$$I(h, k, l) = CpLA|F(h, k, l)|^2 \quad (3.5)$$

onde C é uma constante de escala arbitrária que inclui todos os termos constantes do experimento. Logo o fator de estrutura pode ser determinado pela equação (3.6).

$$|F(h, k, l)| = \sqrt{\frac{I(h, k, l)}{CLpA}} \quad (3.6)$$

3.3 O Problema das Fases - $\alpha(h, k, l)$

Como citado acima, a informação a respeito das fases $\alpha(h, k, l)$ são completamente perdidas no processo de medida, os detectores não são capazes de detectar as fases, somente as intensidades são mensuradas, de modo que a densidade eletrônica não é corretamente calculada apenas pelo módulo do fator de estrutura. Se acaso o fator de estrutura fosse completamente determinado a partir das intensidades integradas, equação (3.6), a determinação da densidade eletrônica seria um processo puramente computacional. Diversos métodos foram desenvolvidos para resolver o problema das fases, neste trabalho será descrito resumidamente os fundamentos de dois métodos específicos: o método de Patterson e o de substituição molecular.

Método de Patterson

Historicamente, Arthur Lindo Patterson foi o primeiro a determinar as fases $\alpha(h, k, l)$ para um conjunto de dados de difração (PATTERSON, 1934). O método consiste no fato de que, se um ou mais átomos pesados estiverem presentes na estrutura, a sua

contribuição para o espalhamento dos raios X é preponderante sobre os demais, e pode-se considerar o fator de estrutura inicial e aproximado, considerando somente os átomos pesados. Patterson modificou a função da densidade eletrônica, dada na equação (3.3), substituindo o fator de estrutura $F(h, k, l)$ e a fase $\alpha(h, k, l)$ pelo quadrado das amplitudes do fator de estrutura $|F(h, k, l)|^2$. Essa modificação é expressa por meio da equação (3.7).

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h, k, l)|^2 \cos 2\pi[hu + kv + lw] \quad (3.7)$$

A equação (3.7) é centrossimétrica, enquanto a densidade eletrônica possui picos nas respectivas posições atômicas, a equação (3.7) possui picos correspondendo a vetores interatômicos. No espaço de Patterson $P(u, v, w)$ também periódico e definido por uma cela unitária idêntica a cela unitária do cristal, são definidas coordenadas genéricas $(u; v; w)$, de forma que qualquer par de átomos no cristal, localizados em (x_1, y_1, z_1) e (x_2, y_2, z_2) , serão localizados por um máximo no mapa de Patterson com coordenadas $u = x_1 - x_2; v = y_1 - y_2; w = z_1 - z_2$ (PATTERSON, 1934).

A função de Patterson fornece um mapa de vetores interatômicos, correlacionado com as posições atômicas absolutas. Um máximo do mapa de Patterson é proporcional ao número de elétrons do átomo, implicando que a presença de átomos pesados facilita a localização de suas coordenadas na estrutura e as respectivas fases. A intensidade de um pico no mapa de Patterson é proporcional ao produto dos números atômicos (Z) dos átomos correspondentes ao pico. A constante de proporcionalidade (k) necessária para relacionar um pico ao correspondente produto dos números atômicos é obtida no pico na origem $P(0; 0; 0)$, cuja intensidade é dado por:

$$P(0; 0; 0) = k \sum_{j=1}^N Z_j^2 \quad (3.8)$$

A maior importância da equação (3.8) é que os correspondentes vetores entre átomos pesados aparecem com seus pesos proporcionais ao quadrado de seus números atômicos, sendo facilmente indentificados entre os picos correspondentes a átomos pesado-leve e leve-leve. Uma vez que o mapa de Patterson é calculado e interpretado, determina-se

as posições atômicas, por meio da equação (eq. (3.3) na página 43), determina-se as fases iniciais $\alpha(h; k; l)$.

Método de Substituição Molecular

Nas próximas páginas serão descritos alguns passos para a obtenção de um modelo estrutural para a estrutura de uma proteína através do Método de Substituição Molecular.

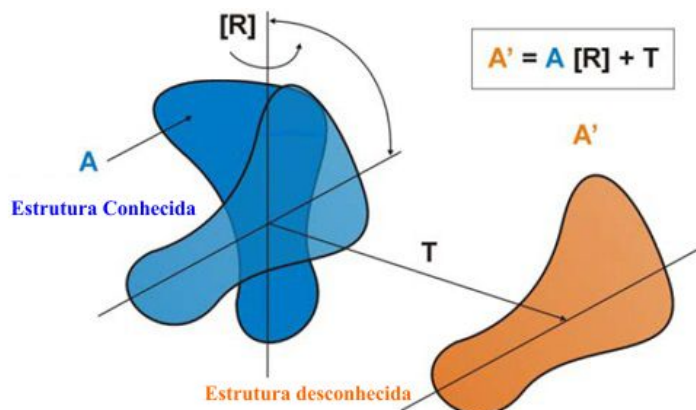
O método de substituição molecular baseia-se na comparação do mapa de Patterson da estrutura cristalina de uma proteína homóloga conhecida, com mapa de Patterson obtido experimentalmente para a proteína em estudo. A estrutura tridimensional de proteínas é mais conservada durante a evolução do que a sequência de aminoácidos (KOONIN; GALPERIN, 2003). Quando dispõe-se de um modelo estrutural pré-determinado e um percentual de homologia de aminoácidos superior que 25 pode ser utilizado para se determinar as fases dos fatores de estrutura da estrutura cristalina estudada (CROWTHER; BLOW, 1967). Criando um modelo estrutural inicial da estrutura da proteína estudada, que posteriormente será refinado.

Muitas vezes conhecemos a estrutura de uma proteína e queremos determinar sua estrutura em um diferente grupo espacial. Outras vezes, têm-se razões suficientes para acreditar que a conformação da proteína desconhecida é muito similar a outra resolvida anteriormente (alta homologia sequencial). Nos casos acima mencionados, seis variáveis, três rotacionais e três translacionais, descreveram a transformação de um conjunto de coordenadas para o outro. Se denotarmos de \mathbf{A} o conjunto de vetores representando a molécula original e \mathbf{A}' o conjunto de vetores da molécula transladada, a transformação de coordenadas pode ser expressa por meio da equação (3.9).

$$\mathbf{A}' = \mathbf{A}[\mathbf{R}] + \mathbf{T} \quad (3.9)$$

onde $[\mathbf{R}]$ é a matriz que rotaciona as coordenadas \mathbf{A} para a nova orientação, e \mathbf{T} é um vetor de translação. A figura (3.1) ilustra o procedimento realizado no método de substituição molécula.

Figura 3.1: Procedimento realizado no método de substituição molecular para determinação das fases iniciais. Um modelo estrutural conhecido é transladado e rotacionado para a posição da proteína na qual a estrutura deve ser determinada.



Fonte: Adaptado de Ripoll e Cano (2016).

Computacionalmente seria possível encontrar simultaneamente seis variáveis, 3 rotacionais e 3 translacionais, que tornaria a diferença entre os fatores de estrutura observado F_{obs} e calculado F_{cal} mínima, no entanto, na prática esse processo é uma tarefa complicada e demanda de longos tempos computacionais, mesmo com os rápidos computadores disponíveis. A solução do problema computacional foi desenvolvida por Rossmann and Blow (ROSSMANN; BLOW, 1962), eles detectaram a possibilidade de determinar a orientação de subunidades similares na cela unitária, sem o conhecimento do vetor de translação. O procedimento adotado utiliza-se da função de Patterson, equação (3.7).

Função de Rotação [R] de Rossmann-Blow

Considere duas moléculas idênticas, dispostas espacialmente em diferentes orientações. A função de Patterson de uma estrutura deste tipo será constituído por três conjuntos de vetores: O primeiro conjunto de vetores será um conjunto de inter-vetores de Patterson de uma molécula, estes são vetores interatômicos escritos no interior da molécula. O segundo conjunto será inter-vetores de Patterson da outra molécula, porém estão rotacionados devido a diferença de orientação. O terceiro conjunto serão intra-vetores de Patterson, ou conjunto de vetores interatômicos que podem ser escritos de uma molécula

em relação a outra.

Supondo que a função de Patterson de uma molécula é rotacionada sobre a outra, não deve existir nenhuma concordância se o conjunto de auto-vetores de Patterson possuir uma diferente orientação relativo as duas moléculas. O máximo da rotação será numa condição de concordância ou de perfeita sobreposição entre elas. A função que descreve a sobreposição foi determinada por Rossmann and Blow (ROSSMANN; BLOW, 1962), denominada de função de rotação e é definida por meio da equação (3.10).

$$RF(\mathbf{R}) = \int_U \mathbf{P}_1(\mathbf{x})\mathbf{P}_2(\mathbf{R}\mathbf{x})d\mathbf{V} \quad (3.10)$$

onde P_1 e P_2 são funções de Patterson, \mathbf{R} é o operador de rotação que descreve o sistema de coordenadas de P_2 para P_1 , U é o volume de integração, usualmente esférico e centrado na origem. Se um máximo for encontrado para a função de rotação, provavelmente a orientação da molécula conhecida está corretamente orientada em relação a molécula cuja estrutura é desconhecida.

Considerando P_1 e P_2 a partir da equação (3.7), a função de rotação pode ser re-escrita na forma:

$$RF(\mathbf{R}) = \frac{U}{V^3} |\mathbf{F}(\mathbf{h})|^2 |\mathbf{F}(\mathbf{p})|^2 \mathbf{G}_{\mathbf{hp}} \quad (3.11)$$

onde $|F(h)|$ e $|F(p)|$ são os módulos dos fatores de estrutura das moléculas conhecida e desconhecida, respectivamente. A função $G_{\mathbf{hp}}$ é conhecida como função de interferência, expressa por meio da equação (3.12).

$$G_{hp} = \int_U e^{-2\pi i(h+p[R])u} du = \frac{3(\sin 2\pi HR - 2\pi HR \cos 2\pi HR)}{(2\pi HR)^3} \quad (3.12)$$

onde H é o módulo do vetor no espaço recíproco $H = h + p[R]$. G_{hp} assume valor máximo de 1 quando $h + p[R] = 0$ e rapidamente decresce e oscila em torno de zero para $-\infty < h + p[R] > \infty$.

A Função Translação

Uma vez que a correta orientação do modelo estrutural é determinada, a sua correta posição na cela unitária precisa ser determinada. Para tal procedimento é verificada a correlação entre as intensidades observadas e os intra-vetores de Patterson da unidade relacionada por simetria. Se o modelo estrutural estiver corretamente posicionado, a função translação deve conter picos nos valores correspondentes aos vetores de translação entre as moléculas relacionadas por simetria. Existem diversas descrições matemáticas para a função de translação, porém de forma análoga a função de rotação, pode-se definir a função de translação como:

$$TF(\mathbf{T}) = \int_U P_1(\mathbf{x})P_2(\mathbf{T}\mathbf{x})d\mathbf{V} \quad (3.13)$$

onde P_1 e P_2 são funções de Patterson, \mathbf{T} é o operador de translação que descreve o sistema de coordenadas de P_2 para P_1 .

As funções de translação constroem modelos da proteína desconhecida movendo apropriadamente o modelo estrutural dentro da cela unitária. Para cada translação \mathbf{x} as amplitudes dos fatores de estrutura são calculados $|F_{cal}(h, k, l)|^2$, e a correlação com as amplitudes dos fatores de estrutura observados $|F_{obs}(h, k, l)|^2$ são também determinadas. Crowther e Blow representaram a correlação pela função de translação \mathbf{T} , expressa por meio da equação (3.14).

$$T(\mathbf{x}) = |F_{obs}(h, k, l)|^2|F_{cal}(h, k, l)|^2 \quad (3.14)$$

Para duas moléculas na cela unitária, com fatores de estrutura F_{m1} e F_{m2} , Crowther e Blow expandiram a equação (3.14) em termos de vetores do espaço recíproco e obtiveram a chamada função T de translação, expressa por meio da equação (3.16).

$$T(\mathbf{r}) = \sum_{h,k,l} (|F_{obs}(h, k, l)|^2) F_{m1} F_{m2}^* \exp(-2\pi i(h(\mathbf{r}_1 - \mathbf{r}_2) + k(\mathbf{r}_1 - \mathbf{r}_2) + l(\mathbf{r}_1 - \mathbf{r}_2))) \quad (3.15)$$

o termo da equação acima:

$$F_{m1}F_{m2} \exp(-2\pi i(h(\mathbf{r}_1 - \mathbf{r}_2) + k(\mathbf{r}_1 - \mathbf{r}_2) + l(\mathbf{r}_1 - \mathbf{r}_2))) \quad (3.16)$$

refere-se ao intra-vetores entre as duas moléculas. Os respectivos auto-vetores de cada molécula podem contribuir com um ruído no cálculo, de modo que o $|F_{cal}(h, k, l)|^2$ deve ser subtraído das contribuições dos auto-vetores de cada modelo estrutural para o cálculo fator de estrutura, para ter certeza dessa subtração e a prevalência somente dos intra-vetores, os auto-vetores dos modelos estruturais são subtraídos de $|F_{cal}(h, k, l)|^2$, essa subtração é expressa por meio da equação (3.16).

$$|F_{cal}(h, k, l)|_{cross}^2 = |F_{cal}(h, k, l)|^2 - k(F_{m1} - F_{m2}) \quad (3.17)$$

onde k é fator de escala, dado por $k = \sum \frac{|F_{obs}(h, k, l)|^2}{|F_{cal}(h, k, l)|^2}$.

Índices de Qualidade de Coleta de Dados e Refinamento

Completeza

A completeza é o parâmetro que especifica o número de reflexões cristalográficas únicas medidas em um conjunto de dados, e é expressa como uma porcentagem do número total de reflexões únicas presentes em uma faixa de resolução específica. Como a maioria dos dados são, em última análise utilizados no cálculo da densidade eletrônica ou mapas Patterson via transformada de Fourier, a completeza dos dados é muito importante. A intensidade de uma simples reflexão é um fator na soma da transformada de Fourier, se essa informação não é coletada, distorções significativas na densidade eletrônica e nos mapas de Patterson vão ocorrer e esconder aspectos importantes da estrutura.

Multiplicidade

A multiplicidade das intensidades medidas é definida como o número total de intensidades medidas dividido pelo número total de reflexões únicas medidas, em outras

palavras esse parâmetro avalia o número de vezes que o ponto de difração ou seu equivalente foram sistematicamente medidos (CULLITY; STOCK, 2001).

Fator R_{merge} e a Razão sinal-ruído $I/\sigma(I)$

A qualidade dos dados é usualmente julgada usualmente julgadas pelo fator global R_{merge} , expresso por meio da equação (3.18), seja baseado em intensidade ou no $|F^2|$, este fornece uma média geral da propagação das intensidades de múltiplas reflexões simetricamente equivalentes. Se o cálculo de R_{merge} for realizado sobre um número grande de reflexões é um bom indicativo da qualidade dos dados coletados, pois se esse número for alto pode indicar que os dados possuem muitos erros devido a fatores como absorção dos raios X, espalhamento anômalo e determinação incorreta dos parâmetros cristalinos, idealmente o valor de R_{merge} deve ser $< 5\%$.

$$R_{merge} = \sum_{hkl} \sum_I \frac{|I_{hkl,I} - \langle I_{hkl} \rangle|}{\sum_{hkl} \langle I_{hkl} \rangle} \quad (3.18)$$

Informação complementar ao R_{merge} sobre a qualidade dos dados é dada por meio da razão $[\sum I_{hkl} / \sum \sigma(I_{hkl})]$, onde $[\sigma(I_{hkl})]$ é o desvio padrão das intensidades, e esse seja apropriadamente estimado. Quanto maior a razão $I/\sigma(I_{hkl})$ mais informação se tem dentro dos dados coletados, em princípio, a coleta de dados contém alguma informação para valor maior que 1. Na prática usa-se a razão $I/\sigma(I_{hkl})$ para efetuar o corte de resolução, onde o limite de resolução aceito cai abaixo de $I/\sigma(I_{hkl})=2,0$ (DAUTER, 1999; EVANS; MURSHUDOV, 2013).

Fator R_{work} e R_{free}

No refinamento de estruturas de proteínas é calculado um modelo com diversos parâmetros variáveis, cujos valores são calculados de modo a fornecer a melhor concordância as melhores concordâncias entre os fatores de estrutura observados e calculados. Um método conveniente que permite determinar a concordância é por meio dos fatores R_{work} e R_{free} , esses são utilizados durante o refinamento para indicar a consistência interna entre

os valores de fator de estrutura calculado e observado. Os dois são determinados de forma idêntica, expressos por meio da equação(3.19), no entanto o valor de R_{free} é somente calculado para um determinado conjunto de reflexões, onde 5 a 10% do total das reflexões são excluídas do refinamento, e sobre elas é determinado sem tendenciosidade a probabilidade de o modelo proposto estar mais correto ou não a cada ciclo de refinamento.(KARPLUS; DIEDERICHS, 2012).

$$R_{work,free} = \frac{\sum_{hkl} (|F_{obs}| - |F_{cal}|)}{\sum_{hkl} |F_{obs}|} \quad (3.19)$$

3.4 Superfície de Potencial Eletrostático

Mecânica Molecular também conhecida como cálculos de campo de força, permite calcular a energia e otimizar geometrias moleculares, bem como explorar potencial eletrostático de superfície. Se existe um modelo estrutural pré determinado, como o modelo cristalográfico, por exemplo, as equações de campo de força dão conta de descrever a energia potencial eletrostática para a molécula estudada. As interações eletrostáticas possuem um importante papel na conformação tridimensional, estabilidade, atividade enzimática, energias de ligação e interações proteína-proteína e proteína-ligante, essas interações surgem das seguintes fontes: da presença de cargas locais, da polarização resultante do distribuição não-esférica de densidade de elétrons em torno de átomos, da redistribuição de elétrons provocada por campos elétricos locais de polarização e da reorientação de grupos polares em moléculas de soluto e de solvente em resposta a campos elétricos (SHEINERMAN; NOREL; HONIG, 2000). A descrição do potencial eletrostático na região de superfície se torna muito útil para estudar essas interações ou descrever comportamentos biológicos observados por outras metodologias.

O modelo matemático conhecido como AMBER (*Assisted Model Building with Energy Refinement*) desenvolvido em 1981 como um programa para construir modelos de moléculas e suas interações, é um dos campos de força utilizados para estudos de proteínas, partir desse modelo é possível obter a energia potencial eletrostática de uma

molécula, sendo esse uma superposição de termos de energia potencial de átomos ligados e não ligados (CORNELL et al., 1995), expresso por meio da equação (3.20).

$$\Phi(r, \theta, \phi) = \sum_{lig.} \frac{1}{2} k_r (r - r_0)^2 + \sum_{ang.} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{tor.} \frac{1}{2} k_\phi [\cos(n\phi + \delta) + 1] + \sum_{pares} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right] \quad (3.20)$$

onde o primeiro é o termo energético de estiramento, que descreve as variações de energia potencial em função das vibrações dos átomos, e o segundo é o termo energético de torção angular, que descreve mudanças na energia potencial em função do ângulo formado da ligação entre os átomos. O terceiro termo é o termo energético referente aos ângulos diédricos, todas essas interações são respectivas aos pares de átomos ligados. Os termos k_r , k_θ e k_ϕ são as constantes que caracterizam a força de ligação química. Os termos dependentes de r descrevem as interações entre pares de átomos não ligados covalentemente. Os termos que variam com r_{ij}^{12} e r_{ij}^6 descrevem o potencial do tipo Lennard-Jones, repulsivo e atrativo, respectivamente. O termo que varia com r_{ij} descreve variações na energia potencial por meio do potencial coulombiano entre átomos carregados (CORNELL et al., 1995).

A partir do modelo cristalográfico, o potencial $\Phi(r, \theta, \phi)$ pode ser reconstruído utilizando as coordenadas atômicas (posições e ângulos) do arquivo .pbd gerado. O servidor APBS (DOLINSKY et al., 2007) gera uma superfície de potencial molecular baseado na estrutura do modelo, podendo ser posteriormente utilizada para estudos de interações proteína-proteína e proteína-ligante.

Capítulo 4

Material e Métodos

Para o estudo cristalográfico da estrutura de proteínas, é necessário obter essa com alto grau de pureza, também são necessários processos de purificação que separem a proteína de interesse por suas propriedades físicas e químicas. Neste trabalho foi adotado o sistema de expressão heteróloga para expressar a proteína HsRecA, que posteriormente foi purificada por técnicas cromatográficas de alta resolução e especificidade. Os métodos gerais de manipulação de DNA e proteína estão descritos neste capítulo, bem como os ensaios de atividade biológica, cristalização, coleta de dados de raios X, e as ferramentas de bioinformática utilizadas na solução da estrutura da proteína HsRecA.

4.1 Métodos Gerais para DNA

Construção Plasmídial e Clonagem

O fragmento de DNA contendo o gene *recA* de *H. seropedicae* foi obtido a partir do plasmídeo pAETWT-HMK (GALVÃO et al., 2012), esse gentilmente cedido pela Prof. Dra. Carolina Weigert Galvão do Laboratório de Biologia Molecular Microbiana da UEPG. O plasmídeo foi digerido com as enzimas *NdeI-BamHI* e o fragmento clonado diretamente no vetor pET21a (Novagen) digerido com as mesmas enzimas, originando a construção denominada de pAEW989 que codifica a proteína HsRecA nativa. Ao produto da restrição foi adicionado o tampão de amostra (0,25 % azul de bromofenol, 0,25

% xileno cianol FF e 30 % de glicerol) e analisado em gel de agarose 0,8% (w/v) em tampão TAE (10 mmol EDTA, 40 mmol Tris-AC (80% cation, pH 8), gel foi posteriormente corado em 0,5 mg/mL de brometo de etídeo (Invitrogen), e visualizado sob luz ultravioleta. A integridade do fragmento clonado foi confirmada por sequenciamento do plasmídeo pAEW989.

cssDNA e ldsDNA

O DNA circular simples-fita (cssDNA) do bacteriófago M13mp18 e o DNA linear dupla fita super-helicoidal (ldsDNA) M13mp18 foram gentilmente cedidos pela equipe do Laboratório do Prof. Mike M. Cox, Universidade de Wisconsin-Madison USA. A concentração de DNA foi expressa em μM de nucleotídeos ($\mu\text{M nt}$). A concentração foi determinada pela absorvância em 260 nm empregando a seguinte relação: $D.O_{260} = 1,0$ corresponde a 108 μMnt (0,036 $\mu\text{g}/\mu\text{L}$) ssDNA e 151 μMnt (0,05 $\mu\text{g}/\mu\text{L}$) ldsDNA (SAMBROOK; RUSSELL, 2001).

Transformação Bacteriana por Eletroporação

Foram utilizados as estirpes *E. coli* DH5 α (Invitrogen) e STL2669 [AB1157 contendo ($\Delta\text{recA-srlR}$)306::Tn10 TetR, *xonA2* (*exoI*⁻), gentilmente cedida pelo laboratório do prof. Cox], para armazenamento dos plasmídeos e expressão, respectivamente. A estirpe STL2669 foi modificada de forma a ser deficiente de nucleases e contribui para o processo de purificação da proteína RecA que será descrito posteriormente.

Preparo de células eletrocompetentes

Foram plaqueadas células de *E. coli* DH5 α ou STL2669 em meio Luria-Bertani Agar (LA) (10 g/L Triptona, 5 g/L Extrato de levedura, 10 g/L NaCl, 1,1 mL/L de NaOH (1 M), 15 g/L Ágar) (SAMBROOK; RUSSELL, 2001), uma colônia isolada foi replaqueada e este procedimento ocorreu por 3 vezes com sucessivas colônias isoladas, posteriormente o pré-inóculo foi preparado a partir da inoculação de uma colônia de células em 5 mL de meio Luria-Bertani (LB) (10 g/L Triptona, 5 g/L Extrato de levedura, 10 g/L NaCl e 1,1

mL/L de NaOH (1 M)) e incubadas (37 °C a 150 rpm por 3 horas). O cultivo foi realizado a 37°C e 130 rpm até a saturação da cultura. Quinhentos microlitros desta cultura foram utilizados como inóculo em 100 mL de meio LB. As células foram cultivadas a 37 °C e 130 rpm até atingir a densidade ótica em 600 nm ($D.O_{600}$) entre 0,5-0,7. Em seguida, a cultura foi transferida para tubos plásticos de 50 mL e centrifugadas (5000xg a 4 °C por 5 min). As células foram lavadas duas vezes, com 40 mL de água Milli-Q estéril gelada (4 °C). Após as lavagens, o sedimento foi ressuspensão em 30 mL de glicerol 10% mantido a (4 °C) e centrifugado (5000xg a 4 °C por 10 minutos). Após o descarte do sobrenadante, o sedimento de células foi ressuspensão na solução de glicerol remanescente nas paredes do tubo (cerca de 1 mL), a suspensão de células foi aliqüotada (100 μ L) e armazenada a -80 °C (SAMBROOK; RUSSELL, 2001).

Transformação

Aliqüotas de 100 μ L de células eletrocompetentes foram descongeladas em banho de gelo e utilizadas para eletroporação. Em cada tubo foi adicionado 2 μ L de DNA plasmídial. Após uma leve homogenização da mistura, essa foi transferida para uma câmara de eletroporação e submetida a um pulso elétrico de 1,8 kV durante 5,8 ms, no aparelho Bio Rad MicroPulser, a seguir, as células foram transferidas para frascos de 10 mL contendo 1 mL de meio LB. Aliqüotas de 300 μ L de células foram plaqueadas em meio LA contendo 100 μ g/mL do antibiótico Ampicilina (Amp) para a seleção das células transformantes. As placas foram incubadas a 37°C por 12-24 h. Aquelas colônias que continham o plasmídeo recombinante foram selecionadas e crescidas novamente em 1 mL meio LB contendo 100 μ g/mL de Amp durante 8 - 12 h. A presença do plasmídeo contendo o fragmento de DNA foi confirmada por meio da extração de DNA plasmídial, e posição de migração na eletroforese em gel de agarose 0,8%. As células contendo o plasmídeo recombinante foram centrifugadas e estocadas na presença de glicerol 20% a -80 °C (SAMBROOK; RUSSELL, 2001).

4.2 Purificação de Proteínas

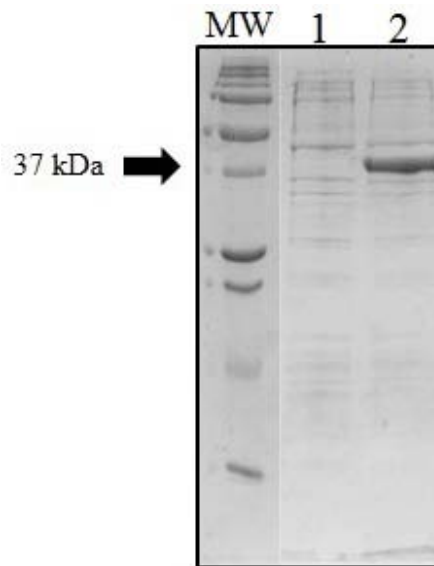
Expressão da Proteína

Para a expressão em larga escala da proteína HsRecA nativa um volume de 250 mL de pré-inóculo foi preparado a partir de uma colônia ou uma alíquota do estoque de bactérias, mantidas em glicerol em meio de cultura Luria-Bertani (LB), suplementada com Amp (100 $\mu\text{g}/\text{mL}$). O pré-inóculo foi mantido durante 8-12h sob agitação de 250 rpm a temperatura de 37 °C .

O pré-inóculo foi então inoculado em um total de 10 L de meio de cultura LB contendo 100 $\mu\text{g}/\text{mL}$ de Amp. As células foram mantidas a 37 °C sob agitação de 180 rpm até a DO_{600} atingir 0,6. Neste momento, a expressão da proteína alvo foi induzida pela adição do agente indutor IPTG 0,5 mmol/L por 4 h, foi mantida a temperatura de 37 °C e agitação de 180 rpm.

A fim de confirmar a superexpressão da proteína de interesse volumes normalizados para $DO_{600} = 0,3$, das frações antes e após a indução foram coletados para análise por eletroforese em gel de poliacrilamida desnaturante (SDS-PAGE; (LAEMMLI, 1970)). A figura (4.1), apresenta um gel de poliacrilamida SDS-PAGE 12,5% representativo da superexpressão.

Figura 4.1: Superexpressão da HsRecA a temperatura de 37 °C por 4 horas. MW - Padrão de massa molecular (Bio-Rad), (1) Fração não induzida e (2) fração induzida. A banda mais intensa do gel corresponde à proteína HsRecA, localizada próxima ao peso molecular de 37 kDa.



Fonte: O autor.

As células contendo a proteína superexpressa foram recuperadas por centrifugação a 3985 *g* por 20 minutos a 4 °C, resfriadas em nitrogênio líquido e armazenadas a -80 °C até o momento de uso.

Purificação da Proteína

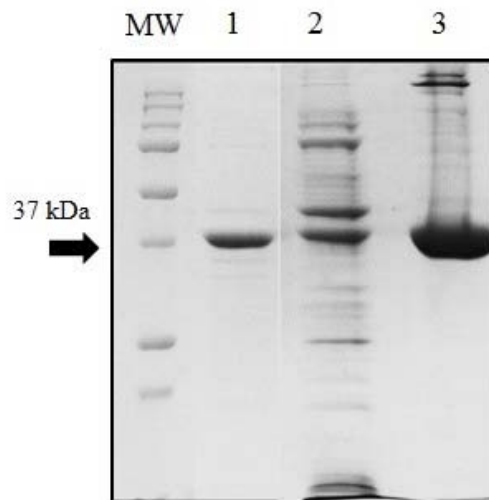
Para purificação da proteína HsRecA um protocolo foi desenvolvido com base em procedimentos descritos anteriormente (COX; LEHMAN, 1981; SHAN; COX; INMAN, 1996). As modificações utilizadas neste trabalho estão abaixo descritas.

As proteínas EcRecA e EcSSB foram gentilmente cedidas pelo laboratório do Prof. Dr. Michael M. Cox (Universidade de Wisconsin-Madison, USA), o processo de purificação e determinação da concentração podem ser consultados nos protocolos previamente descritos (COX; LEHMAN, 1981; SHAN; COX; INMAN, 1996).

Lise das células e precipitação seletiva de proteínas

Todas as etapas de purificação foram realizadas a 4 °C para evitar a desnaturação da proteína. O sedimento de células (aproximadamente 13 g) contendo a proteína HsRecA foi descongelada durante 8-12 h em uma solução de lise contendo 250 mmol/L Tris-AC (80% cátion, pH 7.8) e 25% (w/v) sacarose, onde o volume foi ajustado para 20% (w/v) e mantida sob agitação. As células foram então congeladas novamente em nitrogênio líquido e descongeladas em banho térmico a 37 °C, esse procedimento se repetiu por duas vezes. Em seguida, foi adicionado solução de 2,5 mg/mL Lisozima em 250 mmol/L Tris-AC (80% cation, pH 7.8) e agitado por 1 h. Após agitação foi adicionado 0,02 mL de 500 mmol/L EDTA por mL de volume final de solução de lise adicionado de células, e agitado por outros 30 min. A figura (4.2), mostra proteínas presentes no precipitado e sobrenadante que foram visualizadas após eletroforese em gel de poliacrilamida SDS-PAGE 12,5%.

Figura 4.2: Lise celular. MW - Padrão de massa molecular (Bio-Rad), (1) Fração induzida, (2) precipitado da lise celular e (3) sobrenadante da lise celular . A banda mais intensa do gel corresponde à proteína HsRecA.



Fonte: O autor.

A suspensão contendo o lisado foi sonicado por 20 min, utilizando ciclos de 30 s *on* / 30 s *off*, 60% *output* da potência máxima do sonicador, e então centrifugado por (38000xg a 4 °C por 1h30min) para remover debris celulares.

Nas próximas etapas foram utilizadas a polietilenoimina e o sulfato de amônio. A polietilenoimina é um polímero positivamente carregado que ao interagir com proteínas ácidas ou com DNA induz a precipitação de ambos. A polietilenoimina ao interagir com o DNA também induz a precipitação de proteínas ligadas ao DNA. E o sulfato de amônio induz a precipitação das proteínas por *salting out*.

Proteínas que se ligam ao DNA e a proteína HsRecA foram então precipitadas do sobrenadante com a adição de 0,111 mL de 5% (w/v) polietilenoimina (pH ajustado para 7,5 com 12 mol/L HCl) por mL de lisado e incubado por 1h sob leve agitação e centrifugado (13000xg a 4 °C por 15 min).

O material precipitado foi lavado com tampão-R (20 mmol/L Tris-AC (80% cá-tion, pH 7.8), 10% glicerol, 0,1 mmol/L EDTA e 1 mmol/L ditioneitol (DTT) fresco) mantido em 15 min de agitação e centrifugado (13000xg a 4 °C por 15 min). A proteína HsRecA foi extraída do precipitado pela adição de 10 mL de tampão-R contendo 300 mmol/L sulfato de amônio por grama de material precipitado, essa etapa foi realizada sob leve agitação por 30 min e centrifugado (13000xg a 4 °C por 15 min). Esse procedimento de extração se repetiu por duas vezes.

A resina BioRex-70 (BioRad) (pH ajustado para 7,5 utilizando 12 mol/L HCl) foi equilibrada com tampão-R contendo 300 mmol/L sulfato de amônio, e então utilizada para remover a polietilenoimina ligada a proteína HsRecA. A solução contendo a proteína HsRecA foi então homogeneizada por agitação com a resina por 2 h e recuperada via filtração por gravidade em filtros de 115 mL Nalgene.

Ao filtrado foi adicionado 0,28 g de sulfato de amônio por mL de solução e mantida sob agitação durante 8-12 h. A solução foi novamente centrifugada (25000xg a 4 °C por 15 min). Esse processo precipita todas as proteínas presentes na solução mantendo sua conformação nativa.

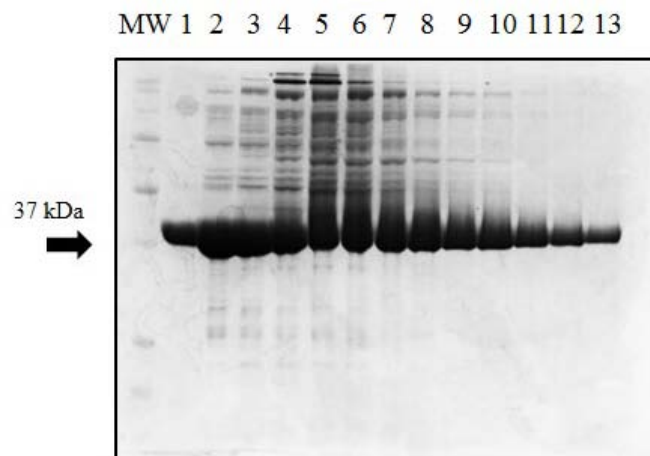
Cromatografia na coluna DEAE Sepharose

A coluna DEAE Sepharose *fast flow column* (GE Healthcare) foi utilizada como primeiro passo cromatográfico de purificação. Essa é constituída de resina para cromato-

grafia de troca iônica fraca e apresenta carga elétrica positiva. A proteína HsRecA possui carga negativa (pI de 5,5, determinado teoricamente pela sequência de aminoácidos) em pH 7,5, ligando-se a coluna e podendo ser eluída por meio de um gradiente de sal.

coluna utilizada possui volume de 120 mL foi previamente equilibrada com 5 volumes de coluna (VC) de tampão-R contendo 50 mmol/L KCl e 1 mmol/L DTT fresco. O material precipitado após adição de sulfato de amônio foi resuspenso no mesmo tampão de equilíbrio da coluna, e injetado num fluxo máximo de 1 mL/min, ocasionalmente o fluxo era reduzido para manter a pressão na coluna constante. Após a aplicação da amostra, a coluna foi lavada com tampão de equilíbrio (1,5 VC), num fluxo de 1 mL/min. A proteína HsRecA foi eluída da coluna com 7 VC de um gradiente linear de KCl (0,05 - 1 mmol/L) em tampão R. As frações de 10 mL eluídas durante o gradiente linear de sal foram coletadas e analisadas por eletroforese SDS-PAGE 12,5%. Um gel representativo da cromatografia na coluna DEAE é mostrado na figura (4.3). Aquelas frações eluídas que apresentavam similar composição de contaminantes foram agrupadas.

Figura 4.3: Análise das frações coletadas durante a cromatografia na coluna DEAE. MW - Padrão de massa molecular (Bio-Rad), (1-13) frações eluídas.



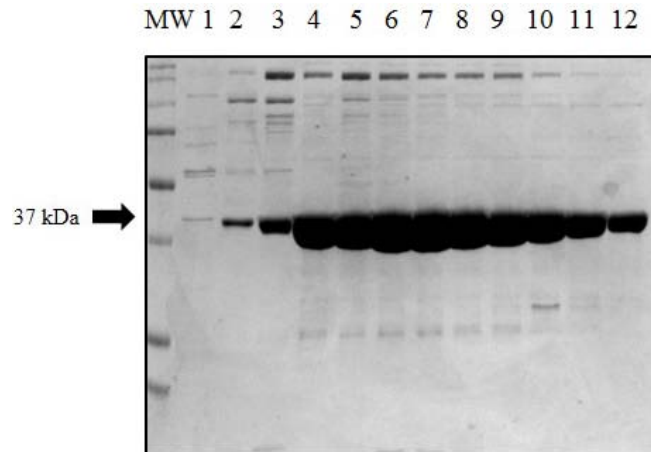
Fonte: O autor.

Cromatografia na coluna cerâmica HAP

A coluna cerâmica de hidroxapatita (HAP) (BIO-RAD) é formada por fosfato de cálcio e utilizada na separação de proteínas devido as naturezas físico-químicas da matriz. A hidroxapatita possui dois tipos de sítios de ligação para proteínas, um positivamente carregado possuindo um grupo de Cálcio no sítio e outro negativo possuindo um grupo fosfato. Estes sítios de ligação estão uniformemente distribuídos sobre a matriz cerâmica. A proteína alvo interage sobre troca catiônica com os grupos fosfatos e/ou afinidade metálica via os átomos de Cálcio.

A coluna utilizada possui 20 mL e foi equilibrada com 2 VC de tampão-P (20 mmol/L fosfato de potássio, 10% glicerol, 0,1 mmol/L EDTA e 1 mmol/L DTT fresco). As frações agrupadas após a coluna DEAE foram dialisadas contra o tampão-P (2 x 2L por 4 h e uma vez com 2 L durante 8-12 h e injetadas na coluna num fluxo de 1 mL/min. Após a aplicação da amostra, a coluna foi lavada com tampão-P de equilíbrio (2 VC), num fluxo de 1 mL/min. A proteína HsRecA foi eluída da coluna com 10 VC de tampão-P contendo um gradiente linear de fosfato de potássio (0,02 - 1 mol/L). As frações de 4 mL eluídas durante o gradiente linear de sal foram coletadas e analisadas por eletroforese SDS-PAGE 12,5%. A figura (4.4), apresenta um gel representativo da cromatografia na coluna HAP. Aquelas frações eluídas que apresentavam similar composição de contaminantes foram agrupadas.

Figura 4.4: Análise das frações coletadas durante a cromatografia na coluna HAP. MW - Padrão de massa molecular (Bio-Rad), (1-12) frações eluidas.



Fonte: O autor.

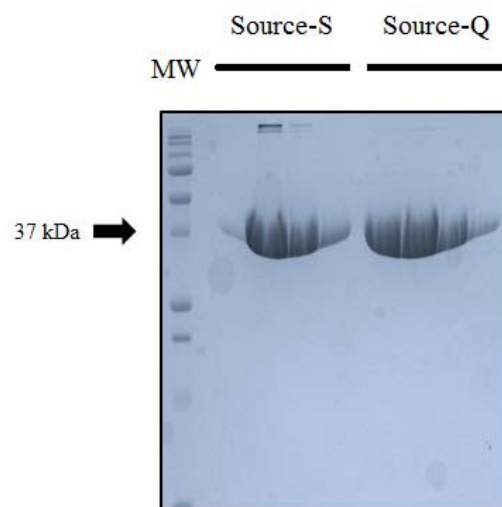
Cromatografia na coluna Source 15-S e Source 15-Q

Diferentemente da coluna DEAE, as colunas Source 15-Q e Source 15-S (GE Healthcare) são fortes trocadoras de ânions e cátion, respectivamente. Em ambos os casos a proteína alvo pode ser eluída por meio de um gradiente de sal. No caso da proteína HsRecA devido sua carga negativa liga-se fortemente na coluna Source 15-Q e não à coluna Source 15-S. Essa é uma etapa importante do processo de purificação pois seleciona os contaminantes devido sua natureza de carga elétrica.

As colunas Source 15-Q e Source 15-S possuem volume de 20 mL e foram equilibradas com 2 VC de tampão-R contendo 50 mmol/L KCl e 1 mmol/L DTT. As frações similares provenientes da coluna cerâmica HAP foram então dialisadas contra tampão-R contendo 50 mmol/L KCl e 1 mmol/L DTT (2 x 2 L por 4 h e uma vez com 2 L durante 8-12 h), então aplicadas na coluna Source 15-S a um fluxo constante de 0,5 ml/min. A proteína HsRecA não é retida na coluna e é coletada na fração que passa durante a injeção. A fração de injeção é então aplicada diretamente na coluna Source 15-Q mantendo mesmo fluxo de 0,5 ml/min. Após a aplicação da amostra, a coluna foi lavada com tampão-R

contendo 50 mmol/L KCl e 1 mmol/L DTT fresco de equilíbrio (2 VC). A proteína Hs-RecA foi eluída da coluna com 10 VC de tampão-P contendo um gradiente linear de KCl (0,05 - 1 mol/L). As frações de 4 mL eluídas durante o gradiente linear de sal foram coletadas e analisadas por eletroforese SDS-PAGE 12,5%, figura (4.5), apresenta um gel representativo da cromatografia nas colunas Source 15-S e Source 15-Q.

Figura 4.5: Análise das frações coletadas durante a cromatografia nas colunas Source-15S e Source-15Q. MW - Padrão de massa molecular (Bio-Rad), as frações eluídas de cada coluna estão indicadas abaixo da identificação de cada coluna no gel.



Fonte: O autor.

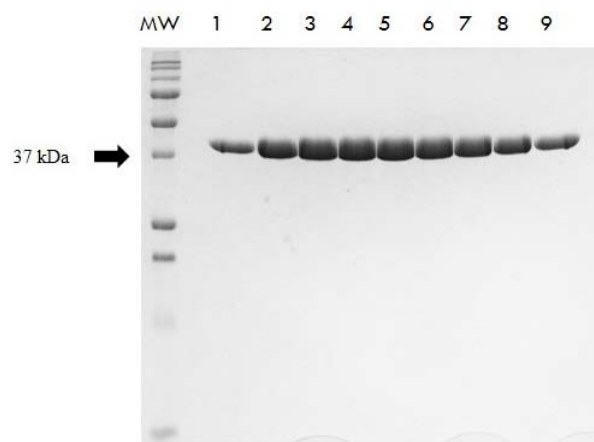
Cromatografia na coluna Butil Sepharose

A coluna Butil Sepharose (GE Healthcare) baseia-se na seleção da proteína alvo via hidrofobicidade, a matriz da coluna possui grupos hidrofóbicos que podem aumentar a interação com o aumento da concentração de sal. Após a proteína de interesse estar ligada a coluna essa pode ser eluída por meio de um gradiente decrescente de sal.

A coluna utilizada possui 18 mL e foi equilibrada com 2 VC de tampão-R contendo 1 mol/L de sulfato de amônio a um fluxo de 1 mL/min. As frações similares provenientes da coluna Source 15-Q foram ajustadas para 1 mol/L de sulfato de amônio e injetadas na coluna. Após a injeção da amostra, a coluna foi lavada com tampão-R contendo 1

mol/L de sulfato de amônio (2 VC), num fluxo de 0,5 mL/min. A proteína HsRecA foi eluída da coluna com 10 VC de um gradiente linear decrescente de sulfato de amônio (1 - 0,05 mol/L) em tampão R. As frações de 2 mL eluídas durante o gradiente linear de sal foram coletadas e analisadas por eletroforese SDS-PAGE 12,5%, figura (4.6), apresenta um gel representativo da cromatografia na coluna Butil Sepharose. Nesta etapa não foi detectada a presença de nenhum contaminante com bandas visíveis no gel e foi verificada uma pureza da proteína de 95%.

Figura 4.6: Análise das frações coletadas durante a cromatografia na coluna Butil. MW - Padrão de massa molecular (Bio-Rad), (1-9) frações eluídas.



Fonte: O autor.

Dosagem da Concentração de Proteína e Estocagem

A dosagem da concentração de proteína HsRecA foi determinada por espectroscopia de absorção no ultravioleta em 280 nm, utilizando o coeficiente de absorção de $1,60 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$, em uma cubeta de quartzo com caminho ótico de 1 cm. A proteína HsRecA foi dialisada contra tampão tampão-R contendo 100 mmol/L de NaCl e 1 mmol/L DTT (2 x 2L por 4 h e uma vez com 2 L durante 8-12 h), após esse processo a dosagem foi realizada. Utilizando o dispositivo centricon (Millipore) e conhecendo a concentração e volume inicial de proteína foi determinado o volume final para que a concentração final

fosse de 4 mg/mL, a concentração final foi confirmada por absorção no 280 nm. Aliquotas de 10-50 μL foram congeladas em nitrogênio líquido e mantidas em freezer a $-80\text{ }^\circ\text{C}$.

Teste de Ausência de Nucleases

A fim de verificar se a proteína purificada estava livre da presença de nucleases foi realizado um ensaio de degradação de três estruturas de DNA (Circular simples fita, Linear duplar fita e Circular dupla fita). No ensaio de 20 μL (volume final), 20 μM de cada DNA foram incubados com 5 $\mu\text{mol/L}$ da proteína HsRecA, 1mmol/L DTT em tampão-RecA (25 Tris-AC (80% cátion, pH 7.8), 10 mmol/L MgOAc, 3 mmol/L glutamato monossódico (Kglu), 5% (w/v) glicerol e 1 mmol/L DTT) por 2 h a $37\text{ }^\circ\text{C}$. A reação foi interrompida com a adição de 3 μL de tapão *Stop-Load* (15% Ficoll, 0,25% azul bromophenol, 0,25% xileno cyanol, 25 mmol/L EDTA e 10% SDS). Toda a reação foi corrida em gel de agarose 8% em tapão TAE e analisada, se nenhuma banda característica de arraste era detectada a proteína purificada era certificada como livre de nucleases.

Ensaio de Atividade ATPásica

A hidrólise de ATP mediada pela proteína HsRecA foi medida em um ensaio de espectrometria como descrito anteriormente por Morrical, Lee e Cox (1986), Brenner et al. (1987), no tampão RecA a $37\text{ }^\circ\text{C}$. Os ensaios foram realizados em um espectrômetro Varian Cary 300 de feixe duplo, equipado com controlador de temperatura e 12 células de leitura com troca de posição automatizada.

Um sistema de regeneração de ATP (10 U/mL de piruvato quinase e 3 mmol/L fosfoenolpiruvato) e outro de oxidação de NADH (2mmol/L NADH e 10 U/mL lactato desidrogenase) foram acrescentados a reação. A conversão de NADH para NAD^+ pode ser monitorada pelo decréscimo na absorbância em 380 nm, embora a máxima absorbância para o NADH ocorra em 340 nm, a absorbância foi medida em 380 nm, preservando o limite linear de detecção de absorbância do espectrômetro para o longo tempo requerido para realizar os ensaios.

A quantidade de ATP hidrolisado é a mesma que aquela de conversão de NADH

para NAD^+ e foi calculada utilizando o coeficiente de extinção do NADH, $1,21 \text{ M}^{-1}\text{cm}^{-1}$. As concentrações de HsRecA, EcRecA, DNA, SSB e ATP estão indicadas nas legendas das figuras do Capítulo de Resultados e Discussões. Foram realizados ensaios em triplicata e os resultados mostrados são a média dos três ensaios. A proteína SSB de *E. coli* foi utilizada para estimular a atividade das proteínas HsRecA e EcRecA quando devido.

Reação de Troca de Três Fitas de DNA

A reação de troca de de fitas foi realizada como descrito anteriormente por Cox e Lehman (1981), Bedale e Cox (1996), no tampão RecA a 37°C . Utilizando um sistema de regeneração de ATP (10 U/mL de piruvato quinase e 2,5 mmol/L fosfoenolpiruvato) também foi incluído. As proteínas HsRecA ou EcRecA foram previamente incubadas com o cssDNA M13mp18 (7249 mer) durante 20 min, a proteína SSB e ATP foram então adicionadas, seguido de outros 10 min de incubação. A reação foi então iniciada pela adição do DNA linear dupla-fita ldsDNA M13mp18 e incubada por 60 min. Para parar a reação $5 \mu\text{L}$ do tampão *Stop-Load* foi adicionado para cada $10 \mu\text{L}$ de alíquota coletada em diferentes intervalos de tempo. As alíquotas foram então analisadas por eletroforese em gel de agarose 0,8% no tampão TAE. As respectivas concentrações utilizadas de HsRecA, EcRecA, DNA, SSB e ATP estão indicadas na figura do Capítulo de Resultados e Discussões.

4.3 Microscopia Eletrônica

Um método modificado de Alcian foi utilizado para visualizar os filamentos formados pelas proteínas HsRecA e EcRecA sobre o cssDNA M13mp18. O material de deposição da amostra "*grids*" foi preparado como descrito previamente por Egger, Lu-setti e Cox (2003). A preparação das amostras para deposição e análise por microscopia seguiu o procedimento a seguir.

Todas as incubações aconteceram a 37°C . As proteínas EcRecA ou HsRecA ($6,7 \mu\text{mol/L}$) foram previamente incubadas com $20 \mu\text{mol/L}$ ssDNA M13mp18 no tampão RecA

na presença do sistema de regeneração de ATP (10 U/mL de creatina fosfoquinase e 12 mmol/L de fosfocreatina - PC) por 20 min. Então, 3 $\mu\text{mol/L}$ de ATP com o sem 2 $\mu\text{mol/L}$ da proteína SSB foi adicionada ao sistema. Finalmente, 3 mmol/L de ATP γ S foi adicionado por outros 3 min de incubação a fim de estabilizar os filamentos.

Para depositar a amostra sobre o *grid* a mistura foi diluída para a concentração de DNA de 0,4 ng/ μL no tampão de diluição (200 mmol/L acetato de amônio, 10 mmol/L HEPES (pH ajustado para 7,5 com HCl) e 10% glicerol) e adsorvida em um *grid* de Alcian por 3 min. O *grid* foi então encostado em uma gota do tampão de diluição, posteriormente se manteve o *grid* flutuando sobre uma nova gota no mesmo tampão durante 1 min, esse procedimento é realizado para retirada de excesso de material Egger, Lusetti e Cox (2003).

Para revelação da amostra, o *grid* foi então encostado-o em uma gota contendo 5% de acetato de uranil, seguido da flutuação do *grid* em uma nova gota durante 30 s. Finalmente o *grid* foi lavado encostando-o em água destilada, seguido da imersão em *beakers* de 10 mL de água destilada, por duas vezes. Posteriormente, a amostra foi seca, esse procedimento foi realizado via técnica *rotary-shadowed*, onde o metal *platinum* é evaporado e depositado sobre a superfície contendo as moléculas que se quer visualizar (SHERRATT et al., 2009).

A captura e registro das imagens foram realizadas com um microscópio de transmissão eletrônica TECNAI G2 12 Twin Electron Microscope (FEI Co.) equipado com a câmera GATAN 890 CCD. Do departamento de Engenharia de Materiais da Universidade de Wisconsin-Madison USA. Imagens digitais dos filamentos de nucleoproteínas foram tomadas com 15000 X de magnificação. Nesta etapa a técnica de laboratório Sindhu Chitteni-Pattu foi responsável pela coleta de imagens no equipamento.

O comprimento de dez filamentos formados pelas proteínas EcRecA e HsRecA foram medidos três vezes, com rotação da imagem de 0, 45 e 90°, utilizando o software Metamorph, o comprimento médio foi calculado em nm. A barra de escala de 0,5 μm foi utilizada como padrão para calcular o número de pixel por μm .

4.4 Determinação da Estrutura Tridimensional da proteína HsRecA

Os procedimentos utilizados para cristalização e determinação da estrutura da proteína estudada estão apresentados por passos:

Passo 1 - Cristalização da proteína HsRecA

Todos os experimentos de cristalização foram realizados utilizando o método da matriz esparsa (JANCARIK; KIM; IUCR, 1991), e técnica da difusão de vapor pela gota suspensa (*hanging-drop*), em placas de 24 poços a 20 °C. Todas as condições foram testadas com a proteína HsRecA sozinha e em complexos com os ligantes ATP γ S e ADP e ssDNA dT18. Na busca por condições iniciais de cristalização foi utilizado um reservatório contendo 1 mL dos componentes dos kits comerciais de cristalização Cryo Suit (QIAGEN), PEG Ion, Natrix, Crystal Screen 1 e 2 e Crystal Screen Lite (Hampton Research). Cada gota continha 1 μ L da solução do reservatório e um volume igual de solução da proteína HsRecA (4 mg/mL) adicionado dos ligantes (ATP γ S e ADP e ssDNA dT18).

Inicialmente, foram observados aglomerados de múltiplas agulhas na condição 0,2 M CaCl₂, 20% w/v PEG 3350, pH 6,8 após 4 dias. Uma primeira otimização foi realizada para esta condição, inicialmente variando as concentrações de CaCl₂ e PEG 3350, no entanto, os cristais mantiveram o formato de agulha, indicando um crescimento preferencial em uma dimensão. Realizando uma segunda otimização com aditivos do kit comercial Additive Screens (Hampton Research), foi observada uma melhora significativa no crescimento dos cristais com a adição de polipropileno glicol 400 (PPG 400). Uma terceira otimização foi realizada pela busca da concentração ideal de CaCl₂, PEG 3350 e PPG 400. Os melhores cristais foram finalmente obtidos na condição 0,25 M CaCl₂ e 14% w/v PEG 3350, com adição de 5-8% w/v de PPG 400 diretamente na gota de cristalização. Os melhores cristais se formaram a partir da solução contendo 0,1 mmol \cdot L⁻¹ HsRecA, 10 mmol \cdot L⁻¹ MgCl₂ e 2 mmol \cdot L⁻¹ ADP, após 4 dias. A solução adiconada dos ligantes foi preparada com 24 h de antecedência e permaneceu a 4 °C até o momento de mistura com a gota.

Passo 2 - Coleta de dados de raios X

A coleta de dados de difração de raios X foi realizada utilizando um difratômetro de geometria κ , no Laboratório de Cristalografia de Proteínas do Professor Dr. James Kerck na Universidade de Wisconsin-Madison. Os dados de difração foram coletados utilizando a radiação CuK_{α} em um difratômetro da Bruker, equipado com anôdo rotatório e um gerador de raios X operando na voltagem de 40 kV e corrente de 40 mA, acoplado a um detector CCD Smart-6000.

Os cristais analisados foram rapidamente resfriados em nitrogênio líquido (100 K) utilizando a solução do reservatório como crioprotectante, adicionado de 25% etileno glicol. O melhor cristal difratou para um máximo de resolução atômica de 1,7 Å. Um total de 1056 imagens de dados foram coletadas com um ângulo de oscilação de 0,5°, com tempo de exposição de 120 s por imagem coletada e uma distância cristal-detector de 60 mm.

Passo 3 - Processamento dos dados de raios X

Nesta etapa um arquivo (.MTZ) obtido experimentalmente a partir da interface do programa Proteum2 (Bruker AXS (2010) PROTEUM2, Version 2010.5, Bruker AXS Inc., Madison, Wisconsin, USA), contém a informação geral de cada uma das imagens de difração coletada. Desta forma é necessário colocar todas as imagens medidas em uma única escala, bem como reduzir os dados medidos para que contenham apenas um conjunto único de reflexões, excluindo reflexões geradas por simetrias, esse procedimento é chamado de redução e escalonamento dos dados. O programa Proteum2 realiza essa tarefa, e fornece o arquivo de saída (.MTZ) com as intensidades de cada conjunto hkl mensurado.

Passo 4 - Substituição molecular

O faseamento inicial foi obtido pelo método de substituição molecular, que permite atribuirmos fases preliminares para um conjunto de amplitudes experimentais uti-

lizando para isto a estrutura tridimensional de uma estrutura já determinada (modelo inicial). A rotação e translação do modelo inicial foram realizadas no programa Phaser (MCCOY et al., 2007). O modelo cristalográfico da proteína RecA de *E. coli* (PDB:1xmv) foi utilizado como modelo inicial de busca. Em um primeiro faseamento, a utilização do modelo mostrou-se insatisfatória, a densidade eletrônica da região C-terminal não apresentava qualidade suficiente para construção do modelo para a proteína HsRecA, então optou-se em separar o modelo de busca em duas partes: Regiões N-terminal e central (resíduos 3 a 282) e a região C-terminal (resíduos 283 a 328). Nesta etapa é gerado um arquivo .pdb que contém as informações estruturais de um primeiro modelo da estrutura que deseja-se determinar.

Passo 5 - Refinamento

A estrutura inicial foi construída com o programa phenix.autobuild (TERWILLIGER et al., 2008), refinada e modelada com ciclos manuais de refinamento com o programa COOT (EMSLEY et al., 2010) e computacionais no phenix.refine (ADAMS et al., 2010). Nesta etapa foi utilizado uma rotina computacional escrita pelo Prof. Dr. Jorge Iulek da Universidade Estadual de Ponta Grossa - Dep. de Química, para refinamento utilizando o programa phenix.refine.

Passo 6 - Validação

Esta etapa é crucial na determinação de estrutura, nela deve-se avaliar a qualidade do modelo calculado, e também da presença de possíveis erros estruturais em regiões localizadas, bem como erros em parâmetros estereoquímicos. A fim de verificar se o modelo construído era consistente e não possuía erros de modelagem o programa Molprobitry (CHEN et al., 2010) foi utilizado para validação geométrica. Nesta etapa o comprimento de ângulo de ligações, a planaridade de anéis de cadeias laterais, a quilaridade, as conformações de cadeias laterais, o impedimento estérico entre pares de átomos não ligados são avaliados. E o gráfico de Ramachandran gerado no Molprobitry é utilizado para avaliar a disposição dos ângulos torcionais ϕ/ψ . As ligações ϕ (entre o grupo NH e carbono alfa)

e ψ (entre o carbono alfa e o grupo carbonila) não possui nenhuma restrição e possui rotação livre. Contudo, nem todas as combinações $\phi + \psi$ são permitidas estericamente, algumas combinações causam colisões desfavoráveis entre grupos da cadeia principal e/ou as cadeias laterais.

Se o modelo gerado não possui nenhum erro de modelagem então a estrutura final pode ser depositada no banco de dados pdb, no qual também ocorre uma validação comparativa com as estruturas já depositadas no banco (<http://www.rcsb.org/pdb/home/home.do>). A estrutura da HsRecA foi depositada com o código de acesso pdb:5JRJ.

A fim de verificar se o modelo construído era consistente e não possuía erros de modelagem o programa Molprobitry foi utilizado para validação geométrica (CHEN et al., 2010).

A estrutura final foi depositada no banco de dados pdb com o código de acensão pdb:5JRJ.

Ferramentas de bioinformática

Os modelos de estrutura tridimensional foram preparados utilizando o programa STRID para assinatura da estrutura secundária (FRISHMAN; ARGOS, 1995), e visualizada na interface do programa PyMOL (SCHRÖDINGER, 2010).

O alinhamento da sequência de aminoácidos levando a estrutura secundária das proteínas foi realizado utilizando os programas MultiProt e T-Coffee (SHATSKY; NUSINOV; WOLFSON, 2004; NOTREDAME; HIGGINS; HERINGA, 2000), e os resultados foram visualizados na interface *on-line* ESPript - <http://esprpt.ibcp.fr> (ROBERT; GOUET, 2014).

As superfícies de potencial eletrostático foram calculadas no servidor Adaptive Poisson-Boltzmann Solver (APBS) <http://www.poissonboltzmann.org/> (BAKER et al., 2001; DOLINSKY et al., 2007), por meio da solução da equação de Poisson-Boltzmann, onde esse descreve interações eletrostáticas entre solutos em meio aquoso. Nesta etapa o arquivo .pdb do modelo é depositado no servidor APBS, após o cálculo da superfície de

potencial o servidor retorna um arquivo .PQR que contém as informações para gerar a superfície, essa visualizada no programa PyMOL (SCHRÖDINGER, 2010).

Capítulo 5

Resultados e Discussões

Este capítulo apresentará a caracterização estrutural e funcional da proteína HsRecA, sendo dividido em duas partes. Primeiramente, serão apresentados os resultados da cristalização e determinação da estrutura cristalográfica, bem como comparações estruturais. Posteriormente, serão apresentados os resultados de atividade de reação de troca de fitas, atividade ATPásica e microscopia eletrônica.

5.1 Estrutura Tridimensional da Proteína HsRecA

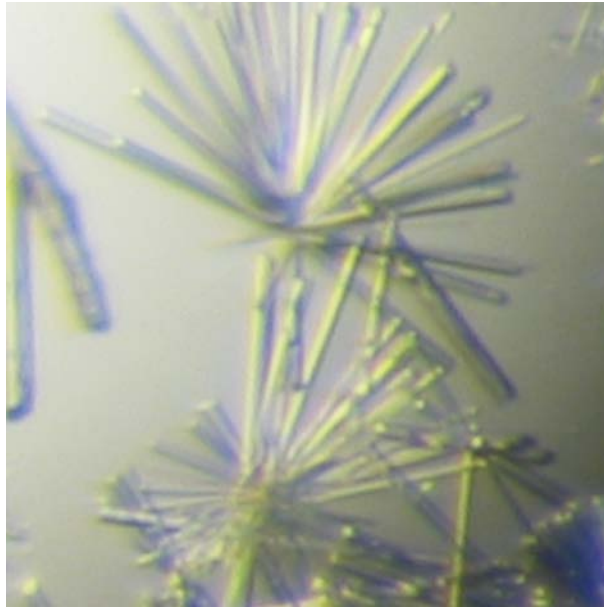
Cristalização, Coleta de Dados de Difração de Raios X

Em busca de condições de cristalização 6 kits comerciais foram utilizados (aproximadamente 576 condições). Foram observados cristais em diversas condições, no entanto após 4 dias a condição contendo 0,25 mol/L $CaCl_2$ e 14% (w/v) polietileno glicol 3500 (PEG 3500) do kit PEG ION (Hampton Research), apresentou os primeiros cristais com dimensões adequadas, a morfologia desses cristais estão apresentados na figura (5.1a), suas dimensões não foram determinadas. Esses cristais não permitiram uma coleta completa de dados de difração de raio X, aliada a baixa resolução do padrão de difração. Após triagem e refinamento com o kit Additive Screen (Hampton Research) constatou-se a adição direta de polipropileno glicol 400 (PPG 400) na concentração final de 5-8% (w/v), o crescimento dos cristais foi otimizado e esses apresentaram melhoria significativa nas

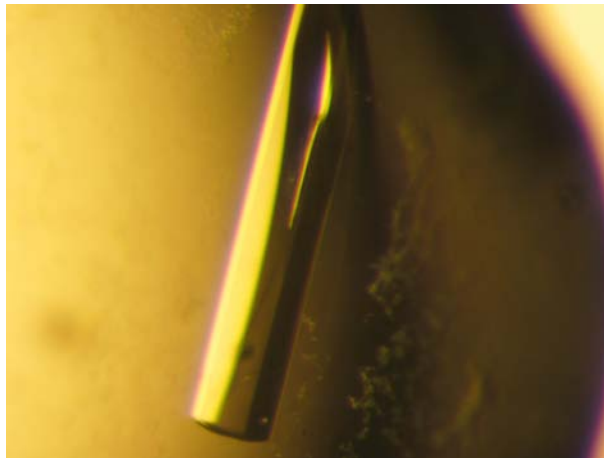
dimensões e morfologia. Os melhores cristais se assemelham ao mostrado na figura (5.1b), onde esse possui dimensões 0,8 mm x 0,2 mm x 0,2 mm.

Figura 5.1: Cristal da proteína HsRecA obtidos pelo método da difusão de vapor em gota suspensa. Condição de cristalização: (a) 0,25 M $CaCl_2$ e 14% (w/v) polietileno glicol 3500 (PEG 3500), (b) 0,25 M $CaCl_2$ e 14% (w/v) polietileno glicol 3500 (PEG 3500) e 5% (w/v) de polipropileno glicol 400 (PPG 400).

(a)



(b)

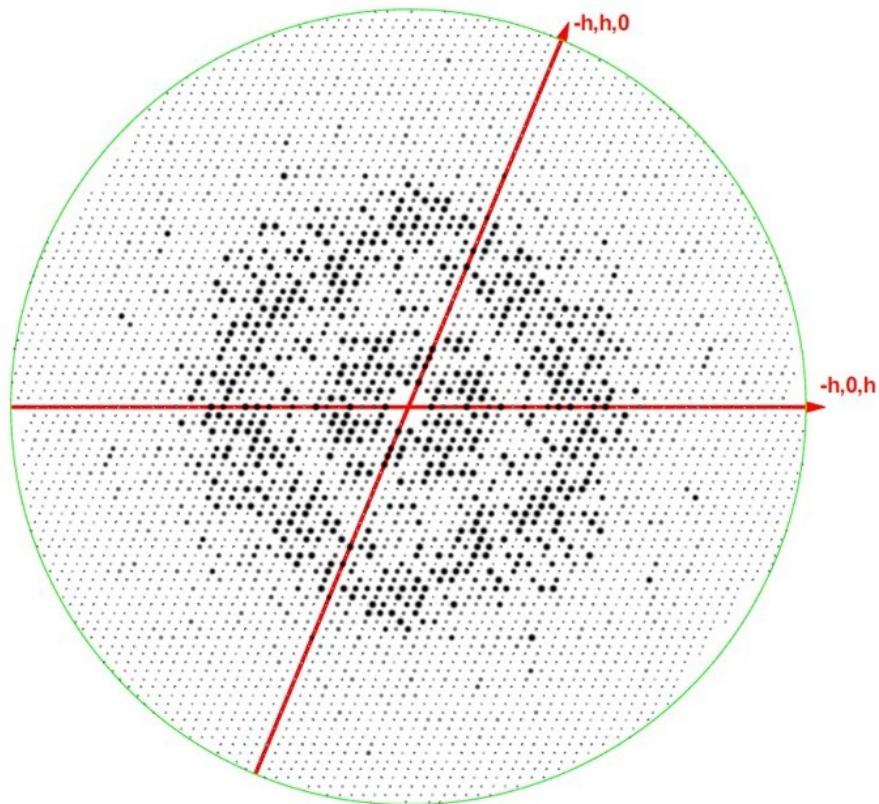


Fonte: O autor.

O cristal da figura (5.1b) difratou para o máximo de resolução atômica de 1,70 Å, uma imagem representativa do seu padrão de difração é apresentado na figura . O

crystal analisado pertence ao grupo espacial $P6_1$, os valores dos parâmetros de rede são $a = b = 87,41\text{\AA}$ e $c = 91,35\text{\AA}$. Uma estatística detalhada da coleta de dados de raios X desse cristal é apresentada na tabela (5.1)

Figura 5.2: Padrão de difração do cristal da HsRecA. A linha verde indica a região de mais alta resolução, que foi de $1,7\text{\AA}$.



Fonte: O autor.

Tabela 5.1: Estatísticas de processamento e coleta de dados de raios X para o cristal da proteína HsRecA. Os valores entre parênteses se referem a faixa de maior resolução de 1,73 - 1,70 Å.

Número de imagens	1056
Comprimento de onda (Å)	1,542
Faixa de resolução (Å)	75,71 - 1,70 (1,73 - 1,70)
Grupo espacial	$P6_1$
Cela unitária	
a; b; c (Å)	87,41; 87,41; 91,35
α ; β ; γ (°)	90; 90; 120
Total de reflexões	
Reflexões únicas	42574 (3467)
Multiplicidade	5,18 (1,83)
Completeza (%)	97,67 (79,50)
Média de $I/\sigma(I)$	31,03 (2,08)
Wilson B-factor	16,94
R_{merge}	0,0467 (0,6285)

Fonte: O autor.

Na tabela (5.1), o alto valor da média de $I/\sigma(I)$ e o baixo valor de R_{merge} indica que os parâmetros cristalinos foram corretamente determinados, o consenso é que $I/\sigma(I) \leq 2$ e $R_{merge} \geq 0,5$ para considerar que uma determinada faixa de resolução contenha somente ruído e deva ser descartada. O baixo valor do Wilson B-factor obtido está correlacionado diretamente a média $I/\sigma(I)$, pois é um indicativo que o movimento atômico foi reduzido devido a baixa temperatura na qual o cristal foi coletado (-173,15 °C). Adicionalmente, a completeza próximo de 100% mostra que as imagens coletadas possuem informação suficiente para resolver a estrutura, uma excelente coleta de dados atinge 99% ou 100% de completeza.

A proteína HsRecA nativa possui 351 aminoácidos, uma massa molecular teórica de 37,47 kDa. O coeficiente de Matthews indicou a presença de um único monômero de

HsRecA na unidade assimétrica ($V_m = 2,73 \text{ \AA}^3 \text{ Da}^{-1}$; conteúdo de solvente = 55,01 %) (MATTHEWS, 1968).

Faseamento Inicial e Refinamento

A estrutura tridimensional da proteína HsRecA foi resolvida pelo método de substituição molecular, utilizando o programa **PHASER** (MCCOY et al., 2007). As coordenadas atômicas da proteína EcRecA (pdb:1xmv, (XING; BELL, 2004)) foram utilizadas como estrutura de referência. O alinhamento da sequência (figura (5.5), na página 91) de aminoácidos da HsRecA e EcRecA, mostrou que essas proteínas dividem 68,12% de identidade e 80,6% de similaridade. O alto valor de identidade que essas proteínas dividem justifica o uso da proteína EcRecA como estrutura de referência.

Em uma primeira tentativa de substituição molecular a região C-terminal da proteína EcRecA mostrou-se estar em diferente orientação em relação a proteína HsRecA. Para contornar o problema optou-se em dividir as coordenadas atômicas do arquivo pdb:1xmv em dois modelos: o Modelo 1 possui os resíduos 3-282, compreendendo a região N-terminal até domínio central (core ATPásico), e o Modelo 2 os resíduos 283-328 (região C-terminal). Esse procedimento conferiu maior liberdade translacional e rotacional ao Modelo 2 e mostrou-se adequado para a solução via substituição molecular. A melhor solução para cada modelo está apresentada na tabela (5.2), juntamente com os respectivos *scores* (RFZ e TFZ) de cada função. O manual do programa PHASER indica que uma possível solução é encontrada para aqueles valores de RFZ menores que 4, e para os valores de TFZ maiores que 6.

A partir da solução composta (Modelo 1 + Modelo 2) o mapa de densidade eletrônica inicial gerado pela substituição molecular, contendo as fases iniciais aproximadas, apresentou qualidade suficiente para construir aproximadamente 65% da cadeia polipeptídica da HsRecA presente na unidade assimétrica. Neste mapa foi aplicada uma construção automatizada para a cadeia polipeptídica, utilizando o programa **phenix.autobuild**, por meio de refinamento de corpo rígido. O modelo inicial gerado a partir do mapa apresentava valores de $R_{work} = 35,46\%$ e $R_{free} = 56,73\%$, após o refinamento automatizado esses

Tabela 5.2: Valores de rotação, translação e os scores para os modelos utilizados pela substituição molecular na resolução da estrutura da proteína HsRecA. A estrutura da proteína EsRecA (pdb 1xmv), separada em dois modelos, foi utilizado como estrutura de referência para os cálculos das fases iniciais. Na tabela α° , β° e γ° , são os ângulos euclidianos da operação de rotação. T_x , T_y e T_z correspondem as respectivas operações de translações nos eixos x; y; z. RFZ e TFZ representam os scores da função de rotação e translação, respectivamente.

Modelo Estrutural	α°	β°	γ°	$T_x(\text{Å})$	$T_y(\text{Å})$	$T_z(\text{Å})$	RFZ	TFZ
Modelo 1	55,9	177,1	216,8	0	0	0,02	3,6	6,7
Modelo 2	347,7	119,2	227,8	0,77	0,77	-0,35	2,5	7,0

Fonte: O autor.

convergiram para $R_{work} = 22,11\%$ e $R_{free} = 25,11\%$, contendo inicialmente 229 resíduos no modelo, de um total de 351 constituintes da proteína HsRecA.

O modelo gerado por **autobuild** foi transferido para a interface do programa COOT, e refinado manualmente utilizando o programa **Phenix.refine**. Foram impostas restrições geométricas para eliminar qualquer tendensidade do modelo. Os mapas de densidade eletrônica foram contornados em $mF_o - DF_c = 3$ e $2mF_o - DF_c = 1$ e utilizados durante o refinamento manual. Aqueles resíduos que não estavam resolvidos na estrutura da proteína EcRecA e que não puderam ser modelados via **autobuild** foram incluídos manualmente na cadeia polipeptídica. Uma estatística detalhada do refinamento e da qualidade do modelo final é apresentada na tabela (5.3).

Tabela 5.3: Estatísticas de processamento da coleta de dados da proteína HsRecA. Os valores entre parênteses se referem a faixa de resolução de 1,73 - 1,70 Å.

Reflexões usados para R_{free}	2025(109)
R_{work}	17,55 % (0.3151)
R_{free}	21,07 % (0.3775)
Número de átomos	
Átomos exceto hidrogênios	2743
Proteína	2359
Ligantes	59
Águas	325
Resíduos	308
Desvios r.m.s	
Comprimento de ligação (Å)	0,015
Ângulos de ligação (°)	1.62
B -factor médio (Å ²)	23,30
B-Proteína (Å ²)	22,30
B-Ligantes (Å ²)	21,70
B-Solvente (Å ²)	31,00

Fonte: O autor.

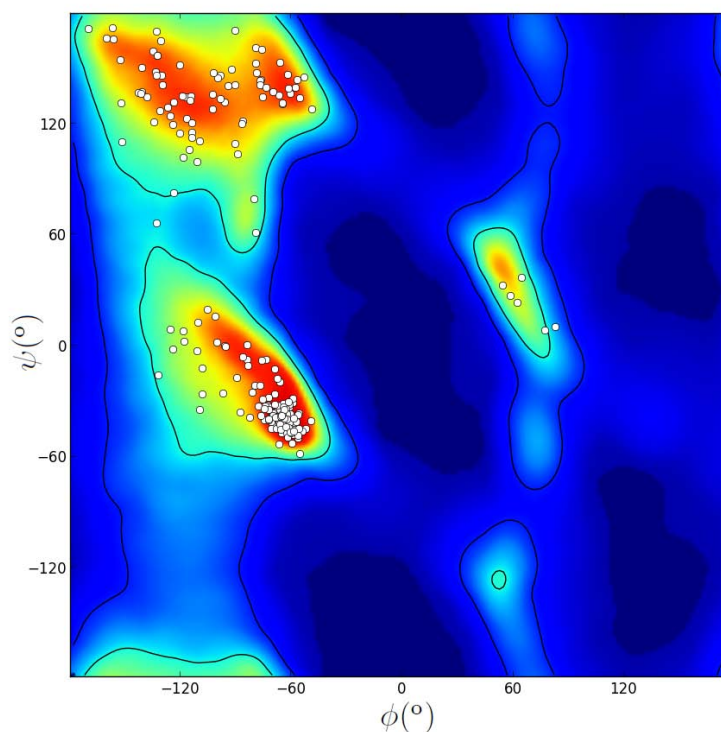
A partir do refinamento no COOT foi possível construir um modelo com 308 resíduos, aproximadamente 87% da cadeia polipeptídica da HsRecA. O mapa de densidade eletrônica não foi observado para os resíduos 1-5, 41-43, 164-175, 201-218 e 343-351, essas regiões não modeladas são indicativos de alta flexibilidade conformacional.

Inicialmente os valores dos índices residuais R_{work} e R_{free} eram de 35,46% e 56,73%, respectivamente. A grande discrepância entre esses valores é justificada pela não inconsistência do modelo sobre os dados experimentais. Após o refinamento via **autobuild**

e o manual esses índices residuais convergiram para $R_{work} = 17,55\%$ e $R_{free} = 21,07\%$. Estes índices devem convergir para valores próximos ao final do refinamento, implicando que o modelo estrutural gerado é consistente com os dados experimentais, e a estrutura foi resolvida com sucesso.

Adicionalmente, aos valores de R_{work} e R_{free} é possível realizar uma validação geométrica do modelo gerado, observando a disposição espacial dos ângulos ϕ/ψ . Observa-se pela figura (5.3) que nenhum resíduo está presente em regiões proibidas no gráfico de Ramachandran, com 97,58% dos resíduos presentes na região mais favorecida (RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963). Indicando que os ângulos torcionais $\phi(^{\circ})$ e $\psi(^{\circ})$ estão corretamente ajustados no modelo para todos os resíduos refinados. Também pode-se observar pela tabela (5.4) que os critérios que determinam a qualidade final do modelo estrutural (de acordo com valores ideais) foram todos alcançados.

Figura 5.3: Gráfico de Ramachandran para o modelo estrutural refinado da proteína HsRecA. Em vermelho são as regiões permitidas, em amarelo adicionalmente permitidas, azul claro generosamente permitas e azul escuro as proibidas. Os pontos brancos sobre o mapa referem-se a combinação de ângulos torcionais $\phi(^{\circ})$ e $\psi(^{\circ})$ dos 308 resíduos do modelo. Gráfico gerado pela validação no programa Molprobitry.



Fonte: O autor.

Tabela 5.4: Estatísticas de validação geométrica do modelo estrutural da proteína HsRecA via Molprobitry.

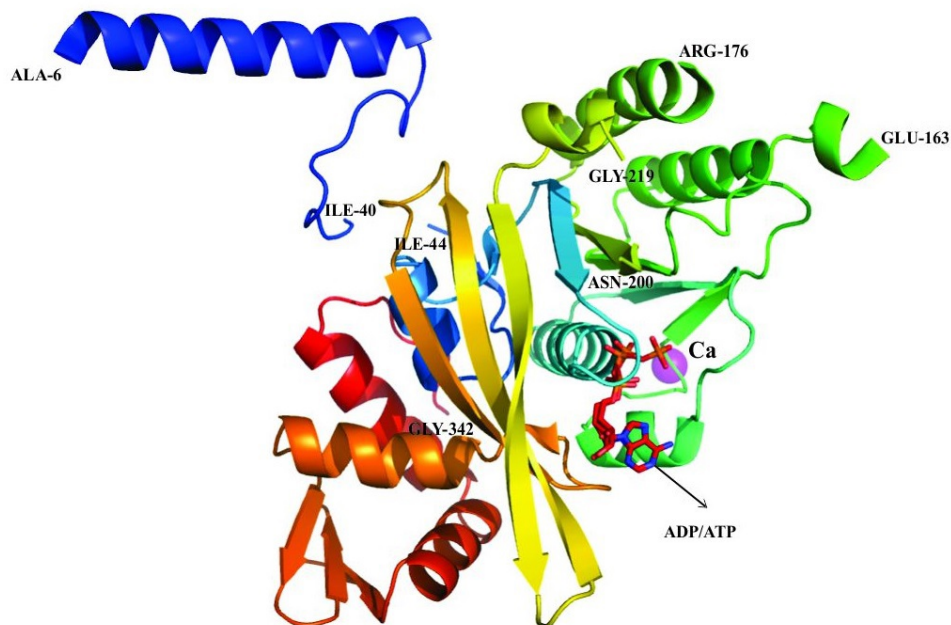
		Valor ideal
Ramachandran favorecido (%)	97,58	> 98
Ramachandran proibido (%)	0	< 0,2
Rotâmeros pobres (%)	0,4	< 1,0
Desvios de $C\beta$	0	0

Fonte: O autor.

5.2 Descrição Geral da Estrutura da Proteína HsRecA

Como predito anteriormente por Story, Weber e Steitz (1992) estrutura terciária da proteína HsRecA pode ser subdividida em três domínios: Um pequeno domínio N-terminal (NTD), um domínio central contendo um sítio ATPásico e e um grande domínio C-terminal (CTD). A figura (5.4) mostra o diagrama de fitas para o monômero do cristal da proteína HsRecA.

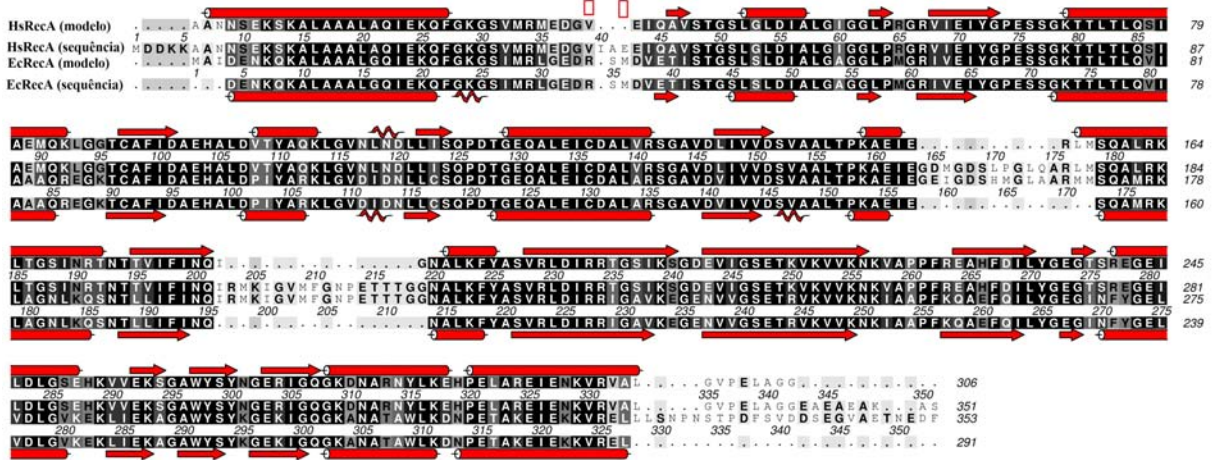
Figura 5.4: Diagrama de fitas da estrutura cristalográfica do monômero da proteína HsRecA. Regiões que não puderam ser modeladas estão indicadas pelo último resíduo-número. O monômero é composto por um domínio N-terminal (NTD), um domínio central contendo um sítio ATPásico e e um grande domínio C-terminal (CTD). No domínio central está presente um íon de Ca^{2+} (esfera magenta). O sítio ATPásico está parcialmente ocupado pelos nucleotídeos ATP e ADP.



Fonte: O autor.

A comparação estrutural do monômero da proteína HsRecA e da EcRecA (pdb:, XMV) mostrou que elas diferem com um RMSD (*root-mean square deviation*) de 1,324 Å para os C_{α} , 1,335 Å para a cadeia principal e 1,486 Å para todos os átomos. A figura (5.5) apresenta um alinhamento com a sequência de aminoácidos da estrutura cristalográfica e da sequências das proteínas HsRecA (pdb:5JRJ) e EcRecA (pdb:1xmv).

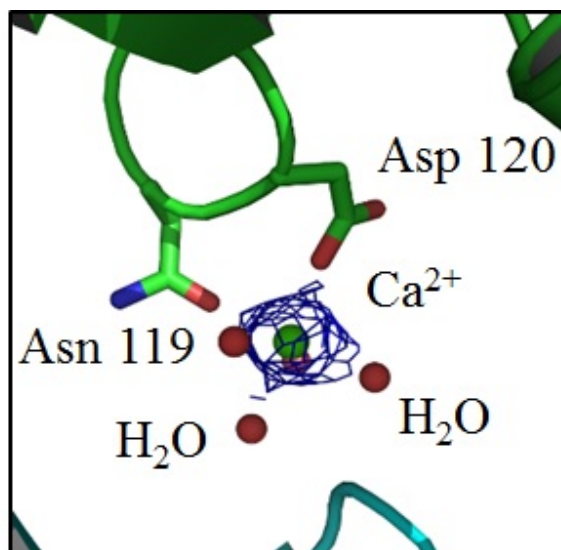
Figura 5.5: Alinhamento da sequência de aminoácidos presente na estrutura cristalográfica contra a sequência de aminoácidos das proteínas HsRecA e EcRecA. A assinatura da estrutura secundária está indicada acima e abaixo de cada sequência.



Fonte: O autor.

A estrutura tridimensional da HsRecA é similar a de suas homólogas, possuindo os mesmos elementos de estrutura secundária, como observado nas RecAs de *E. coli*, *D. radiodurans*, *M. smegmatis*, *M. tuberculosis* (STORY; WEBER; STEITZ, 1992; STORY; STEITZ, 1992; DATTA et al., 2003; CHANDRAN et al., 2015; RAJAN; BELL, 2004). O NTD (6 - 40 aa) consiste de uma α -hélice e um motivo *random coil*. O CTD (276 - 343 aa) consiste de 3 α -hélices e um motivo β -strand antiparalelo. O domínio central (44 - 275 aa) consiste de um conservado sítio ATPásico, intercalando α -hélices e β -strand, que nele contêm os sub-motivos Walker A (72 - 81 aa) e B (147 - 157 aa), e também os sítios de ligação ao DNA L_1 (163-176 aa) e L_2 (202-219 aa), os sítios de ligação ao DNA não estavam presentes na densidade eletrônica devido ao alto grau de desordem. Na estrutura também está ligado um íon de Ca^{2+} , coordenado pelos resíduos Asparagina 119 (Ans119) e Aspartato 120 (Asp120). Esses resíduos fazem parte do domínio central, e a presença do íon foi confirmada pela sua coordenação geométrica, verificada no servidor CheckMyMetal (CMM), adicionado da alta densidade eletrônica no sítio de ligação. A figura (5.6) mostra o sítio de interação do íon Ca^{2+} . A presença desse íon ligado ao sítio ATPásico ainda não havia sido relatado para nenhuma estrutura de proteína RecA.

Figura 5.6: *Stereoview* do sítio de ligação do íon Ca^{2+} e dos resíduos que coordenam este íon. Os valores mostrados em verde se referem as distâncias interatômicas em Å do íon. As águas 141, 222 e 304 fazem a coordenação íon juntamente com os resíduos Asparagina 119 (Asn119) e Aspartato 120 (Asp120).



Fonte: O autor.

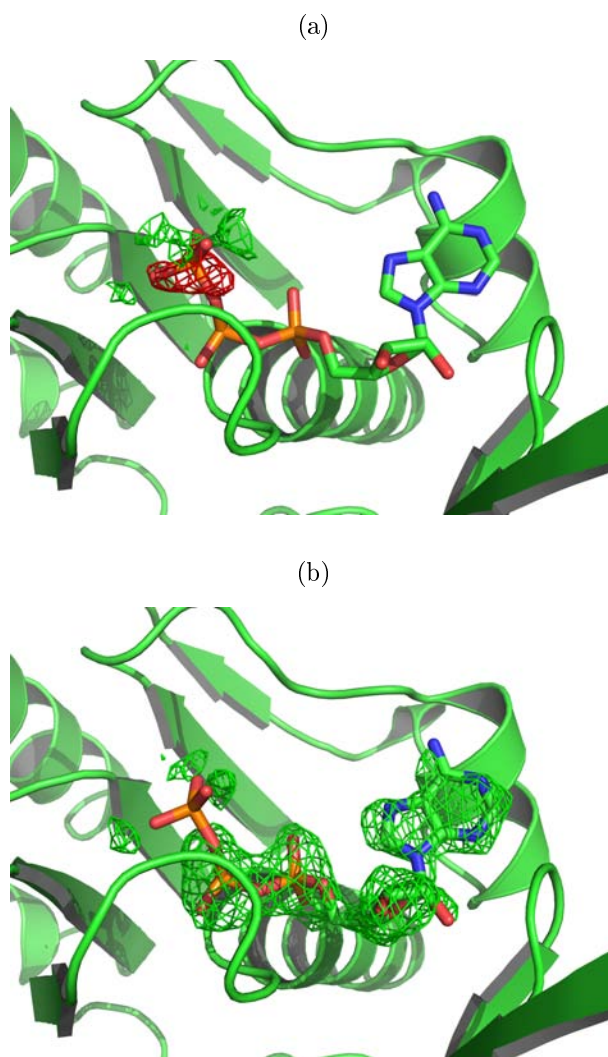
5.3 O Sítio ATPásico

A solução de proteína contendo 0,1 mmmolLHsRecA - 10 mmolLMgCl₂ - 2 mmolLADP, possibilitou que o nucleotídeo ADP fosse co-cristalizado na estrutura. Por meio da análise das vizinhanças de ligação do ADP no sítio foi possível verificar que esse interage diretamente com os seguintes resíduos: Ser76, Ser77, Thr80, Thr81, Asp107, Tyr110 e Gly272 na estrutura da proteína HsRecA. O íon Mg²⁺ é um cofator importante para a atividade ATPásica da proteína RecA, porém apesar da solução de proteína possua 10 mmmolLMgCl₂, não foi observada densidade eletrônica para o íon Mg²⁺ ao redor do ADP de forma a solucionar apropriadamente sua coordenação. A ausência do íon Mg²⁺ também foi observado nas RecAs de *Deinococcus radiodurans* e *Mycobacterium tuberculosis* co-cristalizadas com nucleotídeos trifostafos (CHANDRAN et al., 2015; RAJAN; BELL, 2004).

Durante o refinamento foi observado uma densidade eletrônica extra com conformação aproximadamente tetraédrica acima do fosfato-β do ADP, a qual não pode ser

apropriadamente modelada com águas, ou íons SO_4^{3-} ou PO_4^{3-} . Entretanto o refinamento com o íon PO_4^{3-} na posição da densidade eletrônica extra, juntamente com o ADP mostrou que eles poderiam estar covalentemente ligados, pois ainda observou-se a presença de uma densidade eletrônica negativa no mapa de Fourier diferença entre o ADP e PO_4^{3-} , considerando que o íon PO_4^{3-} possui 100 % de ocupação, para melhor entendimento e visualização do acima explicado, as figuras (5.5a) e (5.5b), mostram o sitio ATPásico com 100% de ocupação pelo ATP e com 39% de ocupação, respectivamente. Mediante a última observação foi considerado que o sítio ATPásico poderia estar parcialmente ocupado também pelo ATP. O refinamento com o ADP e ATP dividindo a mesma posição melhorou o ajuste da densidade eletrônica e resultou nas ocupações de 61% and 39%, respectivamente.

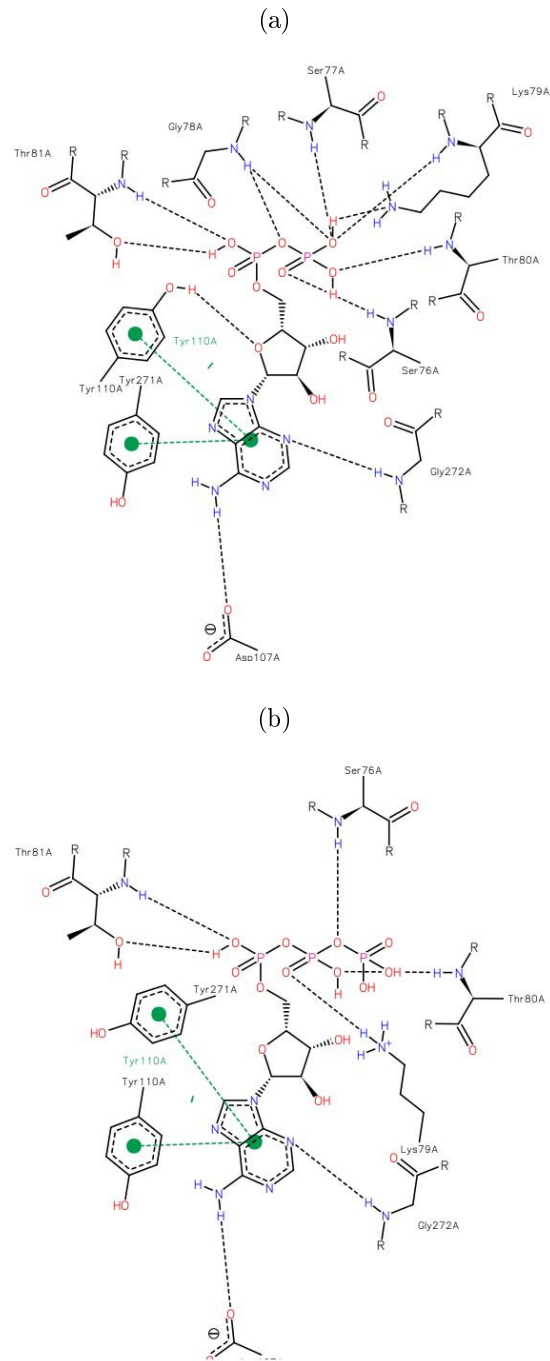
Figura 5.7: Mapas de densidade eletrônica (mFo-DFc) do sítio ATPásico: (a) Refinamento com ATP possuindo 100% de ocupação; a densidade eletrônica negativa sobre o fosfato- γ indica que esta posição não pode estar completamente ocupada. (b) Refinamento com o ATP possuindo 39% de ocupação; a densidade eletrônica positiva sobre o correspondente ADP indica que este deve ocupar a posição sobre a densidade mostrada. Os mapas estão contornados em $+3\sigma$ (verde) e -3σ (vermelho).



Fonte: O autor.

Após verificar a presença dos dois nucleotídeos dividindo a mesma posição, as interações de cada átomo do nucleotídeo com cada resíduo da HsRecA foram identificadas. Essas estão apresentadas nas figuras (5.8a) e (5.8b), para os átomos do ADP e ATP, respectivamente.

Figura 5.8: Interações dos nucleotídeos com aminoácidos do sítio ATPásico da HsRecA. Em (a) interações encontradas para o ADP, em (b) interações encontradas para o ATP.



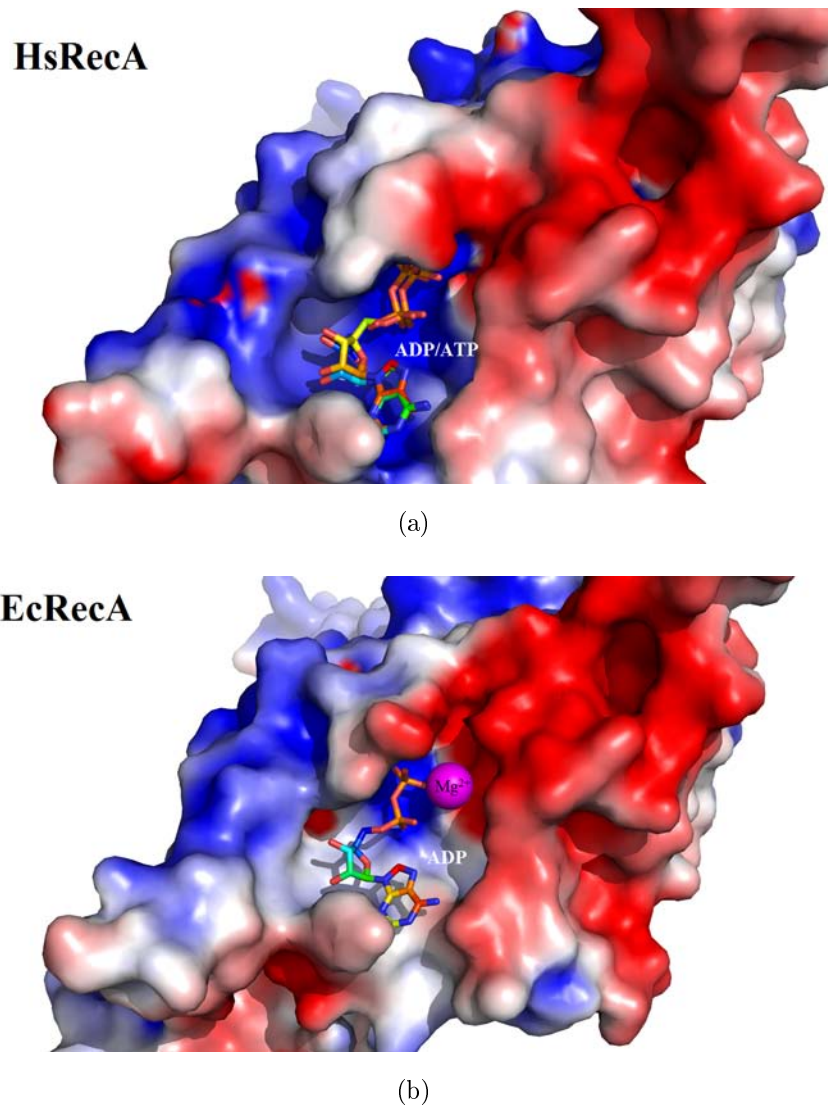
Fonte: O autor.

A glutamina (Gln) é um resíduo altamente conservado no sítio ATPásico, correspondendo as posições 194 e 201 na EcRecA e HsRecA, respectivamente. Esse resíduo não

pode ser mutado para qualquer outro pois inibe completamente a atividade ATPásica. A proposta é que a interação Gln-ATP provoca mudanças conformacionais na RecA (conformação estendida ou comprimida) mediada pela hidrólise de ATP. O íon Mg^{2+} interage com o fosfato β e γ do ATP, de tal modo que o fosfato γ é mantido em posição para interagir com a Gln, desse modo a Gln transmite a informação da ligação do ATP para o *Loop* L_2 e a hélix G, dois elementos estruturais que participam da ligação ssDNA-RecA. Nesta condição a proteína RecA se encontra na forma ativa e estendida, por seguinte o resíduo Glu 96 na EcRecA provome a ativação da molécula de água sobre o fosfato γ , ocasionando a hidrólise de ATP. Uma vez que o ATP está hidrolisado, a Gln não pode manter a conformação do *Loop* L_2 e da hélix G, e o filamento colapsa para a conformação comprimida (KELLEY; KNIGHT, 1997) (Kelley, and Knight 1997, and Hortnagel et al., 1999). Na estrutura da proteína HsRecA o N amida da Gln201 está a 2,88 Å do fosfato- γ do ATP. Na estrutura do complexo DrRecA-ATP γ S, a Gln está a 3,0 Å do thiophosphato- γ . Ambas glutaminas apresentam a mesma conformação nas proteínas HsRecA e DrRecA, diferente daquela observada na estrutura da proteína EcRecA (pdb:1xmv) no qual somente o ADP está ligado na estrutura. Corroborando com a presença do ATP na estrutura da proteína HsRecA.

A reconstrução da superfície eletrostática do monômero da proteína HsRecA, mostrou de forma qualitativa que a cavidade onde ocorre a ligação do ATP é predominantemente positivamente carregada, figura (5.6a), quando comparada à proteína EcRecA, figura (5.6b). O grupo fosfato do ATP possui carga elétrica negativa, uma cavidade possuindo maior densidade de carga positiva aumentaria a interação RecA-ATP, essa interação é importante pois a ligação RecA-DNA é dependente da presença do ATP.

Figura 5.9: Distribuição de potencial eletrostático de superfície na região do sítio ATPásico. Em (a) é mostrado o sítio ATPásico da proteína EcRecA, em (b) o sítio ATPásico da proteína HsRecA. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).



Fonte: O autor.

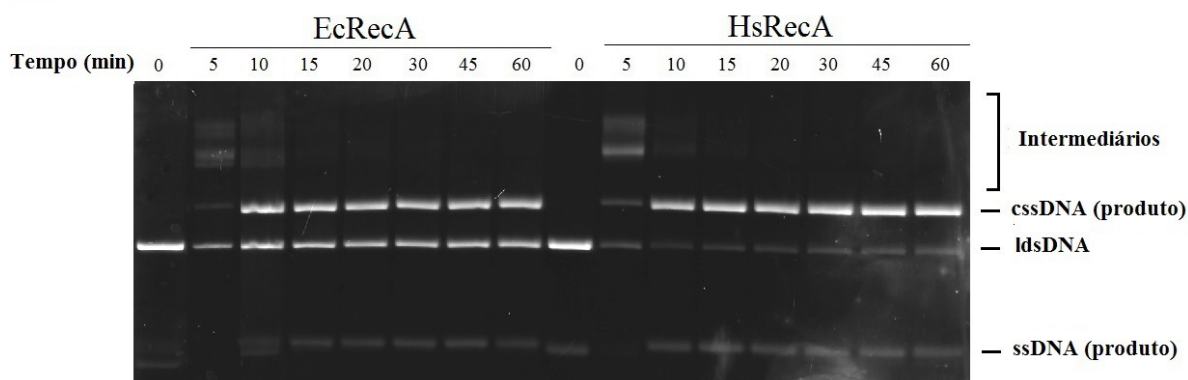
5.4 Reação de Troca de Fitas de DNA *in vitro*

A habilidade das proteínas EcRecA e HsRecA em catalizar a reação de troca de fitas entre o cssDNA e o duplex linear dsDNA (ldsDNA) para formar o duplex circular cdsDNA (Produto) foi examinada entre 5 e 60 min de incubação RecA-DNA na presença

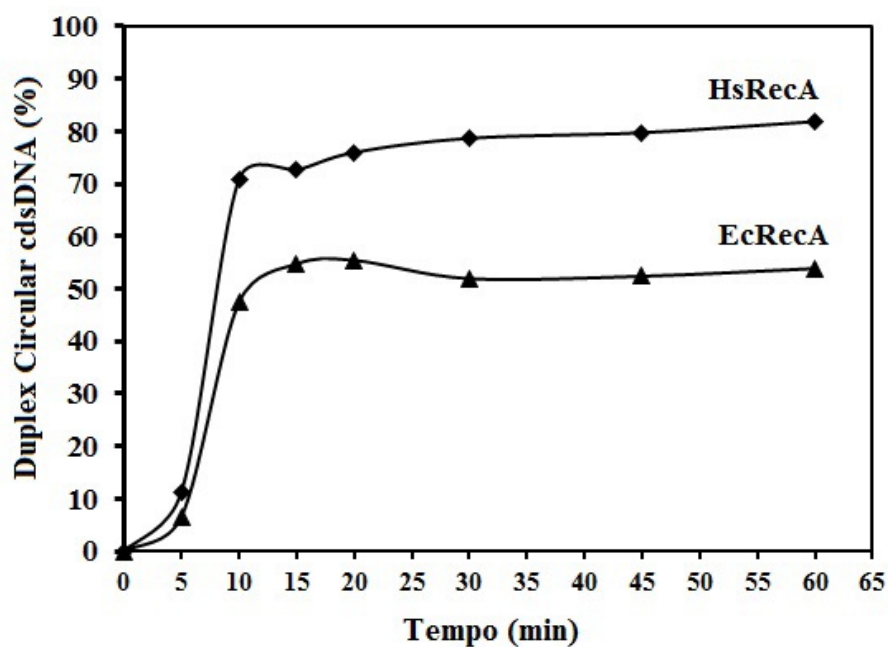
de um sistema de regeneração de ATP e da proteína SSB. A reação de trocas de fitas promovida pela proteína HsRecA foi mais eficiente do que a mesma reação promovida pela proteína EcRecA. As velocidades de reação parecem ser semelhantes para ambas as proteínas, com cdsDNA produtos já detectáveis após 5 min de incubação, bem como moléculas intermediárias. Esse resultado está apresentado na figura (5.11a). Com o tempo adicional as moléculas intermediárias continuaram a ser convertidas em duplex circular cdsDNA, atingindo uma aparente estabilização após 10 min. No entanto, a extensão da reação promovida pela HsRecA sempre foi maior em todos os tempos analisados, conforme quantificação monitorada por meio das intensidades das bandas do cdsDNA formado, figura (5.11b).

No contexto deste trabalho este experimento foi realizado a fim de comprovar que a proteína HsRecA manteve sua central atividade biológica inalterada, mesmo após todo o processo de lise celular e purificação. É essencial que a proteína possua atividade ATPásica e de reação de troca de fitas.

Figura 5.10: Reação de troca de fitas de DNA promovida pelas proteínas HsRecA e EcRecA. (A) Eletroforese em gel de agarose 0,8% do decurso temporal da reação de troca de três fitas, realizada como descrito no capítulo de Material e Método e abaixo, a ilustração dos substratos e produtos da reação de troca de fitas. A reação contendo 10 μM nt cssDNA M13mp18 e 3,5 μM HsRecA ou EcRecA, a reação foi incubada por 20 min a 37 °C, e então 3 μM ATP e 1 μM SSB foram adicionados e incubados por outros adicionais 10 min. Os decurso temporal mostrado representa o tempo da reação após adição de 20 μM nt ldsDNA M13mp18. (B) A percentagem de duplex substrato convertido em duplex circular (Produto) em função do decurso temporal.



(a)



(b)

Fonte: O autor.

5.5 A Dinâmica de Formação dos Filamentos sobre o ssDNA

A atividade ATPásica é comumente utilizada para avaliar indiretamente a dinâmica de formação e estabilidade dos filamentos da RecA sobre o ssDNA. A atividade ATPásica da proteína HsRecA foi avaliada por meio da espectroscopia de absorção no Ultra-Violeta em 340 nm, na presença do DNA circular simples fita M13mp18 (*circular single strand-cssDNA*), ATP e um sistema de regeneração de ATP, como descrito no capítulo de Material e Métodos. Todos os ensaios de atividade ATPásica foram realizados concomitantemente à proteína EcRecA, a fim de comparar a dinâmica de formação dos filamentos dessas duas proteínas *in vitro*. Todos os ensaios foram realizados em triplicatas.

Uma reação contendo as proteínas HsRecA ou EcRecA + cssDNA foi previamente incubada por 20 min a 37° C, seguido da adição de ATP e SSB, ambos adicionados no tempo $t = 0$ (*start*). A formação completa dos contínuos filamentos da proteína RecA sobre o cssDNA são preditos no regime estacionário máximo do perfil de hidrólise (BRENNER et al., 1987). O parâmetro cinético aparente k_{cat} foi determinado após o máximo do regime estacionário ser atingido para ambas as proteínas EcRecA e HsRecA, assumindo que um monômero de RecA se liga a cada 3 nucleotídeos de ssDNA. A determinação do valor aparente de k_{cat} se dá pela razão entre a velocidade máxima de hidrólise e o número de sítios de ligação cssDNA/RecA disponíveis na reação, neste trabalho adotou-se um número de sítios igual a 5/3, logo $k_{cat} = (\text{velocidade máxima de hidrólise}) / (5/3)$.

Nas condições da reação acima ensaiada os resultados mostraram um k_{cat} de $25,63 \pm 0,10$ e $28,91 \pm 0,11 \text{ min}^{-1}$, para a EcRecA e HsRecA, respectivamente (Figura (5.11) - Reação 1). A literatura reporta um k_{cat} de aproximadamente 30 min^{-1} para a proteína EcRecA quando essa está ligada ao ssDNA em condições similares (WEINSTOCK; MCENTEE; LEHMAN, 1981; BRENNER et al., 1987; BIANCO; WEINSTOCK, 1996). Observou-se em múltiplos experimentos que a atividade ATPásica ssDNA-dependente da proteína HsRecA foi consistentemente 10% maior que aquela exibida pela proteína EcRecA.

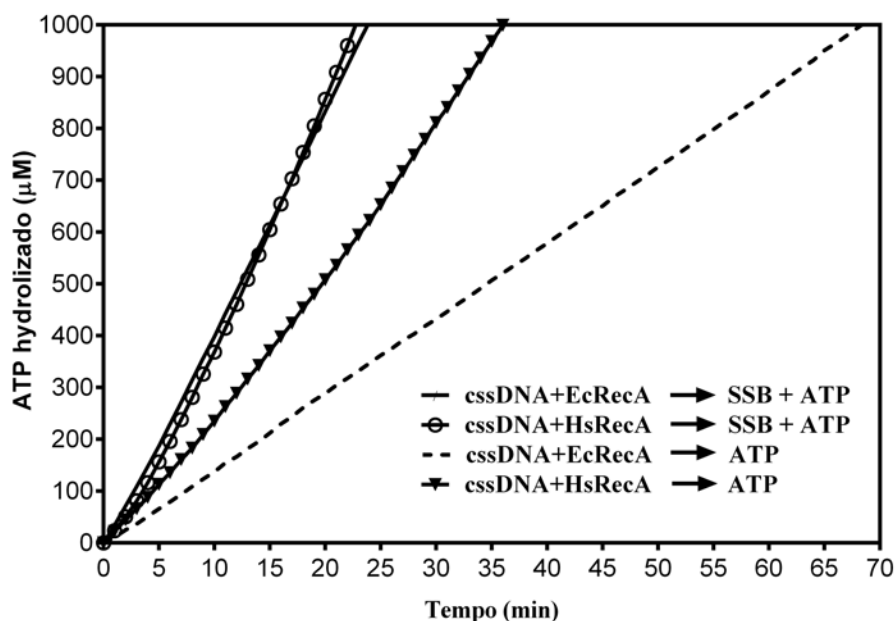
A proteína SSB é tipicamente utilizada em ensaios *in vitro* para remover a estrutura secundária do DNA e facilitar a formação dos contínuos filamentos da proteína RecA sobre o cssDNA quando adicionada após a incubação DNA-RecA. Entretanto, quando a proteína SSB é adicionada antes da RecA, essa fortemente inibe a nucleação da RecA sobre o DNA. Como consequência a atividade ATPásica é inicialmente suprimida seguida de um lento acréscimo (KOWALCZYKOWSKI et al., 1987; UMEZU; KOLODNER, 1994; ROY et al., 2009; SHEREDA et al., 2008). A proteína EcSSB foi utilizada em todos os experimentos, em parte por que a proteína HsSSB não estava disponível. Em estudos anteriores de proteínas RecA de bactérias como *Neisseria gonorrhoeae* (Ng), *Streptococcus pneumoniae* (Sp) e *Deinococcus radiodurans* (Dr) falharam em encontrar qualquer diferença nos efeitos da SSB quando a EcSSB foi trocada pela respectiva cognata SSB do organismo em reações padrões da RecA (STOHL et al., 2011; STEFFEN; BRYANT, 2000; KIM et al., 2002). A sequência de aminoácidos da proteína HsSSB mostrou-se ser a mais próxima da EcSSB (57% identidade/69% similaridade) do que a SSB das três espécies acima mencionadas (50%/63%, 31%/50% e 38%/4% para a NgSSB, SpSSB e DrSSB, respectivamente).

A fim de avaliar os efeitos da SSB sobre a atividade ATPásica da HsRecA, a proteína SSB foi omitida da reação descrita acima (Figura (5.11) - Reação 2). Nesta reação a proteína HsRecA mostrou uma atividade ATPásica 46,08% maior do que a EcRecA, com um k_{cat} de $16,49 \pm 0,11$ e $8,89 \pm 0,01 \text{ min}^{-1}$, respectivamente. Entretanto, nenhuma delas atingiu o mesmo nível de atividade ATPásica registrada na presença pós incubação RecA-DNA (Reação 1) em 70 min de monitoramento. Em contrapartida o perfil linear de hidrólise desde de $t = 0$ e permanecendo assim durante toda a reação indicou que não houve significante acréscimo de ambas as proteínas RecA ligando-se ao cssDNA durante o monitoramento.

Esse resultado corrobora com a informação estrutural obtida a partir da reconstrução da superfície eletrostática no sítio ATPásico dessas proteínas. A proteína HsRecA mostrou possuir um sítio ATPásico com aparente maior afinidade eletrostática pelo ATP, comparada a EcRecA. Uma vez que a auto-polimerização da RecA sobre o DNA é depen-

dente da ligação conjunta do ATP, a maior afinidade do sítio de ligação pelo ATP pode facilitar a auto-polimerização da HsRecA sobre o DNA.

Figura 5.11: Perfil da atividade ATPásica das proteínas HsRecA e EcRecA na presença e ausência da proteína SSB. Reação 1: 5 μM nt M13mp18 cssDNA e 3 μM HsRecA ou EcRecA foram previamente incubados por 20 min a 37 °C, seguindo da adição de 3 μM ATP e 0,5 μM SSB ($t=0, \textit{start}$). Reação 2: Os mesmos procedimentos da Reação 1 foram realizados na ausência da proteína SSB. O tempo $t = 0$ indica a adição de somente ATP. O valor aparente do K_{cat} foi determinado após o estado estacionário da RecA ligada ao cssDNA.



Fonte: O autor.

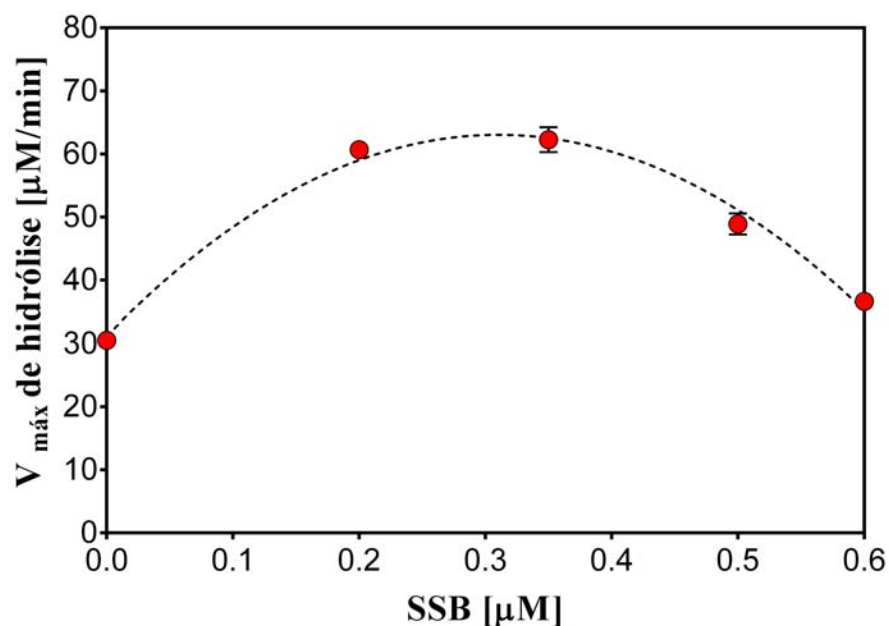
O estímulo na formação de contínuos filamentos pela proteína SSB causado intencionalmente na Reação 1 também foi observado nas proteínas RecA de *E. coli*, *Pseudomonas aeruginosa* e *Neisseria gonorrhoeae* (KOWALCZYKOWSKI et al., 1987; ROY et al., 2009; STOHL et al., 2011; BAITIN et al., 2006). Entretanto, o efeito observado foi diferente em *D. radiodurans*, ambas as proteínas SSB de *E. coli* e de *D. radiodurans* causam a supressão completa da atividade ATPásica da proteína DrRecA *in vitro*, indicando uma inibição da formação dos filamentos dessa proteína na presença da SSB e DNA (NGO et al., 2013).

Baseado no fato que a atividade ATPásica ser correlacionada com a quantidade

de RecA ligada ao DNA, os resultados observados na Reação 2 sugerem que proteína HsRecA possui elevada capacidade de se ligar a estrutura secundária do ssDNA. A omissão da proteína SSB do ensaio de atividade ATPásica causa um decréscimo de 65% (k_{cat} de $8,89 \pm 0,01 \text{ min}^{-1}$) na atividade ATPásica da proteína EcRecA, enquanto para a proteína HsRecA esta redução foi de 43% (k_{cat} de $16,49 \pm 0,11 \text{ min}^{-1}$) (Figura (5.11) - Reação 2). Em condições similares a atividade ATPásica da DrRecA (k_{cat} de $4,81 \pm 0,33 \text{ min}^{-1}$) foi ainda menor que a de EcRecA e HsRecA.

A fim de investigar a competição HsRecA-SSB pela ligação ao cssDNA, a atividade ATPásica dessa proteína foi realizada para diferentes concentrações da proteína SSB, os resultados para a velocidade máxima de hidrólise em função da concentração da proteína SSB estão mostrados na figura (5.12).

Figura 5.12: Velocidade máxima de atividade ATPásica da proteína HsRecA em diferentes concentrações da proteína SSB. As reações contendo $5 \mu\text{M}$ nt M13mp18 cssDNA e $3 \mu\text{M}$ HsRecA, foram previamente incubadas por 20 min a $37 \text{ }^\circ\text{C}$, seguido da adição de $3 \mu\text{M}$ ATP e diferentes concentrações da proteína SSB. Foi utilizado um polinômio de segunda ordem para indicar a tendência dos dados, essa representada pela linha tracejada.



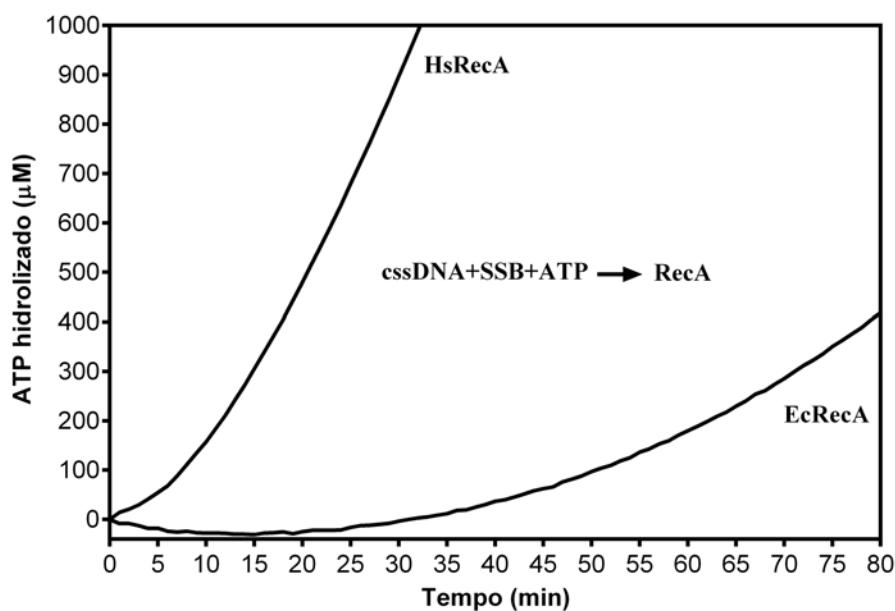
Fonte: O autor.

Observa-se pela figura (5.12) que a proteína SSB inibe a atividade ATPásica para

concentrações maiores que $0,35 \mu\text{M}$ SSB, e é incapaz de estimular para concentrações inferiores a $0,2 \mu\text{M}$. A máxima velocidade da atividade ATPásica foi obtida para $0,35 \pm 0,07 \mu\text{M}$ SSB, indicando que nessa condição a proteína HsRecA é capaz de deslocar a SSB ligada ao cssDNA e formar contínuos filamentos.

Neste trabalho também foi investigado o efeito do momento de adição da proteína SSB sobre a atividade ATPásica da proteína HsRecA. A proteína SSB, cssDNA e ATP foram incubados por 10 min a 37°C , permitindo a ligação da proteína SSB ao cssDNA, seguida da adição da proteína RecA (*start*). O perfil de atividade ATPásica é apresentado para esta condição na Figura (5.13).

Figura 5.13: Perfil de ATPase quando a SSB é adicionada antes da HsRecA ou EcRecA. Na reação $5 \mu\text{M}$ nt M13mp18 cssDNA, $3 \mu\text{M}$ ATP e $0,35 \mu\text{M}$ SSB foram previamente incubados por 10 min a 37°C , seguindo da adição de EcRecA ou HsRecA (tempo $t = 0$)



Fonte: O autor.

Como observado na figura (5.13), a proteína EcRecA apresenta uma baixa atividade ATPásica inicial, seguido de um gradual acréscimo, indicando um lento deslocamento da proteína SSB ligada ao cssDNA, em concordância com resultados encontrados na literatura (KOWALCZYKOWSKI et al., 1987; UMEZU; KOLODNER, 1994; SHEREDA

et al., 2008). Entretanto, observa-se que proteína HsRecA mostrou-se eficiente em transpassar a barreira oferecida pela proteína SSB para a nucleação, mostrando significativa atividade ATPásica em tempos inferiores a 30 minutos de monitoramento. Esse resultado sugere que a proteína HsRecA desloca efetivamente a proteína SSB ligada anteriormente ao cssDNA. A alta taxa de atividade ATPásica observada após de 15 minutos também sugere a formação de contínuos filamentos da proteína HsRecA sobre o cssDNA. A proteína HsRecA apresentou um k_{cat} de $24,66 \pm 1,45 \text{ min}^{-1}$, quase o mesmo valor determinado no ensaio anterior onde a proteína SSB foi adicionada após a incubação DNA+HsRecA. Em comparação a proteína EcRecA não conseguiu atingir o mesmo nível de atividade ATPásica durante os 80 min de monitoramento.

A proteína SSB mostrou-se importante nos processos de reparo de DNA por que estimula a primeira fase, que é a pré-sináptica polimerização da RecA sobre o ssDNA, e a última fase, que é a reação de trocas de fitas (RADDING et al., 1983). Esta ação da proteína SSB sobre o mecanismo de reparo têm se mostrado idêntico em sistemas homologos e heterólogos, ou seja, mesmo utilizando uma proteína SSB de um organismo diferente, o efeito de estímulo sobre a formação dos filamentos de RecA não se altera (STEFFEN; BRYANT, 2000; KIM et al., 2002; STOHL et al., 2011). Estudos anteriores mostram que algumas mutações na proteína EcRecA (*recA730* (E38K) (WANG; CHANG; HUNG, 1993), *recA803* (V37M) (MADIRAJU; TEMPLIN; CLARK, 1988), *recA2020* (T121I) (WANG; SMITH, 1986) e *recA441* (E38K/I298V) (THOMAS; LLOYD, 1983; VOLKERT; MARGOSSIAN; CLARK, 1984; WANG; CHANG; HUNG, 1993)) confere a EcRecA a mesma habilidade de transpassar a barreira oferecida para a nucleação observada na proteína HsRecA selvagem. Na proteína HsRecA esses resíduos mostram-se conservados, exceto pela V37 e E38, nos quais correspondem a I44 e Q45, respectivamente.

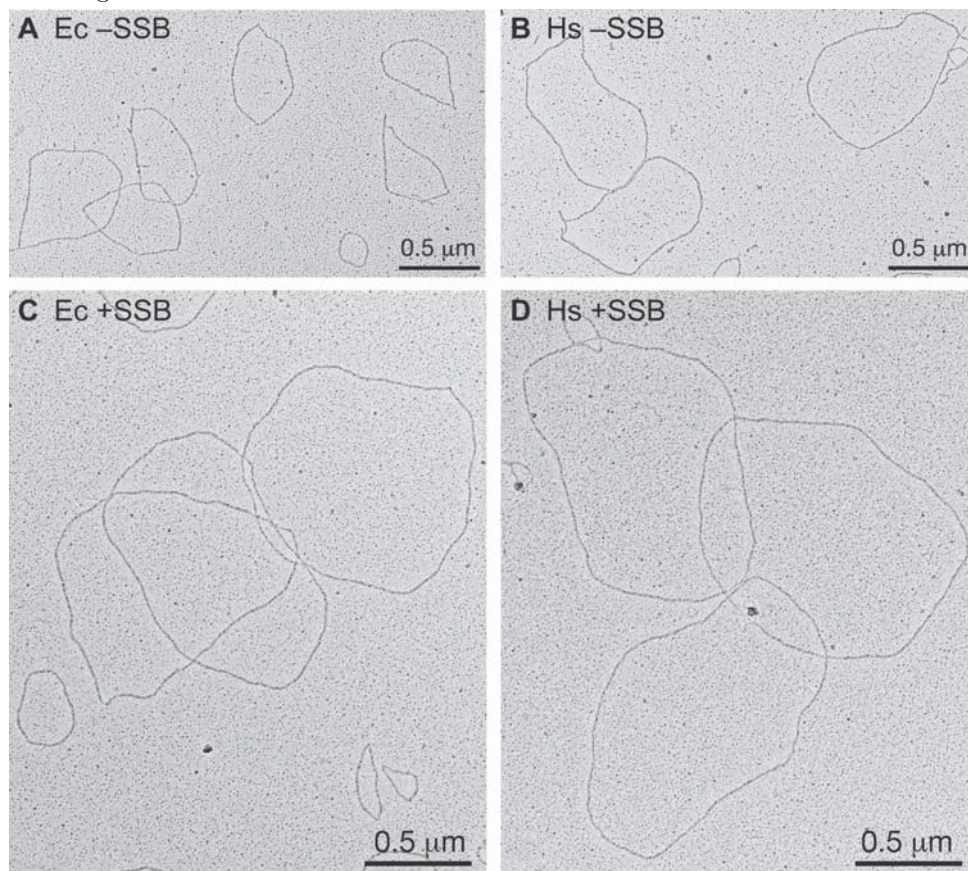
Similar a proteína EcRecA, a HsRecA possui um C-terminal que é predominantemente negativamente carregado (4 ácido glutâmico e uma lisina), sobre 16 amino resíduos da porção C-terminal. A concentração de resíduos negativos no C-terminal da EcRecA é ainda maior (7 de um total de 17 resíduos, com nenhum positivamente carregado). A eliminação do C-terminal da EcRecA também resulta em proteína que desloca rapida-

mente a SSB ligada ao DNA (EGGLER; LUSETTI; COX, 2003). Analisando estes fatos observa-se que a interação entre o domínio C-terminal da proteína RecA (não estruturado em todas as proteínas RecA, incluindo a HsRecA que será apresentada posteriormente) e a superfície na região do resíduo E38 pode mascarar a superfície requerida para deslocar rapidamente a SSB. Ainda, o decréscimo de carga negativa da proteína HsRecA, acoplado com as substituições 44 and 45, acima citadas, talvez diminuam o efeito máscara da região C-terminal e permita o rápido deslocamento da SSB observado nos ensaios de atividade ATPásica.

5.6 Visualização dos filamentos por Microscopia Eletrônica

A microscopia eletrônica foi utilizada para determinar as características dos filamentos formados pelas proteínas EcRecA e HsRecA sobre o cssDNA M13mp18, na ausência da proteína SSB, painéis A e B da figura (5.14), e quando a SSB foi adicionada depois da pré-incubação RecA-cssDNA, painéis C e D. Essas duas condições são similares aquelas exploradas por meio da atividade ATPásica.

Figura 5.14: Imagens de microscopia eletrônica dos filamentos das proteínas EcRecA e HsRecA na ausência e presença da proteína SSB. Em (a) filamentos de EcRecA-cssDNA na ausência da proteína SSB, (b) filamentos de HsRecA-cssDNA na ausência da proteína SSB, (c) filamentos de EcRecA-cssDNA na presença da proteína EcSSB adicionada após 20 min, (d) filamentos de HsRecA-cssDNA na presença da proteína EcSSB adicionada após 20 min. O sistema de reação contendo $6,7 \mu\text{M}$ HsRecA ou EcRecA e $20 \mu\text{M}$ cssDNA M13mp18 foi incubado a 37°C por 20 min, e então $2 \mu\text{M}$ SSB foi adicionada ou não. Após 10 min, os filamentos foram estabilizados com a adição de $3 \mu\text{M}$ ATP γ S, seguido de incubação por 3 min e depositados no grind de Alcian.



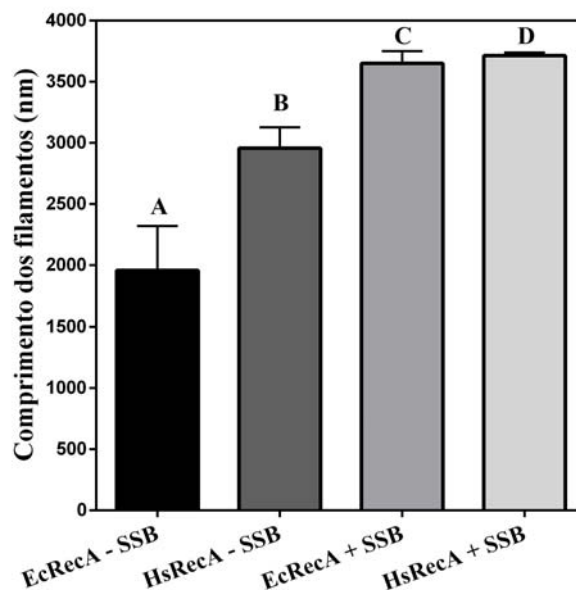
Fonte: O autor.

Na ausência da proteína SSB, a proteína EcRecA forma pequenos filamentos circulares, com algumas regiões de descontinuidades. Por outro lado, a proteína HsRecA forma filamentos mais abertos e com menos descontinuidades. A adição da proteína SSB após a pré-incubação RecA-cssDNA, elimina as diferenças nas características físicas detectáveis entre os filamentos, o que torna ambos os filamentos regulares, grandes e completos. Os longos filamentos formados pela proteína HsRecA na ausência da proteína SSB indica uma capacidade melhorada de estender os filamentos sobre regiões de estrutura

secundária do ssDNA. Importante notar que nestas imagens, somente os filamentos da RecA são visíveis. O ssDNA não é visível a menos que a SSB esteja ligada, dando a essas regiões uma aparência de pontos aglomerados.

As impressões observadas a partir desse primeiro levantamento visual do *grinds* na ME foi confirmada pela medida do comprimento dos filamentos. O comprimento de dez filamentos da EcRecA e HsRecA-*cssDNA* foram determinados na ausência e presença da SSB. Na presença da SSB, as proteínas HsRecA e EcRecA exibiu comprimentos de filamentos indistinguíveis. Entretanto, se a proteína SSB está ausente a redução nos comprimentos dos filamentos foram de 46.39 % e 20.37% para a EcRecA e HsRecA, respectivamente. Os filamentos HsRecA-*cssDNA* mostraram serem mais longos do que os de EcRecA-*cssDNA*, esses resultados podem são apresentados no gráfico da figura (5.15).

Figura 5.15: Comprimento médio de 10 filamentos das proteínas EcRecA e HsRecA sobre o *cssDNA* M13mp18, na presença e ausência da proteína SSB. As letras acima de cada barra se referem aos painéis A-D da figura (5.14).



Fonte: O autor.

Uma medida indireta pode ser realizada para determinar o tamanho do passo helicoidal do filamento a partir dos dados de ME, na condição dos painéis C e D da

figura (5.14), a proteína RecA supostamente ocupa todos os sítios de ligação sobre o cssDNA M13mp18, e pode ser considerada como o estado ativo da proteína. O cssDNA M13mp18 possui 7249 pb, dividindo o tamanho total do filamento pelos 7249 pb, teríamos aproximadamente 0,512 nm/pb para a proteína HsRecA, vezes os 18 pares de base do hexâmero o tamanho do passo helicoidal seria de 92,2 Å, similar aos 91,3 Å determinados pelos dados de DRX.

5.7 A Polimerização da Proteína RecA no Cristal

Embora a estrutura da proteína EcRecA tenha fornecido informações importantes no que diz respeito as suas funções no sistema de reparo de DNA, proteínas de diferentes organismos que exercem funções semelhantes a da RecA foram caracterizadas funcionalmente e estruturalmente. Essas proteínas similares a RecA também formam filamentos ao redor do DNA, catalisam a reação de trocas de fitas de DNA e utilizam a energia proveniente da hidrólise do ATP como fonte de energia.

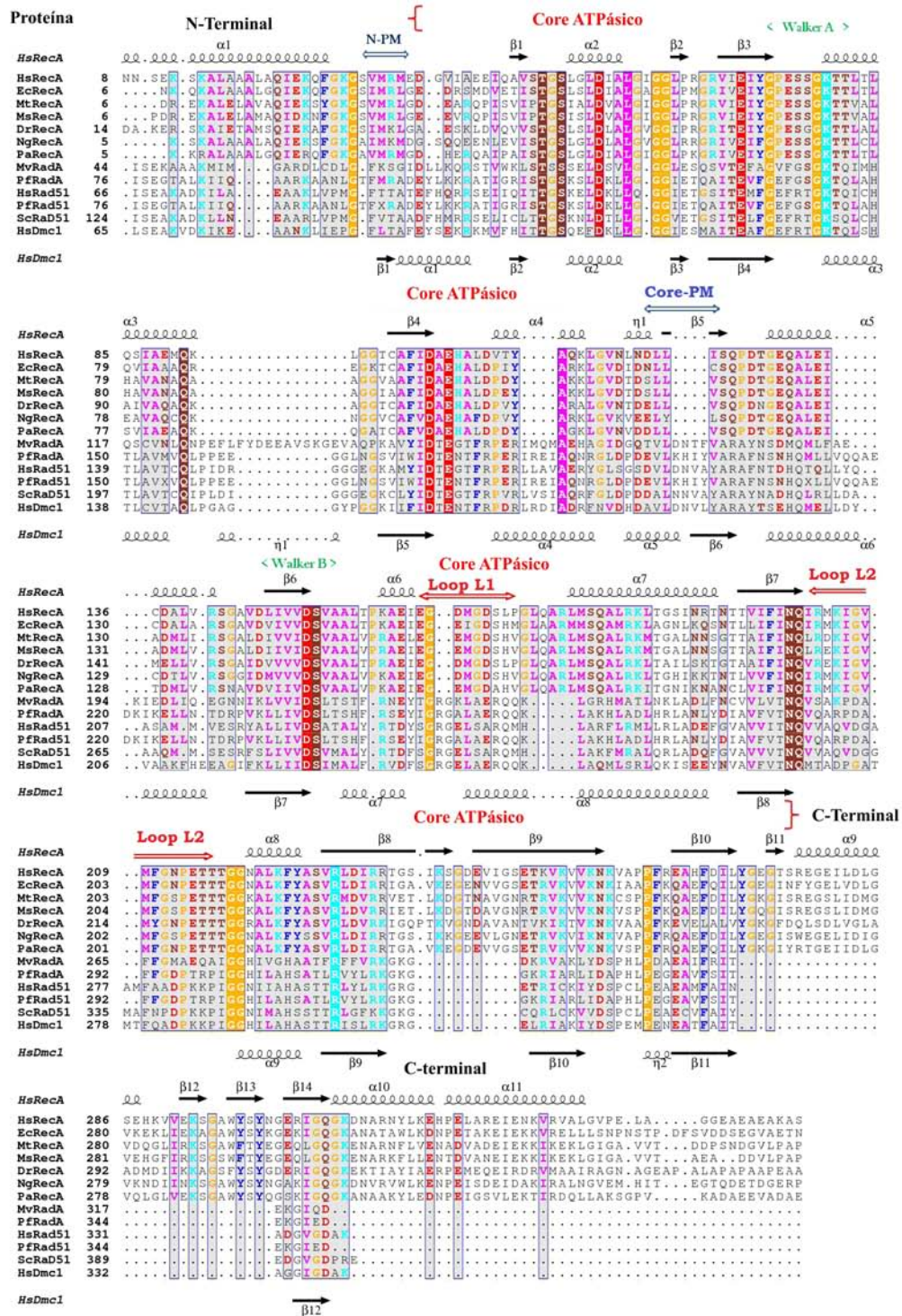
A proteína RecA não é um caso isolado, mas ela faz parte de uma família de proteínas encontradas nos três domínios da vida. O core ATPásico presente no domínio central, está altamente conservado na RecA bacteriana e em proteínas homólogas a RecA, incluindo a proteína RadA de Arquea e a proteínas Rad51 e Dmc1 de Eucariotos, bem como outras similaridades funcionais e estruturais (CHEN et al., 2007; WU et al., 2004; KINEBUCHI et al., 2004; OKOROKOV et al., 2010; SANDLER et al., 1996; CHANG et al., 2009).

Na busca por outras similaridades funcionais e estruturais, proteínas RecA de bactéria e proteínas membros da família RecA (RadA, Dmc1 e Rad51), foram alinhadas utilizando suas estruturas primária e terciária, a fim de comparar regiões em comum. O alinhamento está mostrado na figura (5.16).

Todas as sequências das proteínas alinhadas na figura (5.16) desempenham um papel no sistema de reparo a danos no DNA. No alinhamento parte do N-terminal foi removido para melhor visualização, dando prioridade as sequências de bactéria. Motivos

funcionais estão indicados acima de suas correspondentes sequências de aminoácidos: N-MP e core-MP, N-terminal e core motivos de polimerização, respectivamente, estes serão melhores discutidos abaixo. Os motivos de ligação do ATP Walker A e B, os *loops* de ligação do DNA L_1 e L_2 também estão indicados acima das respectivas sequências de aminoácidos. A posição de início e término dos domínios N-terminal, core ATPásico e do C-terminal estão indicados acima das sequências. A estrutura secundária das proteínas HsRecA e HsDmc1 estão indicadas acima e abaixo, respectivamente.

Figura 5.16: Alinhamento sequencial e estrutural das proteínas de RecA de *H. seropedicae* (HsRecA), *E. coli* (EcRecA), *M. tuberculosis* (MtRecA), *M. smegmatis* (MsRecA), *D. radiodurans* (DrRecA), *N. gonorrhoeae* (NgRecA), *P. aeruginosa* (PaRecA), e da família RecA de *M. gonorrhoeae* (MvRadA), *P. furiosus* (RadA), *H. sapiens* (HsRad51), *P. furiosus* (Rad51), *S. cerevisiae* (ScRad51 e Dmc1).



Fonte: O autor.

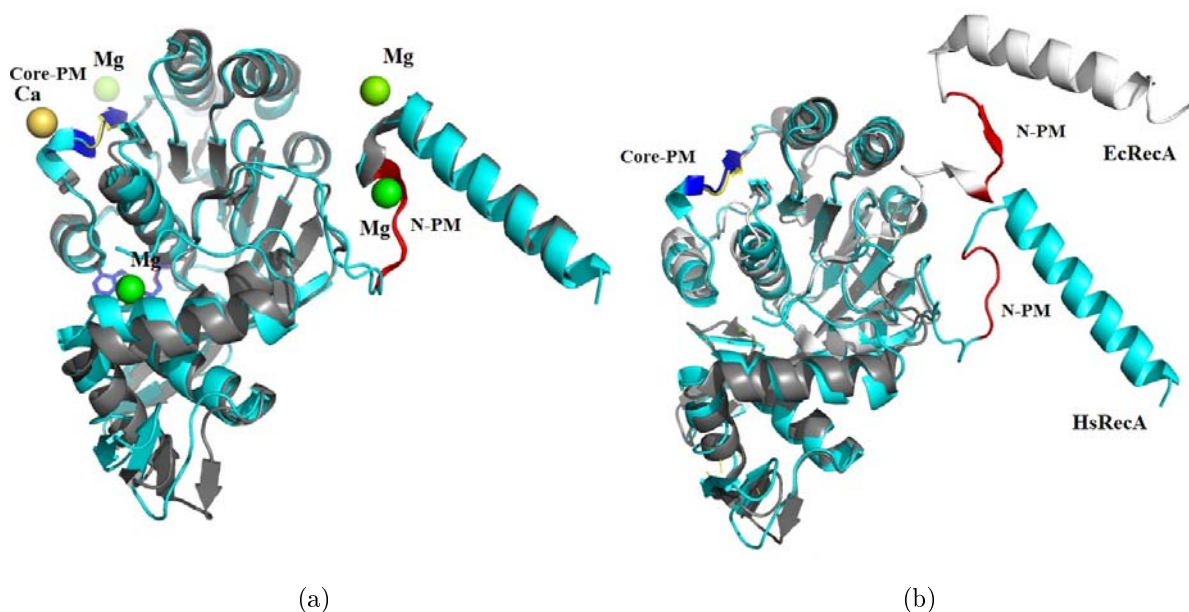
As regiões responsáveis pela polimerização foram primeiramente identificadas nas proteínas RadA/Rad51/Dcm1 (CHEN et al., 2007; WU et al., 2004; LEE; WANG, 2009; WANG; CHEN; WANG, 2008; DU; LUO, 2012). Elas compreendem resíduos nos domínios N-terminal (figura (5.16), região N-MP) e no core ATPásico (figura (5.16), região Core-MP). A estrutura tridimensional de diversas proteínas da família RecA revelou que a região N-MP é um motivo de polimerização, que consiste de uma pequena fita β que está imediatamente precedente ao core ATPásico e possui resíduos hidrofóbicos conservados.

Utilizando o alinhamento e a estrutura tridimensional das proteínas RecAs de *E. coli* (pdb:3cmu, 1xms, 1u98 e 1xmv), *D. radiodurans* (pdb:1xp8), *M. smegmatis* (pdb:1ube, 1ubf e 1ubg), *M. tuberculosis* (pdb:1mo3, 1mo5, 1mo6 e 1g18) e a estrutura da proteína HsRecA, observou-se que em bactérias este motivo de polimerização (N-MP) consiste de uma sequência (Serina, Valina ou Isoleucina, Metionina, Arginina ou Lisina, Leucina, Glicina; resíduos 25-30 na sequência da EcRecA, depois de eliminar a Metionina do N-terminal, e resíduos 30-35 na sequência da HsRecA). Esta sequência interage com outra fita β do core ATPásico (Core-MP) do adjacente monômero em outra conservada sequência (Aspartato, Asparagina, Leucina, Leucina, Leucina, Valina, Cisteína, Serina; resíduos 113-118 e 118-123 nas sequências da EcRecA e HsRecA, respectivamente).

No motivo N-MP, a hidrofóbica e conservada fenilalanina encontrada nas proteínas RadA, Rad51 e Dmc1 está substituída por outro resíduo hidrofóbico, valina ou isoleucina, nas estruturas da proteína RecA. Contudo, nenhuma dessas substituições afetam as ligações de hidrogênio com os resíduos do Core-MP. A conservada Metionina encontrada no N-MP de bactérias, talvez possua uma importante função na ligação/reconhecimento da fita β hidrofóbica do motivo de polimerização presente no core ATPásico, promovendo a polimerização dos filamentos da proteína RecA sobre o DNA substrato. Ainda, observando as estruturas das proteínas RecA inativas, a pequena fita β do N-MP assume a conformação de um motivo em *loop*, chamado de β -*loop*, figura (5.15a); este motivo também é observado na estrutura da proteína HsRecA. Já na estrutura do filamento pré-sináptico EcRecA-ssDNA este motivo permanece como uma conformação de fita β , como encontrado na família das proteínas RecA, figura (5.15b) e este motivo é uma fita

antiparalela com respeito ao segundo motivo no filamento pré-sináptico EcRecA-ssDNA.

Figura 5.17: Superposição da estrutura monomérica da proteína EcRecA inativa e pré-sináptica sobre a inativa da proteína HsRecA. As proteínas EcRecA e HsRecA estão coloridas em tons de cinza e ciano, respectivamente. (a) Superposição da estrutura monomérica inativa da proteína EcRecA (pdb:1xmv) sobre a proteína HsRecA. Na estrutura da EcRecA o íon Mg^{2+} interage com o N-MP, e na estrutura da proteína HsRecA o íon Ca^{2+} interage com Core-MP. (b) Superposição da estrutura monomérica pré-sináptica da proteína EcRecA (pdb:3cmu:A) sobre a proteína HsRecA. O ADP, ATP e ADP- AlF_4 e os íons Mg^{2+} e Ca^{2+} , o ssDNA ligado nas estruturas foram removidos para melhor clareza.



Fonte: O autor.

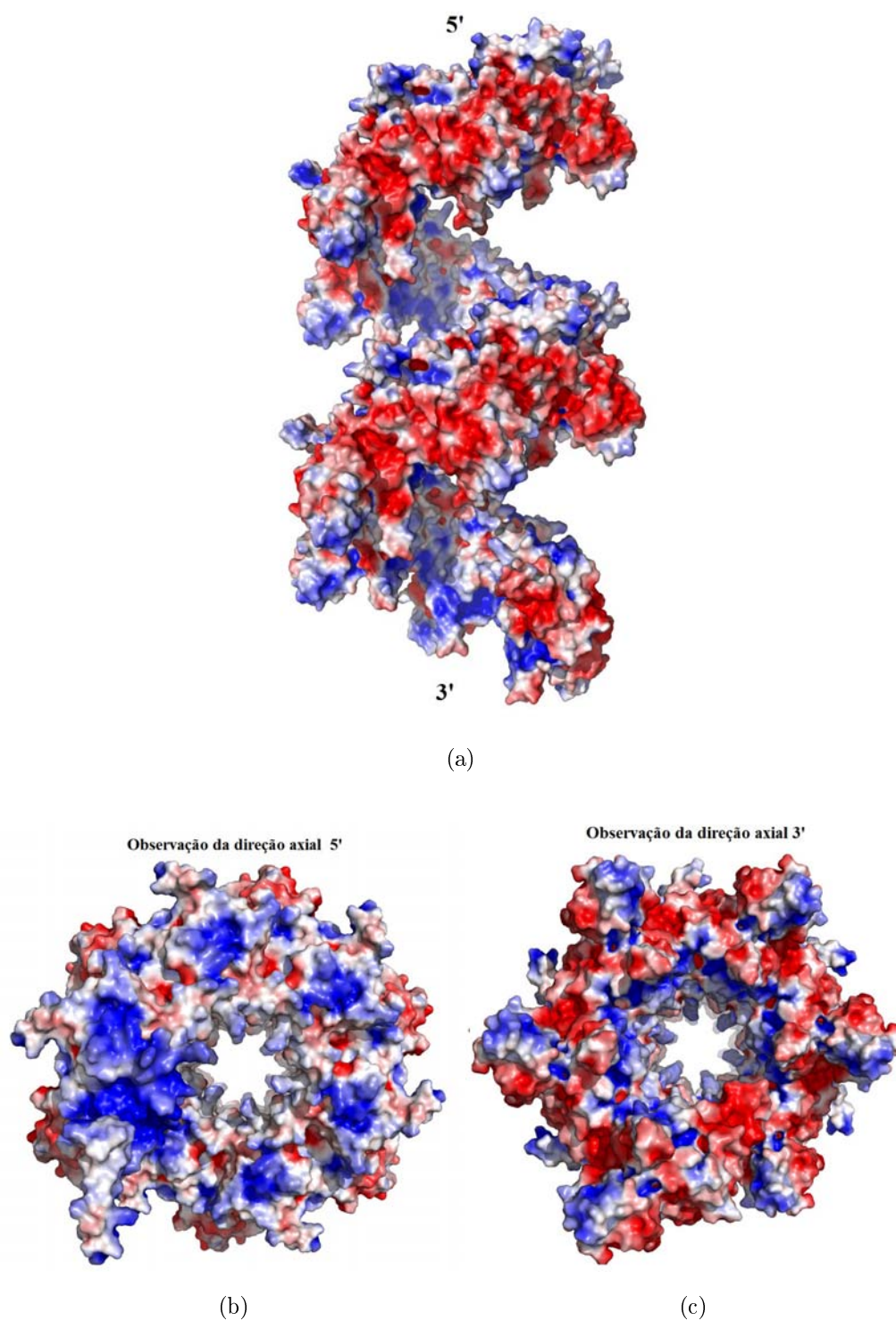
O íon Ca^{2+} no core-MP da proteína HsRecA está coordenado pelos resíduos Asparagina 119 (Asn119) e Aspartato 120 (Asp120), e o íon Mg^{2+} na estrutura da proteína EcRecA está coordenado pelos resíduos Glutamina 18 (Glu18), Lisina 23 (Lys23) e Isoleucina 26 (Ile26) (XING; BELL, 2004). Ambos são íons divalentes e possivelmente possuem um importante papel na polimerização e estabilidade dos filamentos de RecA como um filamento de nucleoproteína helicoidal. Estudos anteriores mostraram que o excesso de Mg^{2+} (relativo a quantidade de ATP disponível) ajuda a produzir a conformação estendida do filamento da proteína RecA, requerida para o pareamento e reação de trocas de fitas *in vitro* (LUSETTI; SHAW; COX, 2003). Outro ponto observado é que o resíduo Glutamato 36 (Glu36) da proteína HsRecA está localizado na fronteira de interface

monômero-monômero, posteriormente ao N-MP, esse talvez interaja com os íons em excesso, pois apresenta carga negativa em pH fisiológico, e também com outros resíduos próximos, possivelmente aumentando a interação de superfície subunidade-subunidade.

Para tentar entender as interações monômero-monômero, foi reconstruída a superfície da proteína HsRecA com base no seu potencial eletrostático. Devido a porção C-terminal de cada monômero, exposta externamente do filamento e a região de solvente, existe uma região relativamente grande carregada negativamente na superfície externa do filamento da proteína HsRecA. Também existe uma visível diferença de cargas positivas e negativas entre a região interna e externa do filamento, figura (5.16a). Observa-se que o filamento da proteína HsRecA possui um aumento de carga positiva dentro do filamento, que é consistente com o fato do ssDNA ligar-se na parte interna desse. Especula-se que a alta capacidade da proteína HsRecA ligar-se a estrutura secundária do ssDNA observada nos ensaios de atividade ATPásica decorre em partes da natureza eletrostática interna do seu filamento.

Importantes domínios que afetam a polimerização do filamento estão ordenados a longo do eixo do filamento, observa-se que há uma mudança de polaridade de negativo para positivo relativo a direção 3' e 5' de ligação ao ssDNA no filamento (comparar figuras (5.16b e 5.16c). Especula-se que as interações eletrostáticas facilitarão a polimerização do filamento, após os monômeros serem estabilizados por interações hidrofóbicas entre o N-PM de um monômero e o core-MP do subsequente monômero.

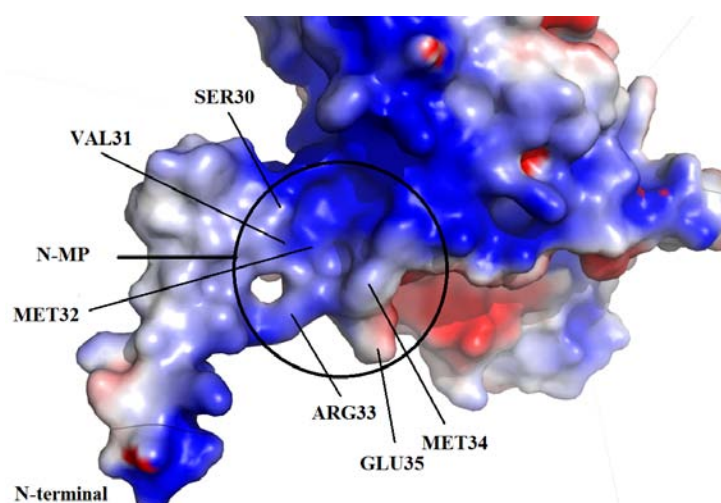
Figura 5.18: Distribuição de potencial eletrostático de superfície da proteína HsRecA. (a) visão lateral de dois subsequentes hexâmeros. (b) e (c) observação nas direções 5' e 3', relativo a direção axial do filamento, respectivamente. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).



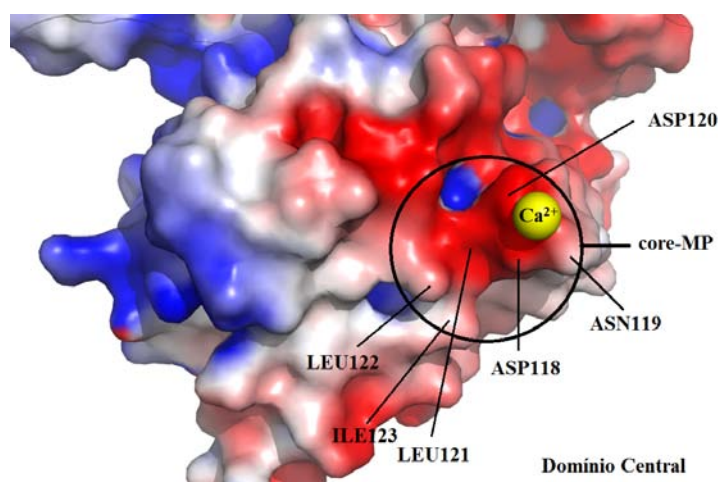
Fonte: O autor.

Os motivos de polimerização do monômero da proteína HsRecA mostrou-se possuir cargas opostas na superfície, sendo que o N-PM possui carga positiva e o core-PM carga negativa, figuras (5.17a e 5.17b), respectivamente, esse resultado corrobora com a suposição que interações eletrostáticas facilitariam a polimerização do filamento, aliado a geometria da região de interação ser do tipo chave-fechadura.

Figura 5.19: Distribuição de potencial eletrostático na região dos motivos de polimerização N-MP e core-MP para a proteína HsRecA. Os níveis de carga na representação do potencial eletrostático de superfície variam de $-3kT/e$ (vermelho) para $+3kT/e$ (azul).



(a)



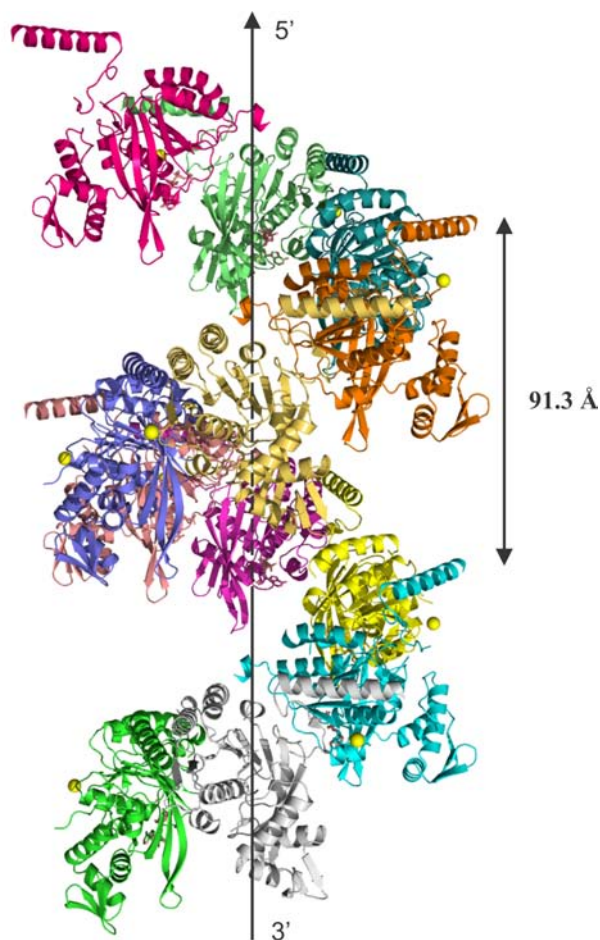
(b)

Fonte: O autor.

5.8 Passo helicoidal Hexamérico

Na estrutura do cristal os monômeros da proteína HsRecA empacotam-se em um filamento helicoidal, devido a simetria cristalográfica de ordem 61 forma-se um hexâmero que possui passo helicoidal de 91,3 Å (eixo-c) na estrutura cristalina. A unidade biológica da proteína RecA formada em solução também apresenta-se na forma predominante de hexâmeros, logo o hexâmero do cristal possui informações importantes a respeito da forma funcional da proteína RecA. A figura (5.20) mostra dois consecutivos hexâmeros reconstruídos com o modelo cristalográfico da figura (5.4) para a proteína HsRecA.

Figura 5.20: Dois consecutivos hexâmeros do cristal da proteína HsRecA. Na figura é mostrado o passo helicoidal do hexâmero.



Fonte: O autor.

O valor do passo helicoidal de 91,3 Å para o hexâmero da HsRecA do cristal, e 92,2 Å via ME, está na faixa de valores de 90-100 Å determinados para os filamentos da proteína EcRecA, na sua conformação estendida, na presença de DNA, ATP γ S ou ATP (EGELMAN; STASIAK, 1986). Nas estruturas com conformação comprimida o passo helicoidal varia de 65-85 Å, sendo formado pela RecA sozinha, ligada ao ADP, tanto na ausência de DNA ou ligada ao ssDNA ou dsDNA (STORY; STEITZ, 1992; GALLETTO et al., 2006; RAJAN; BELL, 2004; DATTA et al., 2003; CHANDRAN et al., 2015; ELLOUZE et al., 1995). No entanto o complexo HsRecA-ADP/ATP do cristal forma um filamento que possui característica de uma forma estendida da proteína RecA. Nos estudos anteriores de Egelman e Stasiak (1986) constatou-se que quando a RecA encontra-se ligada ao ATP (conformação ativa) existe uma re-orientação de todo o domínio ATPásico, e o sítio ATPásico move-se para a interface entre os monômeros. Entretanto, na presença de ADP (conformação inativa) o sítio ATPásico fica exposto a região de solvente. Essas observações também foram verificadas na estrutura da EcRecA em complexo do ssDNA e dsDNA (CHEN; YANG; PAVLETICH, 2008).

O empacotamento estrutural apresentado na figura (5.20) é o mesmo observado no estado inativo, porém apresenta um excepcional passo alongado. Algumas hipóteses foram levantadas a respeito do comportamento observado. Estudos anteriores mostraram que alta concentração de sal induz a hidrólise de ATP da proteína RecA e facilitaria a cristalização da conformação estendida na ausência de ssDNA (HEWAT; RUIGROK; DICAPUA, 1991; CAPUA et al., 1992; PUGH; COX, 1988a). Curiosamente, polietileno glicol (PEG) de vários tamanhos também aumentaram a atividade ATPásica da proteína RadA de *M. voltae* na ausência de DNA (WU et al., 2004). Baseado nestes fatos, é plausível de sugerir que a alta concentração de sal e PEG utilizada no solução de cristalização (0,25 M $CaCl_2$, 14 % *w/v* PEG 3350 and 8 % *w/v* PPG 400) poderia explicar a cristalização da proteína HsRecA na forma de um filamento estendido, mesmo na ausência de DNA.

Capítulo 6

Conclusão e Trabalhos Futuros

A estrutura tridimensional da proteína HsRecA, resolvida numa resolução de 1,7 Å, é formada por um pequeno domínio N-terminal, um domínio central contendo um sítio ATPásico e um grande domínio C-terminal. No cristal, a proteína HsRecA forma um filamento helicoidal dextrogiro. A interação entre monômero-monômero no filamento ocorre por meio de dois motivos hidrofóbicos, localizados na região N-terminal (N-MP) e no core ATPásico (core-MP). Interações eletrostáticas entre os hexâmeros facilitam a polimerização, posteriormente as interações hidrofóbicas entre o N-MP e o core-MP estabilizam o filamento. Íons divalentes como Ca^{2+} e Mg^{2+} , ligados ao core-MP e N-MP, assumem posições ocupadas por resíduos, favorecem a polimerização por meio da estabilização das interações entre subunidade-subunidade. O hexâmero da proteína HsRecA determinada neste trabalho possui excepcional passo helicoidal estendido, característico da forma funcional dessa proteína.

A proteína HsRecA polimeriza-se espontaneamente sobre o ssDNA na presença de ATP e ATP γ S e apresenta a atividade ATPásica dependente de ssDNA 10% superior a EcRecA; A proteína HsRecA apresenta um k_{cat} de $28,91 \pm 0,11 \text{ min}^{-1}$ enquanto EcRecA apresenta um k_{cat} de $25,63 \pm 0,10 \text{ min}^{-1}$. A atividade de trocas de fitas também mostrou-se ser maior para a proteína HsRecA comparada a EcRecA.

Se a proteína SSB é adicionada após a proteína HsRecA no sistema, a proteína SSB facilita a polimerização da HsRecA e quando é adicionada antes, não é observado o

bloqueio da polimerização, como ocorre nas demais proteínas RecA homólogas caracterizadas. A dinâmica de formação dos filamentos também revelou que a proteína HsRecA acessa de forma mais aficiente a estrutura secundária do DNA, e é menos dependente da proteína regulatória SSB para polimeriza-se sobre ssDNA que a EcRecA.

Segundo a literatura, a polimerização e dissociação da proteína RecA do DNA é regulada por diversas outras proteínas. Os resultados encontrados neste trabalho sugerem que a proteína SSB pode controlar a polimerização da proteína RecA sobre o DNA de uma forma não específica e independente da espécie.

Proteínas RecA de bactérias são dependentes das proteínas RecO e RecR para deslocar a SSB do DNA, mas a proteína HsRecA parece exigir menos assistência. As proteínas SSB, RecO e RecR de *H. seropedicae* podem possuir uma nova função nos processos de reparo de DNA dessa bactéria, que ainda não foram explorados.

Para trabalhos futuros experimentos envolvendo a caracterização estrutural e funcional das proteínas SSB, RecO e RecR de *H. seropedicae* precisam ser realizados, a fim investigar os mecanismos de atuação dessas proteínas na dinâmica de polimerização e estabilidades dos filamentos da proteína HsRecA.

Referências Bibliográficas

ADAMS, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. **Acta crystallographica. Section D, Biological crystallography**, International Union of Crystallography, v. 66, n. Pt 2, p. 213–21, 2010.

BAITIN, D. M. et al. Distinguishing characteristics of hyperrecombinogenic RecA protein from *Pseudomonas aeruginosa* acting in *Escherichia coli*. **Journal of bacteriology**, v. 188, n. 16, p. 5812–20, 2006.

BAKER, N. A. et al. Electrostatics of nanosystems: application to microtubules and the ribosome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 18, p. 10037–41, 2001.

BAYNTON, K.; FUCHS, R. P. Lesions in DNA: hurdles for polymerases. **Trends in biochemical sciences**, v. 25, n. 2, p. 74–9, 2000.

BEDALE, W. A.; COX, M. Evidence for the coupling of ATP hydrolysis to the final (extension) phase of RecA protein-mediated DNA strand exchange. **The Journal of biological chemistry**, v. 271, n. 10, p. 5725–32, 1996.

BELL, C. E. Structure and mechanism of *Escherichia coli* RecA ATPase. **Molecular microbiology**, v. 58, n. 2, p. 358–66, 2005.

BERMAN, H. M. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 2000.

BIANCO, P. R.; WEINSTOCK, G. M. Interaction of the RecA protein of *Escherichia coli* with single-stranded oligodeoxyribonucleotides. **Nucleic Acids Research**, v. 24, n. 24, p. 4933–4939, 1996.

BLOW, D. **Outline of Crystallography for Biologists**. 3. ed. New York: [s.n.], 2002. 248 p.

BRANDEN, C.; TOOZE, J. **Introduction to Protein Structure**. [S.l.]: Garland Pub., 1999. (Introduction to Protein Structure Series).

BRENNER, S. L. et al. recA protein-promoted ATP hydrolysis occurs throughout recA nucleoprotein filaments. **Journal of Biological Chemistry**, v. 262, n. 9, p. 4011–4016, 1987.

CAPUA, E. D. et al. Activation of RecA protein. The open helix model for LexA cleavage. **Journal of molecular biology**, v. 226, n. 3, p. 707–19, 1992.

- CHANDRAN, A. V. et al. Structural studies on Mycobacterium tuberculosis RecA: molecular plasticity and interspecies variability. **Journal of biosciences**, v. 40, n. 1, p. 13–30, 2015.
- CHANET, R. et al. Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase. **Molecular and cellular biology**, American Society for Microbiology (ASM), v. 16, n. 9, p. 4782–9, sep 1996.
- CHANG, C. F. et al. Cryo electron microscopy of unstained, unfixed RecA-*cssDNA* complexes. **Journal of ultrastructure and molecular structure research**, v. 100, n. 2, p. 166–72, 1988.
- CHANG, Y.-W. et al. Three new structures of left-handed RADA helical filaments: structural flexibility of N-terminal domain is critical for recombinase activity. **PloS one**, v. 4, n. 3, p. e4890, 2009.
- CHEN, L. T. et al. Crystal structure of the left-handed archaeal RadA helical filament: Identification of a functional motif for controlling quaternary structures and enzymatic functions of RecA family proteins. **Nucleic Acids Research**, v. 35, n. 6, p. 1787–1801, 2007.
- CHEN, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta crystallographica. Section D, Biological crystallography**, v. 66, n. Pt 1, p. 12–21, jan 2010.
- CHEN, Z.; YANG, H.; PAVLETICH, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. **Nature**, v. 453, n. 7194, p. 489–484, 2008.
- 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, may 1995.
- COSTECHAREYRE, D.; BERTOLLA, F.; NESME, X. Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. **Molecular biology and evolution**, v. 26, n. 1, p. 167–76, 2009.
- COURCELLE, J. et al. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. **Genetics**, v. 158, n. 1, p. 41–64, 2001.
- COX, M. M. RecA Filament Dynamics during DNA Strand Exchange Reactions. **Journal of Biological Chemistry**, v. 272, n. 17, p. 11063–11073, 1997.
- COX, M. M. Recombinational DNA repair of damaged replication forks in *Escherichia coli*: questions. **Annual review of genetics**, v. 35, p. 53–82, 2001.
- COX, M. M. The bacterial RecA protein as a motor protein. **Annual review of microbiology**, v. 57, p. 551–577, 2003.
- COX, M. M.; LEHMAN, I. R. Directionality and polarity in recA protein-promoted branch migration. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 10, p. 6018–22, 1981.

CROWTHER, R. A.; BLOW, D. M. A method of positioning a known molecule in an unknown crystal structure. **Acta Crystallographica**, International Union of Crystallography, v. 23, n. 4, p. 544–548, 1967.

CULLITY, B. D.; STOCK, S. R. **Elements of X-ray diffraction**. [S.l.]: Prentice Hall, 2001.

DATTA, S. et al. Crystal structures of Mycobacterium smegmatis RecA and its nucleotide complexes. **Journal of bacteriology**, v. 185, n. 14, p. 4280–4, 2003.

DAUTER, Z. Data-collection strategies. **Acta Crystallographica Section D Biological Crystallography**, International Union of Crystallography, v. 55, n. 10, p. 1703–1717, 1999.

DICAPUA, E. et al. Complexes of RecA protein in solution. A study by small angle neutron scattering. **Journal of molecular biology**, v. 214, n. 2, p. 557–70, 1990.

DOLINSKY, T. J. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. **Nucleic acids research**, v. 35, n. Web Server issue, p. W522–5, jul 2007.

DREES, J. C. et al. A RecA filament capping mechanism for RecX protein. **Molecular cell**, v. 15, n. 5, p. 789–98, 2004.

DREES, J. C.; LUSETTI, S. L.; COX, M. M. Inhibition of RecA protein by the *Escherichia coli* RecX protein: modulation by the RecA C terminus and filament functional state. **The Journal of biological chemistry**, v. 279, n. 51, p. 52991–7, 2004.

DU, L.; LUO, Y. Structure of a hexameric form of RadA recombinase from *Methanococcus voltae*. **Acta crystallographica. Section F, Structural biology and crystallization communications**, International Union of Crystallography, v. 68, n. Pt 5, p. 511–6, 2012.

EGELMAN, E. H.; STASIAK, A. Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP. **Journal of molecular biology**, v. 191, n. 4, p. 677–97, 1986.

EGGLER, A. L.; LUSETTI, S. L.; COX, M. M. The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. **The Journal of biological chemistry**, v. 278, n. 18, p. 16389–96, 2003.

EGNER, C. et al. Effects of various single-stranded-DNA-binding proteins on reactions promoted by RecA protein. **Journal of bacteriology**, v. 169, n. 8, p. 3422–8, 1987.

ELLOUZE, C. et al. Evidence for Elongation of the Helical Pitch of the RecA Filament Upon ATP and ADP Binding Using Small-Angle Neutron Scattering. **European Journal of Biochemistry**, v. 233, n. 2, p. 579–583, 1995.

EMSLEY, P. et al. Features and development of Coot. **Acta crystallographica. Section D, Biological crystallography**, International Union of Crystallography, v. 66, n. 4, p. 486–501, 2010.

- EVANS, P. R.; MURSHUDOV, G. N. How good are my data and what is the resolution? **Acta crystallographica. Section D, Biological crystallography**, v. 69, n. Pt 7, p. 1204–14, 2013.
- FRIEDBERG, E. C.; WALKER, G. C.; SIEDE, W. **DNA Repair and Mutagenesis**. [S.l.]: ASM Press, 1995. 698 p.
- FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins**, v. 23, n. 4, p. 566–79, 1995.
- FU, H. et al. Dynamics and Regulation of RecA Polymerization and De-Polymerization on Double-Stranded DNA. **PLoS ONE**, Public Library of Science, v. 8, n. 6, p. e66712, 2013.
- GALLETTO, R. et al. Direct observation of individual RecA filaments assembling on single DNA molecules. **Nature**, v. 443, n. 7113, p. 875–8, 2006.
- GALVÃO, C. W. et al. The RecX protein interacts with the RecA protein and modulates its activity in *Herbaspirillum seropedicae*. **Brazilian journal of medical and biological research**, v. 45, n. 12, p. 1127–34, 2012.
- GIACOVAZZO, C. **Fundamentals of crystallography; 3rd ed.** Oxford: Oxford University Press, 2011. (IUCr texts on crystallography).
- GRUBER, A. J. et al. A RecA protein surface required for activation of DNA polymerase V. **PLoS genetics**, v. 11, n. 3, p. 1–37, 2015.
- HENLE, E. S.; LINN, S. Formation, Prevention, and Repair of DNA Damage by Iron/Hydrogen Peroxide. **Journal of Biological Chemistry**, v. 272, n. 31, p. 19095–19098, 1997.
- HEWAT, E. A.; RUIGROK, R. W.; DICAPUA, E. Activation of RecA protein: the pitch of the helical complex with single-stranded DNA. **The EMBO journal**, v. 10, n. 9, p. 2695–8, 1991.
- JANCARIK, J.; KIM, S.-H.; IUCR. Sparse matrix sampling: a screening method for crystallization of proteins. **urn:issn:0021-8898**, International Union of Crystallography, v. 24, n. 4, p. 409–411, 1991.
- JANION, C. Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. **International journal of biological sciences**, v. 4, n. 6, p. 338–44, 2008.
- KARPLUS, P. A.; DIEDERICHS, K. Linking Crystallographic Model and Data Quality. **Science**, v. 336, n. 6084, p. 1030–1033, 2012.
- KELLEY, J. A.; KNIGHT, K. L. Allosteric Regulation of RecA Protein Function Is Mediated by Gln194. **Journal of Biological Chemistry**, v. 272, n. 41, p. 25778–25782, 1997.
- KENDREW, J. C. et al. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. **Nature**, v. 181, n. 4610, p. 662–666, 1958.

- KIM, J.-I. et al. RecA Protein from the extremely radioresistant bacterium *Deinococcus radiodurans*: expression, purification, and characterization. **Journal of bacteriology**, v. 184, n. 6, p. 1649–60, 2002.
- KIM, T. et al. Directed Evolution of RecA Variants with Enhanced Capacity for Conjugational Recombination. **PLoS genetics**, v. 11, n. 6, p. e1005278, 2015.
- KINEBUCHI, T. et al. Structural Basis for Octameric Ring Formation and DNA Interaction of the Human Homologous-Pairing Protein Dmc1. **Molecular Cell**, v. 14, n. 3, p. 363–374, 2004.
- KOONIN, E. V.; GALPERIN, M. Y. **Sequence - Evolution - Function: Computational Approaches in Comparative Genomics**. Boston: Kluwer Academic, 2003. Chapter 2 p.
- KOWALCZYKOWSKI, S. C. et al. Effects of the *Escherichia coli* SSB protein on the binding of *Escherichia coli* RecA protein to single-stranded DNA. Demonstration of competitive binding and the lack of a specific protein-protein interaction. **Journal of molecular biology**, v. 193, n. 1, p. 81–95, 1987.
- KOWALCZYKOWSKI, S. C.; EGGLESTON, a. K. Homologous pairing and DNA strand-exchange proteins. **Annual review of biochemistry**, v. 63, p. 991–1043, 1994.
- KUZMINOV, A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. **Microbiology and molecular biology reviews : MMBR**, v. 63, n. 4, p. 751–813, table of contents, 1999.
- LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, aug 1970.
- LEE, C. D.; WANG, T. F. The N-terminal domain of *Escherichia coli* RecA have multiple functions in promoting homologous recombination. **Journal of biomedical science**, v. 16, n. 1, p. 37, 2009.
- LEE, J. W.; COX, M. M. Inhibition of recA protein promoted ATP hydrolysis. 1. ATP gamma S and ADP are antagonistic inhibitors. **Biochemistry**, v. 29, n. 33, p. 7666–76, 1990. ISSN 0006-2960.
- LEHNINGER, A.; NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. Fifth edition. [S.l.]: W. H. Freeman, 2008. Hardcover.
- LIAO, J.-C. Mechanical transduction mechanisms of RecA-like molecular motors. **Journal of biomolecular structure & dynamics**, v. 29, n. 3, p. 497–507, 2011.
- LINDAHL, T.; NYBERG, B. Rate of depurination of native deoxyribonucleic acid. **Biochemistry**, American Chemical Society, v. 11, n. 19, p. 3610–3618, 1972.
- LINDSLEYS, E. Assembly and Disassembly of RecA Protein Filaments Opposite Filament Ends Occur at. v. 265, p. 9043–9054, 1990.
- LUSETTI, S. L.; COX, M. M. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. **Annual review of biochemistry**, v. 71, p. 71–100, 2002.

LUSETTI, S. L.; SHAW, J. J.; COX, M. M. Magnesium ion dependent activation of the RecA protein involves the C terminus. **The Journal of biological chemistry**, v. 278, n. 18, p. 16381–8, 2003.

MACFARLAND, K. J. et al. RecA as a motor protein. Testing models for the role of ATP hydrolysis in DNA strand exchange. **The Journal of biological chemistry**, v. 272, n. 28, p. 17675–85, 1997.

MADIRAJU, M. V.; TEMPLIN, A.; CLARK, A. J. Properties of a mutant recA-encoded protein reveal a possible role for *Escherichia coli* recF-encoded protein in genetic recombination. **Proceedings of the National Academy of Sciences of the United States of America**, v. 85, n. 18, p. 6592–6, 1988.

MARCEAU, A. H. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. **Methods in molecular biology (Clifton, N.J.)**, v. 922, p. 1–21, 2012.

MATTHEWS, B. W. Solvent content of protein crystals. **Journal of molecular biology**, v. 33, n. 2, p. 491–7, apr 1968.

MCCOY, A. J. et al. Phaser crystallographic software. **Journal of applied crystallography**, International Union of Crystallography, v. 40, n. 4, p. 658–674, 2007.

MORRICAL, S. W.; LEE, J.; COX, M. M. Continuous association of *Escherichia coli* single-stranded DNA binding protein with stable complexes of recA protein and single-stranded DNA. **Biochemistry**, v. 25, n. 7, p. 1482–94, 1986.

NGO, K. V. et al. Regulation of *Deinococcus radiodurans* RecA protein function via modulation of active and inactive nucleoprotein filament states. **The Journal of biological chemistry**, American Society for Biochemistry and Molecular Biology, v. 288, n. 29, p. 21351–66, 2013.

NOTREDAME, C.; HIGGINS, D. G.; HERINGA, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. **Journal of molecular biology**, v. 302, n. 1, p. 205–17, 2000.

OKOROKOV, A. L. et al. Structure of the hDmc1-ssDNA filament reveals the principles of its architecture. **PloS one**, v. 5, n. 1, p. e8586, 2010.

PATTERSON, A. L. A fourier series method for the determination of the components of interatomic distances in crystals. **Phys. Rev.**, American Physical Society, v. 46, p. 372–376, 1934.

PUGH, B. F.; COX, M. M. General mechanism for RecA protein binding to duplex DNA. **Journal of molecular biology**, v. 203, n. 2, p. 479–93, 1988.

PUGH, B. F.; COX, M. M. High salt activation of recA protein ATPase in the absence of DNA. **The Journal of biological chemistry**, v. 263, n. 1, p. 76–83, 1988.

RADDING, C. M. et al. Three phases in homologous pairing: polymerization of recA protein on single-stranded DNA, synapsis, and polar strand exchange. **Cold Spring Harbor symposia on quantitative biology**, v. 47 Pt 2, p. 821–8, 1983.

RAGHUNATHAN, S. et al. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. **Nature Structural Biology**, v. 7, n. 8, p. 648–652, 2000.

RAJAN, R.; BELL, C. E. Crystal structure of RecA from *Deinococcus radiodurans*: insights into the structural basis of extreme radioresistance. **Journal of molecular biology**, v. 344, n. 4, p. 951–63, 2004.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of molecular biology**, v. 7, p. 95–9, jul 1963.

REGISTER, J. C.; GRIFFITH, J. The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. **Journal of Biological Chemistry**, v. 260, n. 22, p. 12308–12312, 1985.

RIPOLL, M. M.; CANO, F. H. **Crystallography**. 2016. Disponível em: <<http://www.xtal.iqfr.csic.es/>>.

ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server. **Nucleic acids research**, v. 42, n. Web Server issue, p. W320–4, 2014.

ROBU, M. E.; INMAN, R. B.; COX, M. M. RecA protein promotes the regression of stalled replication forks in vitro. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 15, p. 8211–8, 2001.

ROCA, A. I.; COX, M. M. RecA protein: structure, function, and role in recombinational DNA repair. **Progress in nucleic acid research and molecular biology**, v. 56, p. 129–223, 1997.

ROSSMANN, M. G.; BLOW, D. M. The detection of sub-units within the crystallographic asymmetric unit. **Acta Crystallographica**, International Union of Crystallography, v. 15, n. 1, p. 24–31, 1962.

ROY, R. et al. SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. **Nature**, Macmillan Publishers Limited. All rights reserved, v. 461, n. 7267, p. 1092–7, 2009.

SAMBROOK, J.; RUSSELL, D. W. **Molecular Cloning: A Laboratory Manual**. 3. ed. Cold Spring Harbor, NY: CSHL Press, 2001.

SANDLER, S. J. et al. RecA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae*. **Nucleic Acids Research**, v. 24, n. 11, p. 2125–2132, 1996.

SCHRÖDINGER, L. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

SHAN, Q.; COX, M. M. RecA filament dynamics during DNA strand exchange reactions. **The Journal of biological chemistry**, v. 272, n. 17, p. 11063–73, 1997.

SHAN, Q.; COX, M. M.; INMAN, R. B. DNA Strand Exchange Promoted by RecA K72R: Two Reaction Phases With Different Mg Requirements. **Journal of Biological Chemistry**, v. 271, n. 10, p. 5712–5724, 1996.

SHATSKY, M.; NUSSINOV, R.; WOLFSON, H. J. A method for simultaneous alignment of multiple protein structures. **Proteins**, v. 56, n. 1, p. 143–56, 2004.

SHEINERMAN, F. B.; NOREL, R.; HONIG, B. Electrostatic aspects of protein-protein interactions. **Current Opinion in Structural Biology**, v. 10, n. 2, p. 153–159, 2000.

SHEREDA, R. D. et al. SSB as an organizer mobilizer of genome maintenance complex. **Critical reviews in biochemistry and molecular biology**, v. 43, n. 5, p. 289–318, 2008.

SHERRATT, M. J. et al. ECM macromolecules: rotary shadowing and transmission electron microscopy. **Methods in molecular biology (Clifton, N.J.)**, v. 522, p. 175–81, 2009.

SHIVASHANKAR, G. V. et al. RecA polymerization on double-stranded DNA by using single-molecule manipulation: the role of ATP hydrolysis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 14, p. 7916–21, 1999.

SIGURDSSON, S. et al. Basis for avid homologous DNA strand exchange by human Rad51 and RPA. **The Journal of biological chemistry**, v. 276, n. 12, p. 8798–806, 2001.

STASIAK, A.; CAPUA, E. D. The helicity of DNA in complexes with RecA protein. **Nature**, v. 299, n. 5879, p. 185–186, 1982.

STASIAK, A.; EGELMAN, E. H. Structure and Dynamics of recA Protein-DNA Complexes as Determined by Image Analysis of Electron Micrographs. **Biophysical journal**, v. 49, n. 1, p. 5–7, 1986.

STASIAK, A.; EGELMAN, E. H.; HOWARD-FLANDERS, P. Structure of helical RecA-DNA complexes. III. The structural polarity of RecA filaments and functional polarity in the RecA-mediated strand exchange reaction. **Journal of molecular biology**, v. 202, n. 3, p. 659–62, 1988.

STEFFEN, S. E.; BRYANT, F. R. Purification and characterization of the RecA protein from *Streptococcus pneumoniae*. **Archives of biochemistry and biophysics**, v. 382, n. 2, p. 303–9, 2000.

STOHL, E. A. et al. Purification and characterization of the RecA protein from *Neisseria gonorrhoeae*. **PloS one**, Public Library of Science, v. 6, n. 2, p. e17101, 2011.

STORY, R. M.; STEITZ, T. A. Structure of the recA protein-ADP complex. **Nature**, v. 355, n. 6358, p. 374–6, 1992.

STORY, R. M.; WEBER, I. T.; STEITZ, T. A. The structure of the *E. coli* recA protein monomer and polymer. **Nature**, v. 355, n. 6358, p. 318–325, 1992.

TANG, M. et al. UmuD'(2)C is an error-prone DNA polymerase, *Escherichia coli* pol V. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 16, p. 8919–24, 1999.

- TERWILLIGER, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. **Acta crystallographica. Section D, Biological crystallography**, International Union of Crystallography, v. 64, n. Pt 1, p. 61–9, 2008.
- THOMAS, A.; LLOYD, R. G. Control of recA dependent activities in *Escherichia coli*: a possible role for the recF product. **Journal of general microbiology**, v. 129, n. 3, p. 681–6, 1983.
- TIMMINS, P. A.; RUIGROK, R. W.; CAPUA, E. D. The solution structure of recA filaments by small angle neutron scattering. **Biochimie**, v. 73, n. 2-3, p. 227–30, 1991.
- TORNALETTI, S.; HANAWALT, P. C. Effect of DNA lesions on transcription elongation. **Biochimie**, v. 81, n. 1-2, p. 139–46, 1999.
- UMEZU, K.; KOLODNER, R. D. Protein interactions in genetic recombination in *Escherichia coli*. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. **The Journal of biological chemistry**, v. 269, n. 47, p. 30005–13, 1994.
- VANLOOCK, M. S. et al. ATP-Mediated Conformational Changes in the RecA Filament. **Structure**, v. 11, n. 2, p. 187–196, 2003.
- VOLKERT, M. R.; MARGOSSIAN, L. J.; CLARK, A. J. Two-component suppression of recF143 by recA441 in *Escherichia coli* K-12. **Journal of bacteriology**, v. 160, n. 2, p. 702–5, 1984.
- WANG, T. C.; CHANG, H. Y.; HUNG, J. L. Cosuppression of recF, recR and recO mutations by mutant recA alleles in *Escherichia coli* cells. **Mutation research**, v. 294, n. 2, p. 157–66, 1993.
- WANG, T. C.; SMITH, K. C. recA (Srf) suppression of recF deficiency in the postreplication repair of UV-irradiated *Escherichia coli* K-12. **Journal of bacteriology**, v. 168, n. 2, p. 940–6, 1986.
- WANG, T.-F.; CHEN, L.-T.; WANG, A. H.-J. Right or left turn? RecA family protein filaments promote homologous recombination through clockwise axial rotation. **BioEssays : news and reviews in molecular, cellular and developmental biology**, v. 30, n. 1, p. 48–56, 2008.
- WEINSTOCK, G.; MCENTEE, K.; LEHMAN, I. Hydrolysis of nucleoside triphosphates catalyzed by the recA protein of *Escherichia coli*. Hydrolysis of UTP. **J. Biol. Chem.**, v. 256, n. 16, p. 8856–8858, 1981.
- WU, Y. et al. Crystal structure of archaeal recombinase RADA: a snapshot of its extended conformation. **Molecular cell**, v. 15, n. 3, p. 423–35, 2004.
- XING, X.; BELL, C. E. Crystal Structures of *Escherichia coli* RecA in Complex with MgADP and MnAMP-PNP. **Biochemistry**, v. 43, n. 51, p. 16142–16152, 2004.
- YU, X.; EGELMAN, E. H. The RecA hexamer is a structural homologue of ring helicases. **Nature Structural Biology**, v. 4, n. 2, p. 101–104, 1997.

YU, X. et al. What is the structure of the RecA-DNA filament? **Current protein & peptide science**, v. 5, n. 2, p. 73–9, 2004.