

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

IVAN ROSA E SILVA

Estudos biofísicos da Selenofosfato Sintetase de *Escherichia coli* e  
investigação de seu papel na via de biossíntese de Selenocisteínas

São Carlos

2012







IVAN ROSA E SILVA

Estudos biofísicos da Selenofosfato Sintetase de *Escherichia coli* e  
investigação de seu papel na via de biossíntese de Selenocisteínas

Dissertação apresentada ao Programa de Pós-  
Graduação em Física do Instituto de Física de  
São Carlos da Universidade de São Paulo para  
obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada  
Opção: Física Biomolecular

Orientador: Prof. Dr. Otavio Henrique  
Thiemann

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC, com os dados fornecidos pelo autor.

Rosa e Silva, Ivan

Estudos biofísicos da Selenofosfato Sintetase de Escherichia coli e investigação de seu papel na via de biossíntese de Selenocisteínas / Ivan Rosa e Silva; orientador Prof. Dr. Otavio Henrique Thiemann - versão corrigida - São Carlos, 2012.

140 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Biomolecular) - Instituto de Física de São Carlos, Universidade de São Paulo, 2011.

1. Selenofosfato Sintetase. 2. SELD. 3. Selenocisteína.  
4. Escherichia coli. I. Thiemann, Otavio Henrique, orient. II. Título.







À minha família e amigos, pelo apoio ao longo  
do período de elaboração deste trabalho.



## AGRADECIMENTOS

Ao Prof. Dr. Otavio Henrique Thiemann, pela confiança e orientação na realização deste trabalho, pelos ensinamentos e discussões construtivas.

Aos meus pais Benedito e Amarilda e à minha irmã Iara, pelo apoio em todos os momentos, em nome de quem agradeço a todos da minha família.

À Mamé e à Susana, por terem me guiado nos meus primeiros passos dentro do Laboratório de Cristalografia. À Lívia Faim, pela confiança, pelas discussões sempre produtivas, pela grande contribuição neste trabalho e pelo companheirismo e amizade. Ao Wally, ao Geraldo, e ao B2, pela amizade e contribuição neste trabalho.

Aos membros do grupo de pesquisa do Prof. Dr. Otavio H. Thiemann: Alexandre, Daiana, Fer, Gi, Gustavo, Iza, Jaque, Ju, Lívia Faim, Lívia Manzine, Mamé, Marcos Michel, Rapha, Teresa, Túlio e Udi, em nome de quem agradeço a todos os amigos do laboratório.

Aos técnicos dos laboratórios do Grupo de Cristalografia e do Grupo de Biofísica, pelo apoio nos experimentos.

Aos amigos da BIOMOL: Atílio, B2, Caio, Geraldo, Hilde, Luciano, Paulinha, Ricardo, Rodrigo, Udi e Wally, pela amizade e pelas discussões científicas ou não, pela união nos momentos difíceis e de descontração, em nome de quem agradeço a todos os que me acompanharam na minha formação.

Aos professores do IFSC-USP, pelos ensinamentos e pelas discussões construtivas. Ao Dr. Mario de Oliveira Neto do IFSC-USP, ao professor Dr. Júlio Cesar Borges do IQSC-USP, ao Dr. Daniel Menezes Saidemberg e ao professor Dr. Mario Sergio Palma, ambos do Instituto de Biociências da UNESP-Rio Claro, pela colaboração com este trabalho.

Às funcionárias do serviço de Biblioteca e Informação do IFSC-USP, pela prestatividade em todos os momentos.

Ao Grupo de Biologia Molecular Estrutural e Cristalografia de Proteínas, pela oportunidade de realização deste trabalho.

Ao Instituto de Física de São Carlos e à Universidade de São Paulo, pela oportunidade de realização do curso de mestrado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e à CAPES, pela concessão de bolsa de mestrado e pelo apoio financeiro para a realização desta pesquisa.

Ao LNLS, LNBio e LNNano, por colocar à disposição grande parte dos equipamentos para a realização dos experimentos.

Enfim, a todos que contribuíram de forma direta ou indireta para a realização deste trabalho.

“The important thing is not to stop questioning”.

**Albert Einstein**



## RESUMO

ROSA E SILVA, I. **Estudos biofísicos da Selenofosfato Sintetase de *Escherichia coli* e investigação de seu papel na via de biossíntese de Selenocisteínas**. 2011. 140 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

A principal forma biológica do selênio em vários organismos é o aminoácido Selenocisteína (Sec, U), que é incorporado em um polipeptídeo emergente em códons UGA específicos. Em *Escherichia coli*, esta incorporação requer os genes que codificam para Seril-tRNA Sintetase (SerRS), Selenocisteína Sintase (SELA), um tRNA<sup>Sec</sup> específico (SELC), Selenofosfato Sintetase (SELD) e um fator de alongação de transcrição específico (SELB). A proteína Selenofosfato Sintetase (EC 2.7.9.3) pertence à família AIRS, de proteínas que têm o ATP como substrato, e produz o composto biologicamente ativo doador de selênio, o monoselenofosfato, a partir de ATP e seleneto. O gene *selD* em *E. coli* tem 1041 pares de bases e codifica uma proteína com 347 aminoácidos e massa molecular de 37 kDa. A fase aberta de leitura do gene *selD* foi amplificada do DNA genômico de *E. coli* e clonada em vetor de expressão pet28a(+) (Novagen). A proteína recombinante foi superexpressa em *E. coli* por indução com IPTG e purificada por cromatografia de afinidade por ligação a metal e a fração eluída foi concentrada por ultrafiltração. Em seguida, o produto foi submetido à clivagem da cauda de histidinas com Trombina. Para purificar o produto de reação de clivagem com protease e para estimar sua massa molecular e estado oligomérico, empregou-se cromatografia de exclusão molecular. A proteína pura foi utilizada em experimentos de Gel Nativo e em estudos das suas propriedades hidrodinâmicas realizados por meio de Espalhamento Dinâmico de Luz (DLS), Espalhamento de Raios-X a Baixo Ângulo (SAXS) e Ultracentrifugação Analítica (AUC). Os resultados obtidos revelam uma mistura de oligômeros em solução, em um equilíbrio dímero-tetrâmero e tetrâmero-octâmero. Um modelo tridimensional para o homodímero de SELD de *E. coli* foi obtido por Modelagem Molecular e suas propriedades hidrodinâmicas previstas concordam com aquelas obtidas experimentalmente. Adicionalmente, triagens de condições de cristalização da proteína revelaram condições em que a proteína cristaliza na forma de pequenas agulhas e ensaios de otimização por variação da concentração de agente precipitante e pH não resultaram em monocristais adequados para difração de raios-X. A análise do papel da SELD na via de

biossíntese de Selenocisteínas levanta a hipótese de que esta proteína deve entregar o monoselenofosfato para o complexo SELA-SELC de modo que o selênio seja incorporado para formação do aminoácido Selenocisteína, já que os compostos de selênio são tóxicos quando estão livres na célula. Portanto, a investigação da interação da SELD com o complexo SELA-SELC foi observada pelo monitoramento da anisotropia de fluorescência do complexo SELA-SELC mediante titulação de SELD. A análise local da interação para manutenção do complexo SELD-SELA-SEC foi feita por meio de espectrometria de massas com troca H/D, que revelou possíveis sítios de interação na superfície da SELD. Os resultados mostrados neste trabalho ampliam o conhecimento sobre a via de biossíntese de Selenocisteína, revelando detalhes da interação da SELD com o complexo SELA-SELC.

Palavras-chave: Selenofosfato Sintetase. SELD. Selenocisteína. *Escherichia coli*.



## ABSTRACT

ROSA E SILVA, I. **Biophysical studies of *Escherichia coli* Selenophosphate Synthetase and investigation of its role in the Selenocysteine biosynthesis pathway.** 2011. 140 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

The main biological form of selenium in several organisms is the amino acid Selenocysteine (Sec, U), which is incorporated into selenoproteins in specific UGA codons. In *Escherichia coli*, it requires the genes that codify to Seryl-tRNA Synthetase (SerRS), Selenocysteine Synthase (SELA), a specific tRNA<sup>Sec</sup> (SELC), Selenophosphate Synthetase (SELD) and a specific translation elongation factor (SELB). Selenophosphate Synthetase (EC 2.7.9.3) belongs to AIRS superfamily of proteins that have ATP as a substrate and this protein produces the biologically active selenium donor compound, monoselenophosphate, from ATP and selenide. The *selD* gene from *E. coli* is 1041 base pairs long and codifies a protein with 347 amino acids and molecular mass of 37 kDa. The open reading frame of *selD* gene was amplified from *E. coli* genomic DNA and cloned into pET28a(+) expression vector (Novagen). The recombinant protein was overexpressed in *E. coli* by IPTG induction and purified by metal affinity chromatography, and the eluted fraction was concentrated by ultrafiltration. The product was used for Thrombin protease cleavage of the 6-His tag. In order to purify the product of proteolysis and to estimate its molecular mass and oligomeric state, we used size exclusion chromatography. The pure protein sample was used for Native Gel Electrophoresis. Hydrodynamic properties of the protein were studied by Dynamic Light Scattering (DLS), Small angle X-ray scattering (SAXS) and Analytical Ultracentrifugation (AUC). The results show an equilibrium between SELD oligomeric forms, as dimer-tetramer and tetramer-octamer association in solution. A tridimensional model of *E. coli* SELD was obtained by Molecular Modelling and its predicted hydrodynamic properties agree with those observed experimentally. In addition, crystal screening revealed crystallization conditions suitable for protein crystallization as small needles, but optimization of these conditions by precipitant agent and pH variation did not result in monocrystals reliable for X-ray diffraction. An analysis of SELD's role in the Selenocysteine biosynthesis pathway indicates that SELD must deliver monoselenophosphate to the SELA-SELC complex so that the selenium is incorporated to the amino acid to form selenocysteyl-SEC, since selenium compounds are

toxic when they are freely available in the cell. This interaction was observed by fluorescence anisotropy. The local analysis of complex formation was monitored by mass spectrometry after H/D exchange and revealed possible sites for this interaction on SELD surface. The results improve our knowledge about the Selenocysteine pathway in the cell, showing details of the interaction between SELD and the SELA-SELC complex.

Keywords: Selenophosphate Synthetase. SELD. Selenocysteine. *Escherichia coli*.

## LISTA DE FIGURAS

- FIGURA 1 - Número de publicações científicas envolvendo o estudo de selenoproteínas (2000 – 2010). Adaptado <sup>24</sup>. ..... 30
- FIGURA 2 - Representações 2D e 3D dos aminoácidos A1-A2) L-cisteína e B1-B2) L-selenocisteína. Em vermelho é representado o oxigênio, em azul o nitrogênio, em amarelo o enxofre e em laranja o selênio. As figuras foram preparadas usando os programas Marvin Sketch <sup>29</sup> e UCSF Chimera <sup>30</sup>. ..... 31
- FIGURA 3 - Estrutura secundária encontrada em A) tRNAs em geral, B) tRNA<sup>Sec</sup> de eubactérias e C) tRNA<sup>Sec</sup> de arqueias e eucariotos. Adaptado <sup>38, 39, 40</sup>. ..... 33
- FIGURA 4 - Estrutura consenso do elemento SECIS em A) eubactérias, B) arqueias e C) eucariotos. Em eucariotos são observados dois tipos de elemento SECIS: I e II. O elemento SECIS tipo II de eucariotos é mostrado na caixa. Adaptado <sup>42, 34</sup>. ..... 33
- FIGURA 5 - A) Organização do *operon selAB* em *E. coli*. B) Estrutura do elemento similar ao SECIS na região 5' não traduzida do operon *selAB*. Na figura, destaca-se a região de ligação a ribossomo e o asterisco indica o início da região codificadora do gene *sela*. Adaptado <sup>40</sup>. ..... 34
- FIGURA 6 - Representação esquemática comparativa das vias de biossíntese e incorporação de selenocisteína em selenoproteínas de A) eubactérias, B) arqueias e C) eucariotos. No caso de eucariotos, representam-se apenas as etapas que ocorrem no citoplasma celular. O ribossomo foi omitido nas figuras. O ponto de interrogação em B) indica que ainda não está determinada a ligação do elemento SECIS de arqueias no complexo selenocisteil-tRNA<sup>Sec</sup>-EFSec. Adaptado <sup>49</sup>. ..... 36
- FIGURA 7 - Reação catalisada pelas Selenofosfato Sintetases. .... 38
- FIGURA 8 - Primeira (A) e segunda (B) etapas da reação catalisada pelas Selenofosfato Sintetases. .... 40
- FIGURA 9 - Representação esquemática da condição da reação de amplificação do gene *selD* de *E. coli*. .... 50
- FIGURA 10 - Vetor de expressão pET28a(+). A cauda N-terminal de 6xHis está marcada em vermelho. As setas indicam as regiões reconhecidas pelas

enzimas de restrição <i>NdeI</i> e <i>XhoI</i> , entre as quais foi inserida a ORF <i>selD</i> de <i>E. coli</i> . .....	51
FIGURA 11 - Purificação por afinidade a metal imobilizado. Adaptado do manual da coluna <i>TALON™ Metal Affinity Resin (BD Biosciences)</i> . .....	54
FIGURA 12 - Decomposição do feixe de luz plano polarizado em duas componentes circulares polarizadas à esquerda (L) e à direita (R), a) antes e b) depois de passar por amostra opticamente ativa. Adaptado <sup>83</sup> . .....	57
FIGURA 13 - Espectro de CD no UV distante. ( ____ ) $\alpha$ -hélice; ( -- -- -- ) folha- $\beta$ antiparalela; ( . . . . ) volta- $\beta$ , ( - - - - ) estrutura irregular. Adaptado <sup>83</sup> . .....	58
FIGURA 14 - Experimentos de Ultracentrifugação Analítica. A) SV e B) SE. Adaptado <sup>96</sup> . .....	67
FIGURA 15 - Método de Difusão de Vapor: A) gota pendurada e B) Gota apoiada. ....	68
FIGURA 16 - Representação dos domínios da SELD de <i>E. coli</i> . .....	74
FIGURA 17 - Alinhamento múltiplo das sequências de aminoácidos das Selenofosfatos Sintetases. Em rosa estão assinaladas a Cys (17) catalítica e a Cys (19) não catalítica, além da Lys (20) catalítica. Em cinza estão assinaldos os aminoácidos que participam da ligação de ATP e, em azul, os aminoácidos que fazem parte da interface de dimerização. Alguns aminoácidos acumulam as funções de ligação de ATP e formação da interface de dimerização. São mostradas 5 sequências representativas de Selenofosfato Sintetases: EC, <i>E. coli</i> ; AA, <i>Aquifex aeolicus</i> ; CP, <i>Clostridium perfringens</i> ; MJ, <i>Methanocaldococcus jannaschii</i> ; HS1, SPS1 de <i>Homo sapiens</i> . .....	75
FIGURA 18 - Gráfico de Ramachandran para o modelo de <i>Threading</i> da SELD de <i>E. coli</i> . .....	76
FIGURA 19 - A- Avaliação do <i>Z-score</i> e B- avaliação da energia por resíduo de aminoácido da estrutura de <i>Threading</i> da SELD de <i>E. coli</i> . .....	77
FIGURA 20 - Alinhamento global da sequência de aminoácidos da SELD de <i>E. coli</i> com as cadeias A e B, respectivamente, do PDB 3FD5. Em vermelho estão coloridas as regiões de estrutura em hélice- $\alpha$ e em azul, as regiões de fitas- $\beta$ . A estrutura secundária representada para a SELD de <i>E. coli</i> foi predita utilizando o servidor <i>PSIPRED</i> <sup>68</sup> . .....	79
FIGURA 21 - Avaliação de DOPE para os modelos gerados por Homologia. ....	80

FIGURA 22 - Gráfico de Ramachandran para os modelos otimizados, obtidos das cadeias A e B, respectivamente. ....	81
FIGURA 23 - Avaliação do <i>Z-score</i> obtido para os modelos de Modelagem por Homologia da SELD de <i>E. coli</i> (A1 – cadeia A; A2 – cadeia B) e avaliação da energia por resíduo de aminoácido da estrutura dos modelos (B1 – cadeia A e B2 – cadeia B). ....	82
FIGURA 24 - Alinhamento tridimensional dos modelos do monômero de SELD de <i>E. coli</i> gerados por meio de Modelagem por Homologia (azul – PDB 3FD5-B) e <i>Threading</i> (vermelho). ....	83
FIGURA 25 - A - Modelo da estrutura dimérica da SELD de <i>E. coli</i> obtido por meio do servidor <i>RosettaDock</i> . B – Superfície acessível à solvatação do modelo do dímero de SELD de <i>E. coli</i> (em azul claro e azul escuro) e interface de interação entre monômeros (em verde e vermelho), obtida a partir de análise do modelo no servidor <i>PISA</i> <sup>77</sup> . As dimensões do modelo da estrutura dimérica foram obtidas através do programa <i>UCSF Chimera</i> <sup>30</sup> ....	84
FIGURA 26 - A - Avaliação do <i>Z-score</i> obtido para o modelo do dímero de SELD de <i>E. coli</i> e B - Avaliação da energia por resíduo de aminoácido da estrutura do modelo. ....	85
FIGURA 27 - Amplificação da ORF do gene <i>selD</i> de <i>E. coli</i> : Resultado analisado em gel de agarose 1%, TAE 1X. M: marcador de massa molecular <i>Gene Ruler 1Kb Ladder Plus</i> (Fermentas Life Sciences). Os números 1 e 2 representam o fragmento amplificado por PCR, referente ao gene <i>selD</i> <i>E. coli</i> , indicado pela seta, sendo que a amostra 2 apresenta mais DNA amplificado (dada a intensidade da banda indicada pela seta). ....	87
FIGURA 28 - Verificação de clonagem da ORF do gene <i>selD</i> de <i>E. coli</i> em vetor de expressão pET28a(+). Resultado analisado em gel de agarose 1% TAE 1X. Representa a digestão da construção <i>selD E. coli</i> /pET28a(+) com as enzimas <i>Nde</i> I e <i>Xho</i> I. A seta indica uma banda entre 1000pb e 1500pb, correspondente ao gene <i>selD</i> <i>E. coli</i> . M: marcador de massa molecular de 1Kb plus DNA Ladder (Fermentas Life Sciences). 1: Controle negativo. 2: Amostra resultante da reação de digestão. ....	87
FIGURA 29 - Alinhamento das extremidades 5´ e 3´ do gene <i>selD</i> <i>E.coli</i> . Os alinhamentos indicados pelas setas no sentido esquerda para direita (SEQ. 1 e SEQ. 2) representam o alinhamento da extremidade 5´ e os	

alinhamentos indicados pelas setas no sentido direita para esquerda (SEQ. 3 e SEQ. 4) representam o alinhamento 3'. SELD <i>E. coli</i> representa a sequência original. ....	88
FIGURA 30 - Expressão e purificação por cromatografia de afinidade. M: padrão de massa molecular; 1: antes e 2: depois de indução com IPTG; 3: pellet e 4: sobrenadante após lise; 5: eluato; eluição com 6: 10 mM, 7: 20mM, 8: 200 mM e 9: 500 mM Imidazol. ....	89
FIGURA 31 - Clivagem com Trombina da proteína SELD de <i>E. coli</i> . Analisado em SDS-PAGE 15%. M: padrão de massa molecular; amostras 1: não submetida e 2: submetida à reação de clivagem. A seta indica a a banda correspondente à amostra clivada. ....	89
FIGURA 32 - Resultado da purificação da proteína SELD <i>E. coli</i> através de cromatografia de exclusão molecular. A - Perfil de eluição de SELD <i>E.coli</i> . B - SDS-PAGE 15% dos picos P1 ( $V_e = 13,50$ ml) e P2 ( $V_e = 15,06$ ml). ....	90
FIGURA 33 - Espectro de Dicroísmo Circular da SELD de <i>E. coli</i> . ....	91
FIGURA 34 - Curvas de calibração da SUPERDEX 200 <sup>TM</sup> 10/300 para estimativa de A – massa molecular e B – raio hidrodinâmico. ....	93
FIGURA 35 - Eletroforese em gel nativo. M: amostra com proteínas-padrão. 1: SDEC <i>E. coli</i> (pico 2da cromatografia de exclusão molecular). A seta preta indica a forma tetramérica e a seta vermelha indica a forma dimérica da SELD de <i>E. coli</i> . ....	94
FIGURA 36 - A - Curva de Calibração para Massas Moleculares a partir de Eletroforese em Gel Nativo e B - Curva de Calibração para Raio Hidrodinâmico a partir de Eletroforese em Gel Nativo. ....	95
FIGURA 37 - Comportamento do coeficiente de difusão translacional com a concentração da amostra de proteína, medido por DLS. ....	96
FIGURA 38 - Curva de espalhamento de raios a baixo ângulo para a SelD de <i>E. coli</i> e Gráfico de Guinier inserido. ....	98
FIGURA 39 - Distribuição de distâncias, $p(r)$ . ....	99
FIGURA 40 - Comparação das curvas de espalhamento de raios-X a baixo ângulo experimental e calculada para o PDB 3FD5. ....	100

FIGURA 41 - Comparação das curvas de espalhamento de raios-X a baixo ângulo experimental e calculada para o modelo da SELD de <i>E. coli</i> obtido através de modelagem molecular. ....	100
FIGURA 42 - Modelo <i>Ab Initio</i> para o envelope molecular da proteína SELD de <i>E. coli</i> alinhado com o modelo tridimensional obtido por Modelagem Molecular por meio do programa SUPCOMB <sup>92</sup> . ....	101
FIGURA 43 - Distribuição de coeficientes de sedimentação, c(s). É mostrado um <i>zoom</i> na região de 6 S a 13 S (inserido). ....	102
FIGURA 44 - Regressão linear dos coeficientes de sedimentação para as 3 espécies identificadas na distribuição de coeficientes de sedimentação (FIG. 43). ....	103
FIGURA 45 - Sedimentação em equilíbrio (SE). São mostradas as curvas de Absorbância da amostra contra o raio da célula e o comportamento dos residuais do modelo utilizado no programa SEDPHAT <sup>98</sup> para ajuste dos dados a um sistema de equilíbrio entre dímero-tetrâmero e tetrâmero-octâmero. ....	105
FIGURA 46 - Pequenos cristais na forma de agulhas obtidos em condição de cristalização com 0.2M AmSO <sub>4</sub> , 0,1M HEPES pH 7,5, 16% (w/v) PEG 4000 e 10% Isopropanol para a SELD de <i>E. coli</i> na presença de 5 mM ADP. ....	106
FIGURA 47 - Proteólise da SELD de <i>E. coli</i> . 1: Quimiotripsina 250 µg/ml. 2: SELD de <i>E. coli</i> 5 mg/ml. 3: Produto de proteólise com Quimiotripsina (250 µg/ml) por 20 minutos. SDS-PAGE 15%. A seta indica o produto de proteólise. ....	107
FIGURA 48 - Resultado da purificação do produto de proteólise da proteína SELD <i>E. coli</i> através de cromatografia de exclusão molecular (preto) e perfil de eluição de SELD <i>E. coli</i> (cinza). ....	108
FIGURA 49 - Eletroforese em gela nativo dos picos 1 e 2 resultantes da cromatografia de exclusão molecular do produto de proteólise da SELD de <i>E. coli</i> com Quimiotripsina. ....	108
FIGURA 50 - Cristais na forma de agulhas obtidos para a SELD de <i>E. coli</i> na ausência de ligantes após proteólise com Quimiotripsina. Condições de cristalização: A- 1,5 M sulfato de amônio / 0,1 M BIS-TRIS pH 6,5 / 0,1 M cloreto de sódio. B- 0,2 M sulfato de amônio / 0,1 M Tris pH 8,5 / 25 % PEG 3350. C- 0,2 M acetato de amônio / 0,1 M HEPES pH 7,5 / 25 % PEG 3350. ....	109

- FIGURA 51 - Cristais na forma de agulhas obtidos para a amostra purificada de SELD de *E. coli* após proteólise com Quimiotripsina, na ausência de ligantes. São mostradas as condições de cristalização em que cresceram cristais, todos na forma de agulha, sob variação de pH, concentração de acetato (A e B) e sulfato de amônio (C). ..... 110
- FIGURA 52 - Variação da anisotropia de fluorescência com a titulação de SELD de *E. coli* em amostra contendo 5  $\mu$ M do complexo SELA-SELC, sendo 10 nM de SELC marcado com fluoresceína. A linha em vermelho indica a razão estequiométrica de 1:1 (mol de monômeros de SELD : mol do complexo SELA-SELC). As barras de erro representam erros intrínsecos a cada medida no equipamento utilizado. .... 112
- FIGURA 53 - Resultado da deconvolução do espectro de massas da SELD de *E. coli* apresentando dois picos, 1 e 2, correspondentes a massas moleculares de  $36232 \pm 14$  e  $72465 \pm 29$  Da, respectivamente. .... 113
- FIGURA 54 - Perfil de incorporação de deutério na SELD de *E. coli* em diferentes tempos de troca H/D:  $t_5 = 5$  min,  $t_{10} = 10$  min,  $t_{15} = 15$  min,  $t_{30} = 30$  min,  $int = 30$  min (no complexo SELD-SELA-SELC). A incorporação de deutério nas amostras  $t_5$ ,  $t_{10}$ ,  $t_{15}$  e  $t_{30}$  foi de 68,7 % enquanto que para a amostra  $int$  foi de 62,5 %. A legenda indica faixas de porcentagem de incorporação de deutério na estrutura da proteína. Em cinza estão marcados os resíduos de aminoácidos catalíticos. .... 115
- FIGURA 55 - Perfil de incorporação de deutério na SELD de *E. coli* com 30 minutos de troca H/D, para A – amostra pura de proteína e B – após interação com o complexo SELA-SELC, indicando apenas as regiões em que houve grande variação na taxa de incorporação de deutério. As setas indicam os N-terminais de cada monômero. Em azul, são representadas regiões com porcentagem de incorporação de deutério de 0 a 14 %, enquanto que em ciano a porcentagem e incorporação é de 15 a 24 %, em amarelo é de 50 a 69 % e em vermelho é de 85 a 100 %. Os números apresentados em A indicam a posição dos resíduos de aminoácidos nas regiões assinaladas, com suas respectivas cores (esses números são indicados em apenas 3 regiões, dado que as outras são simétricas às primeiras). .... 117



## LISTA DE TABELAS

TABELA 1 - Oligonucleotídeos utilizados para amplificação da ORF do gene de interesse por meio de PCR. ....	49
TABELA 2 - Condições da reação de amplificação dos genes. ....	49
TABELA 3 - Reagentes usados nas reações de digestão do vetor de clonagem e do gene de interesse. ....	51
TABELA 4 - Reagentes usados nas reações de ligação do fragmento dos genes de interesse em vetor pET28a(+). ....	52
TABELA 5 - Parâmetros hidrodinâmicos da SELD de <i>E. coli</i> . ....	65
TABELA 6 - Porcentagem de aminoácidos na estrutura primária da SELD de <i>E. coli</i> . ....	73
TABELA 7 - Domínios da SELD de <i>E. coli</i> obtidos ....	74
TABELA 8 - Informações cristalográficas de Selenofosfato Sintetases com estrutura resolvida. ....	78
TABELA 9 - Porcentagem de aminoácidos nas regiões do gráfico de Ramachandran. ....	78
TABELA 10 - Distribuição de aminoácidos no gráfico de Ramachandran para o modelo baseado na cadeia A da estrutura de código PDB 3FD5 como molde. ....	81
TABELA 11 - Ligações de hidrogênio previstas na interface de dimerização (PISA <sup>77</sup> ). ....	85
TABELA 12 - Interações iônicas previstas na interface de dimerização (PISA <sup>77</sup> ). ....	85
TABELA 13 - Propriedades hidrodinâmicas do melhor modelo de SELD de <i>E. coli</i> . ....	86
TABELA 14 - Deconvolução do espectro de CD da SELD de <i>E. coli</i> em comparação com a do modelo tridimensional da mesma, obtido por Modelagem Molecular e com a estrutura de código PDB 3FD5. ....	91
TABELA 15 - Padrões para calibração da SUPERDEX <sup>TM</sup> 200 10/300 (GE). ....	92
TABELA 16 - Padrões para determinação de massa molecular e raio hidrodinâmico a partir de Eletroforese em Gel Nativo. ....	94
TABELA 17 - Parâmetros hidrodinâmicos da SELD de <i>E. coli</i> obtidos por DLS. ....	96
TABELA 18 - Parâmetros hidrodinâmicos para a SELD de <i>E. coli</i> obtidos por SV. ....	103



## LISTA DE ABREVIATURAS E SIGLAS

ADP	adenosina-5'-difosfato
AIRS, PurM	aminoimidazol-ribonucleotídeo sintetase
AMP	adenosina-5'-monofosfato
AMPCP	adenosina-5'-( $\alpha,\beta$ -metileno)-difosfato
AMPCPP	adenosina-5'-( $\alpha,\beta$ -metileno)-trifosfato
AMPPCP	adenosina-5'-( $\beta,\gamma$ -metileno)-trifosfato
ATP	adenosina-5'-trifosfato
AUC	Ultracentrifugação analítica, do inglês <u>A</u> nalytical <u>U</u> ltra <u>C</u> entrifugation
CD	Dicrismo circular, do inglês <u>C</u> ircular <u>D</u> ichroism
<i>D</i>	coeficiente de difusão translacional
DIOs	deiodinases
DLS	Espalhamento dinâmico de luz, do inglês <u>D</u> ynamic <u>L</u> ight <u>S</u> cattering
<i>D<sub>máx</sub></i>	dimensão máxima de uma proteína
DTT	ditiotreitól
EF-Sec	fator de alongação de tradução para inserção de selenocisteína em eucariotos
FGAR-AT ou PurL	formilglicinamida-ribonucleotídeo amidotransferase
GPxs	glutathiona peroxidases
GTP	guanosina-5'-trifosfato
H/D	hidrogênio/deutério
HypE	proteína de maturação de [NiFe]-hidrogenases
I	intensidade de espalhamento de raios-X
IDR	ingestão diária recomendada
IPTG	isopropil-tio-B-D-galactopiranosídeo
LB	<u>L</u> uria- <u>B</u> ertani
LEC	<u>L</u> aboratório de <u>E</u> spectroscopia e <u>C</u> alorimetria
LNBio	<u>L</u> aboratório <u>N</u> acional de <u>B</u> iociências
LNLS	<u>L</u> aboratório <u>N</u> acional de <u>L</u> uz <u>S</u> íncrotron
MM, M	massa molecular
mnm5Se2U	5-metilaminometil-2selenouridina
mRNA	RNA mensageiro
$N_A$	número de Avogadro
ORF	fase aberta de leitura, do inglês <u>O</u> pen <u>R</u> eading <u>F</u> rame
p(r)	função de distribuição de distâncias de centros espalhadores de raios-X
PCR	reação em cadeia da polimerase
PDB	do inglês, <u>P</u> rotein <u>D</u> ata <u>B</u> ank
Pi	fosfato inorgânico
PPi	pirofosfato
PSTK	fosfoseril-tRNA <sup>Sec</sup> -cinase
q	vetor de espalhamento de raios-X
R	constante universal dos gases
$R_g$	raio de giro
RMN	<u>R</u> essonância <u>M</u> agnética <u>N</u> uclear
<i>S</i>	coeficiente de sedimentação
SAXS	Espalhamento de raios-X a baixo ângulo, do inglês <u>S</u> mall <u>a</u> nge <u>X</u> -ray <u>s</u> cattering
SE	Sedimentação em equilíbrio, do inglês <u>E</u> quilibrium <u>S</u> edimentation

Se	selênio
Sec, U	L-selenocisteína
SECIS	sequência de inserção de selenocisteína
SELA	selenocisteína sintase
SELB	fator de alongação de tradução para inserção de selenocisteína em procaríotos
SELC, tRNA <sup>Sec</sup>	RNA transportador de selenocisteína
SELD, SPS	Selenofosfato Sintetase
SeMet	L-selenometionina
SepSecS	selenocisteína sintase
SerRs	seril-tRNA-sintetase
SV	Velocidade de sedimentação, do inglês <u>S</u> edimentaition <u>V</u> elocity
Te	telúrio
ThiL	tiamina monofosfato cinase
tRNA	RNA transportador
TRxRs	tiorredoxina redutases
$\bar{v}$	volume parcial específico
$\eta$	viscosidade
$\rho$	densidade

## Sumário

<b>1 Introdução</b> .....	27
<b>1.1 O elemento químico selênio e seu papel na saúde humana</b> .....	27
<b>1.2 Selenocisteína: biossíntese e incorporação em selenoproteínas</b> .....	30
<b>1.3 Selenosfosfato Sintetase</b> .....	37
<b>2 Objetivos específicos</b> .....	43
<b>3 Materiais e Métodos</b> .....	45
<b>3.1 Análise da proteína SELD de <i>E. coli</i> através de ferramentas de Bioinformática</b> .....	45
<b>3.1.1 Análise da sequência de aminoácidos</b> .....	45
<b>3.1.2 Modelagem das estruturas terciária e quaternária</b> .....	46
<b>3.2 Preparação da Selenofosfato Sintetase de <i>E. coli</i></b> .....	48
<b>3.2.1 Extração de DNA genômico de <i>Escherichia coli</i> DH5a</b> .....	48
<b>3.2.2 Amplificação da fase aberta de leitura do gene <i>selD</i> de <i>Escherichia coli</i></b> .	49
<b>3.2.3 Clonagem do gene <i>selD</i> de <i>Escherichia coli</i> em vetor pET28a(+)</b> (Novagen).....	50
<b>3.2.4 Expressão da proteína SELD de <i>E. coli</i></b> .....	53
<b>3.2.5 Purificação por Cromatografia de Afinidade</b> .....	53
<b>3.2.6 Diálise</b> .....	54
<b>3.2.7 Clivagem da cauda de histidinas com Trombina</b> .....	55
<b>3.2.8 Purificação por Cromatografia de Exclusão Molecular</b> .....	55
<b>3.3 Caracterização biofísica da proteína recombinante SELD de <i>E. coli</i></b> .....	56
<b>3.3.1 Espectroscopia de Dicroísmo Circular</b> .....	56
<b>3.3.2 Estimativa da massa molecular e raio hidrodinâmico por Cromatografia de Exclusão Molecular</b> .....	59
<b>3.3.3 Eletroforese em Gel Nativo</b> .....	59
<b>3.3.4 Espalhamento Dinâmico de Luz (DLS)</b> .....	59
<b>3.3.5 Espalhamento de Raios-X a Baixo Ângulo (SAXS)</b> .....	62
<b>3.3.6 Ultracentrifugação Analítica</b> .....	64
<b>3.3.7 Ensaio de Cristalização</b> .....	67
<b>3.4 Estudo da interação da SELD de <i>E. coli</i> com o complexo SELA-SELC</b> .....	69
<b>3.4.1 Anisotropia de Fluorescência</b> .....	69
<b>3.4.2 Espectrometria de massas com troca H/D</b> .....	70

<b>4 Resultados e Discussão</b> .....	73
<b>4.1 Análise da proteína SELD de <i>E. coli</i> através de ferramentas de Bioinformática</b> .....	73
4.1.1 Análise da sequência de aminoácidos .....	73
4.1.2 Modelagem da estrutura terciária e quaternária.....	76
<b>4.2 Preparação da SELD de <i>E. coli</i></b> .....	86
<b>4.3 Caracterização biofísica da proteína recombinante SELD de <i>E. coli</i></b> .....	90
4.3.1 Espectroscopia de Dicroísmo Circular .....	90
4.3.2 Estimativa da massa molecular e raio hidrodinâmico por Cromatografia de Exclusão Molecular.....	92
4.3.3 Eletroforese em Gel Nativo .....	94
4.3.4 Espalhamento Dinâmico de Luz (DLS) .....	96
4.3.5 Espalhamento de raios-X a baixo ângulo (SAXS).....	97
4.3.6 Ultracentrifugação Analítica (AUC) .....	101
4.3.7 Ensaio de Cristalização.....	105
<b>4.4 Estudo da interação da SELD de <i>E. coli</i> com o complexo SELA-SELC</b> .....	111
4.4.1 Anisotropia de Fluorescência.....	111
4.4.2 Espectrometria de massas com troca H/D .....	112
<b>5 Conclusões</b> .....	117
<b>6 Perspectivas</b> .....	119
<b>Referências</b> .....	121
<b>Anexos</b> .....	131

## 1 Introdução

### 1.1 O elemento químico selênio e seu papel na saúde humana

A descoberta do selênio pelos cientistas suecos J. J. Berzelius e J. G. Gahn em 1817 foi por acaso. Ao examinar um resíduo avermelhado, com odor fétido, encontrado em câmaras de chumbo usadas para produção de ácido sulfúrico, os cientistas isolaram uma substância desconhecida à época, com propriedades químicas similares àquelas apresentadas pelo elemento químico telúrio (do grego, *Telles*, que significa Terra). Essa substância foi, então, chamada de selênio (do grego, *Selene*, que significa Lua) <sup>1-4</sup>.

O elemento químico selênio (Se) pertence ao Grupo 16/VIA e ao Período 4 da Tabela Periódica dos elementos químicos. Portanto, o selênio e seus compostos apresentam propriedades químicas similares aos compostos de enxofre (S) e telúrio (Te). Esta posição na Tabela Periódica é responsável, em grande parte, por muitas de suas interações biológicas com o enxofre, bem como com o arsênio e o fósforo <sup>1,4</sup>.

Como o enxofre, o selênio reage com metais e ametais formando compostos iônicos com o íon seleneto ( $\text{Se}^{2-}$ ). Os estados de oxidação do selênio que ocorrem naturalmente são -2 (como o seleneto de sódio,  $\text{Na}_2\text{Se}$ ), 0 (Se, o selênio elementar), +4 (como o selenito de sódio,  $\text{Na}_2\text{SeO}_3$ ) e +6 (como o selenato de sódio,  $\text{Na}_2\text{SeO}_4$ ) <sup>1</sup>.

O selênio é um dos elementos químicos mais raros, sendo apenas o 70º em abundância entre os 88 elementos que ocorrem naturalmente na crosta terrestre <sup>4</sup>. Contudo, apesar de sua escassez, tem papel fundamental na vida animal e humana <sup>5</sup>.

Como inicialmente era visto como substância tóxica para animais, que chegavam a morrer após ingestão de plantas do gênero *Astragalus*, as quais acumulam selênio em suas folhas, as primeiras pesquisas organizadas sobre o elemento químico selênio foram direcionadas ao estudo de sua toxicidade <sup>6</sup>. Entretanto, os pesquisadores Schwarz e Foltz, em 1957, mostraram que o selênio previne a necrose de fígado em ratos na deficiência de vitamina E <sup>5,7</sup>, o que levou à investigação de seus efeitos benéficos como um micronutriente. Pinsent já havia reportado em 1954 a importância do selênio no crescimento de *Escherichia coli* na presença de glicose, em condições anaeróbias <sup>8</sup>. Desta forma, o selênio passou a ser

reconhecido pela comunidade científica como um importante fator nutricional para procariotos e eucariotos.

Este micronutriente é obtido pelo homem a partir de plantas, que o absorvem em sua forma inorgânica a partir do solo. Nas plantas, o selênio é convertido às suas formas orgânicas, como compostos metilados de selênio e os aminoácidos L-selenometionina (SeMet) e L-selenocisteína (Sec), sendo o primeiro o aminoácido mais abundante em grãos e legumes<sup>9</sup>. Em animais, este aminoácido é o principal precursor de L-selenocisteína<sup>1</sup>.

Para seres humanos, doses diárias acima de 880 µg de selênio são consideradas tóxicas, enquanto doses diárias abaixo de 20 µg são consideradas insuficientes<sup>10</sup>. A Resolução RDC nº 269, de 22 de setembro de 2005, da Agência Nacional de Vigilância Sanitária (ANVISA), regulamenta a ingestão diária recomendada (IDR) de selênio como sendo de 34 µg para adultos<sup>11</sup>.

Apesar da deficiência de selênio na dieta humana ser uma possibilidade bastante remota, devido à sua ampla distribuição geográfica nos solos, alguns casos foram relatados em áreas rurais da China, onde pessoas viviam de alimentos produzidos localmente, em solo deficiente em selênio, consumindo menos de 10 µg/dia deste micronutriente. Deste modo, essas pessoas passaram a desenvolver a doença de Keshan, uma miopatia cardíaca com sintomas parecidos com aqueles apresentados por gados criados em solo deficiente em selênio em algumas regiões dos EUA<sup>2</sup>. A doença de Kashin-Beck, uma osteoartrite endêmica em certas áreas do Tibete, também está associada à ingestão de baixas quantidades de selênio<sup>12</sup>.

Além dessas enfermidades, há evidência de que o cretinismo mixedematoso, caracterizado por retardo mental e de crescimento, esteja relacionado à deficiência em selênio e iodo, o que afeta o metabolismo do hormônio da tireoide<sup>13</sup>. Deficiência moderada de selênio também tem sido relacionada a várias condições patogênicas, como aumento de risco ao câncer, infertilidade masculina, diminuição da atividade do sistema imune e da função da tireoide, além de várias condições neurológicas, incluindo doenças de Parkinson e Alzheimer<sup>5</sup>.

A suplementação da dieta de animais em fazendas nos EUA com selênio e também seu uso em fertilizantes foram adiados por vários anos devido ao seu temido potencial carcinogênico. Porém, o pesquisador americano Clark e seus colaboradores, em 1957, demonstraram que o selênio tinha, na verdade, potencial atividade anti-câncer contra câncer de próstata, pulmão e intestino<sup>14</sup>. Com a suplementação de selênio, a indústria pecuária nos EUA economizou milhões de dólares com a prevenção de patologias animais como a doença muscular bovina, além de ter havido um aumento significativo na reprodução dos animais<sup>15</sup>.



O mecanismo molecular preciso da ação do selênio em condições fisiológicas e em algumas patologias ainda é desconhecido <sup>2</sup>. Entretanto, sabe-se que sua influência na saúde humana se deve, em grande parte, à sua presença em pelo menos 25 proteínas conhecidas como selenoproteínas, compreendendo cinco glutathione peroxidases (GPxs), três tioredoxina redutases (TRxRs), três deiodinases (DIOs) e as selenoproteínas H, I, K, N, O, S, T, W, SelP, SPS2, Sep15, SelM e MsrB1<sup>16</sup>.

A primeira evidência do papel do selênio na função enzimática foi mostrada independentemente por Rotruck e colaboradores nos EUA <sup>17</sup> e Flohé e colaboradores na Alemanha <sup>18</sup>, em 1973, quando descobriram a presença de selênio na enzima glutathione peroxidase, envolvida em processos antioxidativos. Diferentemente de outros metais que interagem com proteínas como cofatores, este elemento está presente em selenoproteínas na forma do aminoácido L-selenocisteína (Sec), como demonstrado por Cone e seus colaboradores em 1976 <sup>19</sup>. Além disso, também foi identificado o nucleotídeo modificado 5-metilaminometil-2selenouridina (mnm5Se2U) em certos tRNAs <sup>20</sup>.

As selenoproteínas têm atraído cada vez mais interesse da comunidade científica nos últimos 10 anos, como se observa pelo crescente número de publicações envolvendo o tópico (FIG. 1). Reflexo disso, o número de proteínas que contêm selênio descritas na literatura vem aumentando ultimamente, principalmente devido aos avanços da Bioinformática. Além das 25 selenoproteínas humanas, são conhecidas pelo menos 15 famílias diferentes de selenoproteínas codificadas por genomas de bactérias <sup>21</sup>. Entretanto, a distribuição de selenoproteínas varia conforme a espécie, sendo aparentemente dependente do ambiente em que o organismo vive e da disponibilidade de selênio. Algumas plantas superiores, certos insetos e fungos perderam toda a maquinaria de inserção de Sec <sup>22</sup>, enquanto outros organismos, como o *Trypanosoma brucei*, dependem de selenoproteínas para sua sobrevivência em determinadas condições de crescimento <sup>23</sup>.

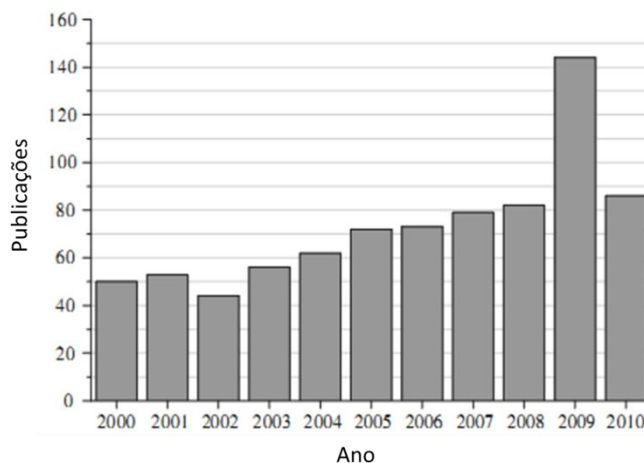


FIGURA 1 – Número de publicações científicas envolvendo o estudo de selenoproteínas (2000 – 2010). Adaptado <sup>24</sup>.

## 1.2 Selenocisteína: biossíntese e incorporação em selenoproteínas

Um grande avanço na bioquímica do selênio ocorreu em 1986, quando se descobriu que os genes das selenoproteínas glutathiona peroxidase de rato <sup>25</sup> e formato desidrogenase de *E. coli* <sup>26</sup> continham um códon TGA, conhecido por corresponder a um códon de parada de tradução no RNA mensageiro, em fase aberta de leitura (do inglês Open ReadinG Frame, ORF), na posição correspondente a uma selenocisteína na estrutura proteica. Pouco tempo depois, Leinfelder e colaboradores identificaram um RNA transportador específico para incorporação de selenocisteína em selenoproteínas, o tRNA<sup>Sec</sup> (SELC) <sup>27</sup>. Desde então, desenvolveu-se o paradigma da biossíntese de uma selenoproteína, levando ao resultado inesperado de que a selenocisteína (Sec, U) podia ser considerada como o 21º aminoácido natural geneticamente codificado<sup>28</sup>.

A estrutura química da selenocisteína é similar à da cisteína, diferindo apenas pela substituição do enxofre pelo selênio na cadeia lateral (FIG. 2). O efeito combinado das diferenças físico-químicas entre os elementos enxofre e selênio, ou seja, o maior raio atômico do selênio em comparação com o enxofre, levando à sua maior polarização e o maior comprimento da ligação Se-Se, determinam as diferenças bioquímicas entre os resíduos de aminoácidos selenocisteína e cisteína encontrados em proteínas <sup>1</sup>.

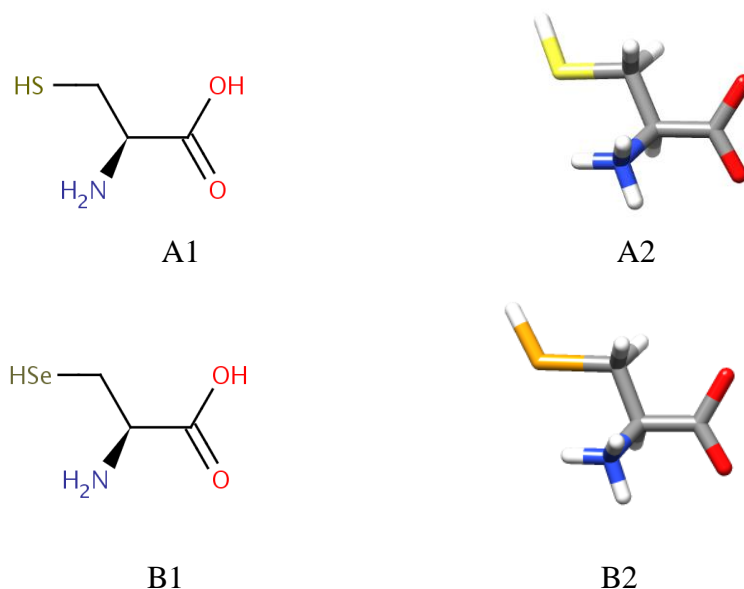


FIGURA 2 - Representações 2D e 3D dos aminoácidos A1-A2) L-cisteína e B1-B2) L-selenocisteína. Em vermelho é representado o oxigênio, em azul o nitrogênio, em amarelo o enxofre e em laranja o selênio. As figuras foram preparadas usando os programas Marvin Sketch <sup>29</sup> e UCSF Chimera <sup>30</sup>.

Uma diferença evidente entre esses aminoácidos é o  $pK_a$  igual a 5,2 para o selenolato da selenocisteína e igual a 8,5 para o tiolato da cisteína. Deste modo, em pH fisiológico (6,5 – 7,5), a molécula de selenocisteína está desprotonada e, portanto, mais propensa a atuar em reações de oxidorredução, uma das principais funções de selenoproteínas. Porém, muitas proteínas homólogas contendo cisteína no lugar de selenocisteína mantêm a capacidade antioxidante <sup>31, 32</sup>. Para tal comparação, deve-se levar em consideração o microambiente na estrutura proteica na qual está inserido o aminoácido selenocisteína. Assim, essa diferença de  $pK_a$  não explica a razão pela qual selenoproteínas são expressas na natureza, substituindo suas homólogas com resíduo de cisteína na posição catalítica <sup>31</sup>. Além disso, a eficiência catalítica de reações enzimáticas envolvendo cisteína no lugar de selenocisteína mostrou-se de 10 a 100 vezes menor em mutantes de selenoproteínas que atuam como oxidoredutases e, portanto, não justifica o maior gasto energético para a produção de selenocisteína, considerando toda a maquinaria de síntese e incorporação desse aminoácido em proteínas <sup>33</sup>.

A hipótese levantada por E. S. J. Arnér <sup>31</sup>, em 2010, é que a característica que torna uma selenoproteína única em sua função é a maior nucleofilicidade do selênio presente na selenocisteína em comparação ao enxofre da cisteína. Deste modo, o autor propõe que a selenocisteína, como resíduo catalítico em selenoproteínas envolvidas em reações de oxidorredução, leve a uma maior taxa de reação química com eletrófilos quando comparada à cisteína na mesma posição, que atua apenas mantendo o equilíbrio redox e não é capaz de mimetizar a ação da selenocisteína mesmo no microambiente químico de uma proteína.

A biossíntese de selenocisteína e sua incorporação em selenoproteínas de maneira específica é extremamente custosa para os organismos que as produzem devido ao número de fatores celulares dedicados a redefinir os códons UGA específicos de terminação da tradução para a inserção de selenocisteína <sup>31</sup>.

Em procariotos, a via de biossíntese e incorporação específica de selenocistena em selenoproteínas tem sido bastante estudada ao longo dos últimos 20 anos, sendo mais bem entendida na enterobactéria *E. coli* <sup>28, 34</sup>. Formalmente, a biossíntese de selenocisteína em *E. coli* segue três passos (resumidos na FIG. 6-A, juntamente com os passos seguintes da via de inserção de selenocisteína em proteínas):

a) Carregamento do tRNA<sup>Sec</sup> com L-serina, consumindo ATP, em uma reação catalisada pela enzima seril-tRNA sintetase (SerRS) <sup>35</sup>.

b) Formação de um intermediário 2,3-aminoacrilil-tRNA<sup>Sec</sup> a partir do seril-tRNA<sup>Sec</sup>, catalisada pela selenocisteína sintase (SELA) na presença do grupo prostético piridoxal-5'-fosfato <sup>36</sup>.

c) Ataque nucleofílico do selênio ao intermediário 2,3-aminoacrilil-tRNA<sup>Sec</sup>, levando à formação do selenocisteil-tRNA<sup>Sec</sup>, sendo que o selênio é obtido a partir de sua forma ativa, o monoselenofosfato, em uma reação catalisada pela enzima Selenofosfato Sintetase (SELD) <sup>37</sup>.

Forchhammer e colaboradores <sup>36</sup>, em 1991, através de um experimento de microscopia eletrônica de alta resolução, mostrou que 5 moléculas de tRNA<sup>Sec</sup> se ligam a um homodecâmero de SELA em *E. coli*. Como o número de moléculas de tRNA<sup>Sec</sup> na célula é cerca de 250 e de decâmeros de SELA é de aproximadamente 150 em *E. coli*, acredita-se que esta enzima atue como um reservatório de selenocisteil-tRNA<sup>Sec</sup>, uma maneira eficiente de utilizar essa molécula na célula <sup>14</sup>.

O tRNA<sup>Sec</sup> é único, tendo características distintas dos demais tRNAs conhecidos. Em *E. coli*, apresenta 95 nucleotídeos, sendo que 8 pares de bases estão no braço acceptor e 5 no braço T, além de 6 no braço D <sup>38</sup>. Em eucariotos e arqueias o número de nucleotídeos é cerca de 90, sendo 9 pares de bases no braço acceptor e 4 no braço T. Além disso, verifica-se a presença do nucleotídeo modificado 5-metilaminometil-2selenouridina (mnm5Se2U) na região do anti-códon do tRNA<sup>Sec</sup> desses organismos <sup>39</sup> (FIG. 3).

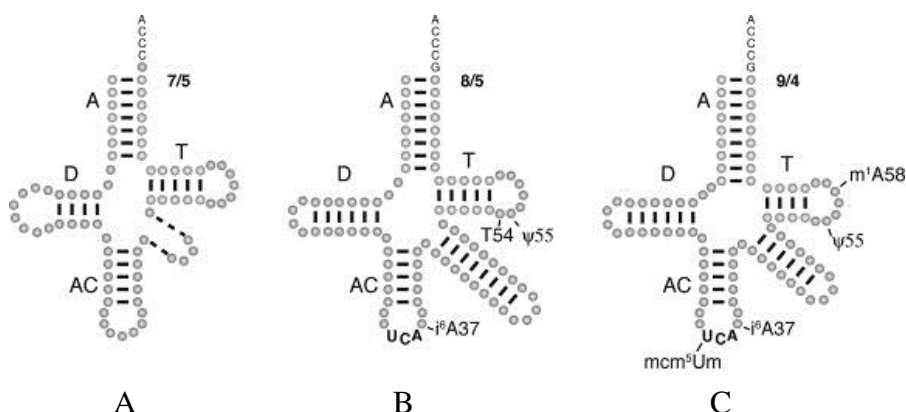


FIGURA 3 - Estrutura secundária encontrada em A) tRNAs em geral, B) tRNA<sup>Sec</sup> de eubactérias e C) tRNA<sup>Sec</sup> de arqueias e eucariotos. Adaptado <sup>38, 39, 40</sup>.

A incorporação específica do aminoácido selenocisteína em selenoproteínas de *E. coli* requer o reconhecimento do códon UGA no mRNA de selenoproteínas como sendo códon para inserção de selenocisteína e não um códon de parada de tradução. Para isso, logo após esse códon existe uma estrutura secundária na forma de grampo conhecida como SECIS (Sequência de Inserção de Selenocisteína). Em *E. coli*, esta estrutura contém 2 hélices e 2 voltas (*loop*), além de dois pares de bases não-Watson-Crick <sup>40</sup> (FIG. 4A). A SECIS é reconhecida por um fator de alongação de tradução, chamado SELB em *E. coli*, em sua porção C-terminal denominada domínio IV. SELB também interage com alta especificidade com o selenocisteil-tRNA<sup>Sec</sup> <sup>41</sup>. A interação com ambos os RNAs leva à formação de um complexo quaternário estável entre SELB, GTP, selenocisteil-tRNA<sup>Sec</sup> e elemento SECIS. A inserção de selenocisteína no polipeptídeo nascente ocorre no sítio A do ribossomo, após a hidrólise de GTP catalisada pela SELB <sup>40</sup>.

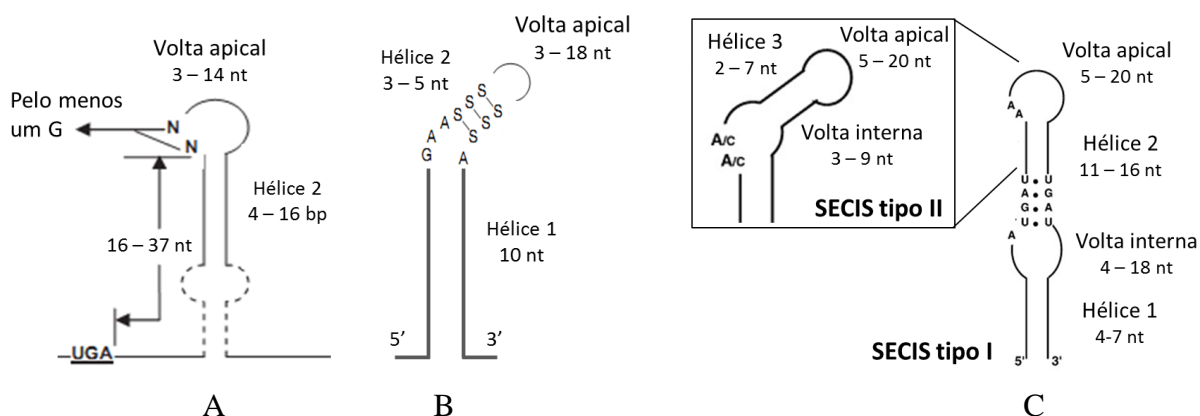


FIGURA 4 - Estrutura consenso do elemento SECIS em A) eubactérias, B) arqueias e C) eucariotos. Em eucariotos são observados dois tipos de elemento SECIS: I e II. O elemento SECIS tipo II de eucariotos é mostrado na caixa. Adaptado <sup>42, 34</sup>.

Excesso de selenocisteil-tRNA<sup>Sec</sup> na célula pode levar a um aumento do nível de selenocisteína livre, por sua hidrólise espontânea, e esta pode ser reciclada e servir como doador de selênio para síntese *de novo* de selenocisteil-tRNA<sup>Sec</sup><sup>43</sup>. Porém, a selenocisteína também pode se infiltrar em vias de biossíntese de proteínas e substituir a cisteína, o que pode levar à inativação da enzima afetada devido à diferença no potencial redox do grupamento selenol, presente na selenocisteína, em relação ao tiol, presente na cisteína<sup>44</sup>. Assim, deve haver regulação da biossíntese do tRNA<sup>Sec</sup> aminoacilado com selenocisteína<sup>40</sup>.

No genoma de *E. coli*, os genes que codificam para os componentes da maquinaria de biossíntese e inserção de selenocisteína em selenoproteínas estão organizados em três unidades de transcrição do genoma, conhecidas como *selAB* (FIG. 5A) (correspondente às proteínas SELA e SELB), *selC* (correspondente ao tRNA<sup>Sec</sup>) e *selD* (correspondente à proteína SELD). A transcrição dessas unidades é constitutiva e, portanto, independente da condição fisiológica. Porém, a tradução da unidade *selAB* está sujeita à regulação<sup>40</sup>.

Estrutura similar ao elemento SECIS na região 5' não-traduzida do transcrito *selAB* de *E. coli* foi identificada em 2002 por M. Thanbichler e A. Böck<sup>40</sup> (FIG. 5B). Essa estrutura se liga especificamente à SELB, com afinidade seis vezes menor quando comparada ao elemento SECIS, formando um complexo quaternário com GTP e selenocisteil-tRNA<sup>Sec</sup>. Como a formação desse elemento de estrutura secundária não influencia na taxa de transcrição, concluiu-se que a formação do complexo quaternário está envolvida na regulação da iniciação da tradução, atuando como repressor de tradução de SELA e SELB.

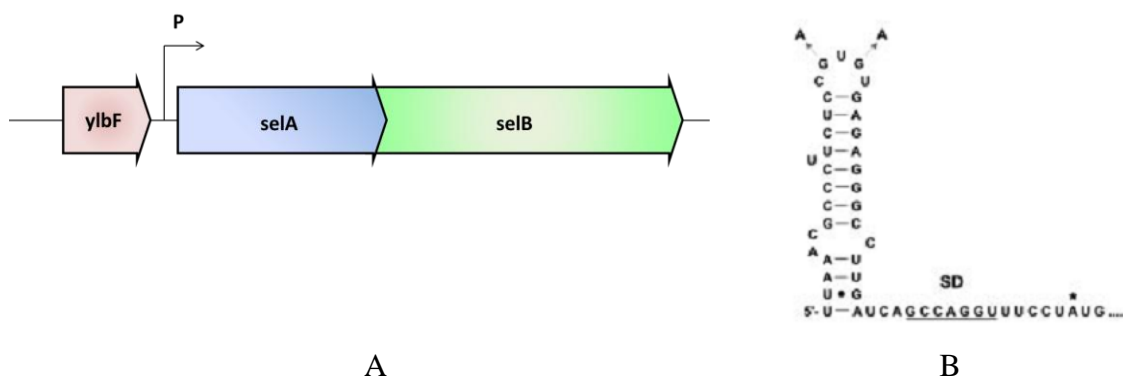


FIGURA 5 - A) Organização do operon *selAB* em *E. coli*. B) Estrutura do elemento similar ao SECIS na região 5' não traduzida do operon *selAB*. Na figura, destaca-se a região de ligação a ribossomo e o asterisco indica o início da região codificadora do gene *selA*. Adaptado<sup>40</sup>.

Vias de biossíntese e inserção de selenocisteína comparáveis à encontrada em bactérias também foram verificadas em arqueias<sup>45</sup> e eucariotos<sup>16, 34</sup>. Porém, no caso desses organismos, a biossíntese de selenocisteil-tRNA<sup>Sec</sup> envolve primeiramente a fosforilação do

seril-tRNA<sup>Sec</sup> pela enzima fosfoseril-tRNA<sup>Sec</sup>-cinase (PSTK), específica para seril-tRNA<sup>Sec</sup> e não para seril-tRNA<sup>Ser</sup>. Em seguida, o fosfoseril-tRNA<sup>Sec</sup> é convertido em selenocisteil-tRNA<sup>Sec</sup>, reação catalisada pela selenocisteína sintase (SepSecS)<sup>34, 39</sup>. Como em procariotos, o monoselenofosfato é o doador ativo de selênio, obtido a partir de seleneto e ATP em uma reação catalisada pela Selenofosfato Sintetase, isoforma 2 (SPS2) em eucariotos<sup>34</sup>. A isoforma 1 (SPS1) de eucariotos não é capaz de sintetizá-lo, o que condiz com o fato de que insetos que apresentam somente a SPS1 não são capazes de sintetizar selenoproteínas<sup>46</sup>.

Em eucariotos e arqueias, também é necessária a presença do elemento SECIS no mRNA das selenoproteínas. Porém, essa estrutura está localizada na região 3' não traduzida (3'-UTR), sendo que são descritos dois tipos de SECIS em eucariotos, que se diferem pela presença de uma terceira hélice e uma volta apical menor no tipo II quando comparado ao tipo I<sup>34</sup> (FIG. 4B e 4C). A maquinaria de incorporação de selenocisteína em selenoproteínas de eucariotos (FIG. 6B) e arqueias (FIG. 6C) também exige um fator de alongação de tradução, conhecido como EF-Sec, assim como procariotos, que apresentam SELB. Porém, apesar de se ligar especificamente a tRNA<sup>Sec</sup>, direcionando-o para o sítio A do ribossomo, não é capaz de interagir com a SECIS, já que não apresenta o domínio IV no C-terminal, observado em SELB.

Um outro fator de alongação de tradução, conhecido como SBP2, é responsável pela interação com SECIS e EF-Sec em eucariotos<sup>47</sup>. Esse fator de alongação não foi identificado em arqueias e não se sabe se existe um intermediador da interação entre SECIS e EF-Sec nesses organismos<sup>39</sup>. Estudos recentes, demonstraram, ainda, que a proteína SECp43 aumenta a interação entre EF-Sec e SBP2 *in vivo*<sup>34, 48</sup> e que a proteína ribossomal L30 se liga ao elemento SECIS e, portanto, deve estar envolvida no mecanismo de decodificação do códon UGA em eucariotos<sup>34</sup>.

Em eucariotos, o processo de transcrição é compartimentalizado no núcleo e o processo de tradução é citoplasmático. Allmang e colaboradores<sup>34</sup>, baseado em evidências experimentais prévias<sup>48</sup>, propôs, em 2009, a formação de um complexo, no núcleo, entre SPS1, SECp43, EF-Sec e selenocisteil-tRNA<sup>Sec</sup> e posterior reconhecimento do complexo formado pela SBP2 e o elemento SECIS. O complexo final é, então, exportado ao citoplasma, onde ocorre a incorporação de selenocisteína a uma selenoproteína em biossíntese no sítio A de um ribossomo.

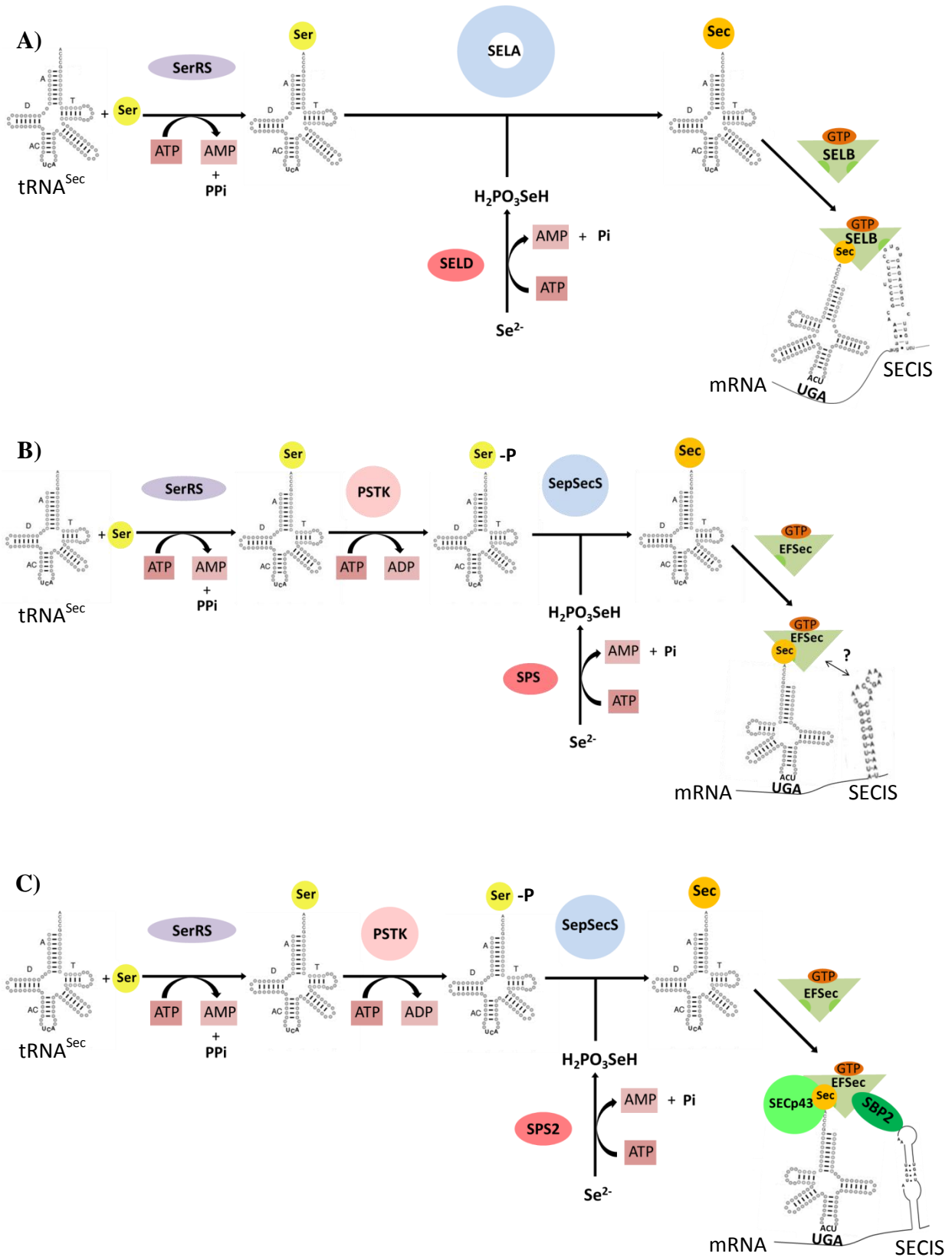


FIGURA 6 - Representação esquemática comparativa das vias de biossíntese e incorporação de selenocisteína em selenoproteínas de A) eubactérias, B) arqueias e C) eucariotos. No caso de eucariotos, representam-se apenas as etapas que ocorrem no citoplasma celular. O ribossomo foi omitido nas figuras. O ponto de interrogação em B) indica que ainda não está determinada a ligação do elemento SECIS de arqueias no complexo selenocisteil-tRNA<sup>Sec</sup>-EF-Sec. Adaptado <sup>49</sup>.



Em contraste com a via específica de metabolismo de selênio, também foi identificada sua incorporação não-específica, resultado da similaridade química do selênio com o enxofre. Quando presente em alta concentração na célula, isto é, em concentrações de selenito acima de 1 $\mu$ M, o selênio entra nas vias de metabolismo de enxofre e é metabolizado a seleneto. O seleneto é, então, usado para a produção de monoselenofosfato, o qual é necessário para a produção de selenocisteína ou 5-metilaminometil-2-selenouridina pela via específica de metabolismo de selênio<sup>50</sup>.

O selenato, quando absorvido pela célula, é carregado pela via de transporte de selenato e reduzido a seleneto via sistema de redução de sulfato. Quando oferecido à célula na forma de selenito, a redução parece ocorrer através da interação com compostos com grupamento tiol, como a glutathione. Bactérias crescidas na presença de selenito apresentam cheiro característico de alho devido à produção de formas metiladas de selênio, que são voláteis<sup>51</sup>.

A primeira forma biológica do selênio formada na célula é a selenocisteína, que pode ser convertida a selenocistationa e selenometionina. Por outro lado, selenocisteína livre também pode ser usada na síntese de selenocisteil-tRNA<sup>Cys</sup> e, assim, há incorporação de selenocisteínas na posição de cisteínas em proteínas. Quando há selenometionina disponível no meio celular, esta é prontamente incorporada em proteínas nas posições de metioninas. Este resultado é bastante empregado para a resolução da estrutura cristalográfica de proteínas pelo método de dispersão anômola em múltiplos comprimentos de onda. Por outro lado, selenocisteína livre é altamente tóxica e, portanto, inibitória para o crescimento celular<sup>50</sup>.

### 1.3 Selenosfosfato Sintetase

A enzima Selenofosfato Sintetase (SELD, SPS) de *E. coli* foi caracterizada por Leinfelder e colaboradores em 1990<sup>20</sup>. Em seu trabalho, a fase aberta de leitura do gene que codifica para SELD em *E. coli* foi clonada em vetor pACYC184 e sequenciada. O gene *selD* em *E. coli* codifica para uma proteína de 347 aminoácidos com massa molecular calculada de 36,687 kDa. Ensaio de complementação funcional de cepas de *E. coli* com o gene *selD* deletado demonstraram que a proteína SELD é necessária para a incorporação de selênio no nucleotídeo modificado 5-metilaminometil-2-selenouridina de tRNA e para a biossíntese de

selenocisteína. Além disso, usando extratos com SELD e SELA superexpressos, observou-se a formação eficiente de selenocisteil-tRNA<sup>Sec</sup> a partir de Ser-tRNA<sup>Sec</sup> 20.

A primeira estrutura cristalográfica de uma Selenofosfato Sintetase foi publicada em 2008 52 para a SPS do procaríoto *Aquifex aeolicus*, truncada no N-terminal (código PDB 2ZAU) e na ausência de ligante (código PDB 2ZAU), revelando uma proteína com enovelamento do tipo  $\alpha/\beta$ , composto por dois domínios homólogos àqueles apresentados pelas proteínas pertencentes à Superfamília AIRS, onde se encontram proteínas que têm ATP como substrato: aminoimidazol-ribonucleotídeo sintetase 53 (AIRS ou PurM), formilglicinamida-ribonucleotídeo amidotransferase 54 (FGAR-AT ou PurL), proteína de maturação de [NiFe]-hidrogenases 55 (HypE), tiamina monofosfato cinase 56 (ThiL), além da Selenofosfato Sintetase 52 (SELD ou SPS). Todas estas enzimas apresentam conservação da sequência Dx<sub>4</sub>GA/GxP no N-terminal, responsável pela ligação de ATP 53.

Um fato interessante é que, com exceção da PurL, todas as outras possuem tamanho similar, com aproximadamente 340 aminoácidos, e apresentam-se em forma ativa dimérica. A enzima Selenofosfato Sintetase apresenta uma característica única dentro desta família de proteínas, já que, enquanto as outras catalisam a hidrólise de ATP a ADP, a SPS consome ambas as ligações fosfoéster de alta energia do ATP 52, 53, formando AMP.

No início da década de 1990, estudos de espectroscopia de NMR mostraram que um derivado lábil contendo selênio, produto da reação catalisada pela enzima Selenofosfato Sintetase de *Escherichia coli*, era, na verdade, um composto que continha selênio ligado a um grupamento fosfato, o monoselenofosfato 37. A reação catalisada pela enzima SELD (EC 2.7.9.3), deste modo, foi proposta como sendo:

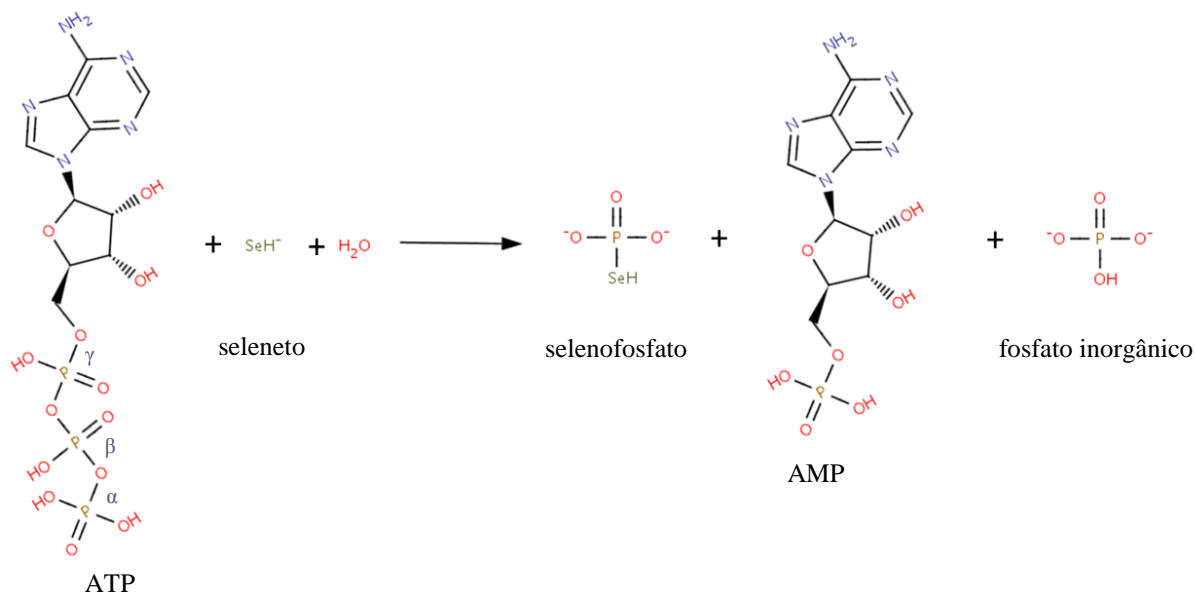


FIGURA 7 – Reação catalisada pelas Selenofosfato Sintetases.

O grupamento fosfato do monoselenofosfato é derivado do  $\gamma$ -fosfato do ATP, enquanto o fosfato inorgânico é derivado do grupamento  $\beta$ -fosfato<sup>57, 58</sup>. A enzima SELD catalisa a hidrólise do ATP a AMP também na ausência de seleneto, o que sugere que a quebra da ligação  $\beta$ - $\gamma$  fosfoanidro deve ser a primeira etapa do mecanismo de reação da enzima, independente da presença do seleneto no sítio ativo<sup>59</sup>.

A estrutura cristalográfica da Selenofosfato Sintetase da bactéria *Aquifex aeolicus* em complexo com um análogo da molécula de ATP, a adenosina 5'-( $\alpha$ - $\beta$ -metileno)-tiofosfato (AMPCPP)<sup>60</sup> (código PDB 2YYE), revelou dois sítios de ligação a ATP, formados na interface de dimerização do homodímero. Quatro resíduos de asparagina coordenam quatro íons  $\text{Co}^{2+}$  que se ligam aos grupamentos fosfato do AMPCPP. Os íons  $\text{Co}^{2+}$  devem ser substituídos *in vivo* pelos íons  $\text{Mg}^{2+}$ , essenciais para a atividade catalítica da SELD. O grupamento  $\gamma$ -fosfato é acessível ao solvente, o que sugere que um nucleófilo putativo poderia fazer um ataque nucleofílico nesta posição para gerar selenofosfato e ADP<sup>60</sup>.

Além disso, Kim e colaboradores, em 1992, mostraram que uma cisteína na posição 17 da estrutura primária da SELD de *E. coli*, localizada na volta N-terminal rica em glicina, é essencial para sua atividade catalítica. As estruturas cristalográficas da Selenofosfato Sintetase da bactéria *Aquifex aeolicus* mostram que o resíduo de cisteína/selenocisteína catalítico no N-terminal móvel pode apontar tanto para o solvente, em condições para aceitar seleneto (na forma “aberta” da proteína), formando perseleneto ( $\text{Sec-Se}^-$ ), quanto para o ATP ligado ao sítio catalítico (na forma “fechada”), podendo entregar o seleneto, quando em forma de perseleneto ( $\text{Sec-Se}^-$ ), para formação de monoselenofosfato. Uma molécula de água próxima ao grupamento  $\beta$ -fosfato poderia atuar como um nucleófilo na subsequente hidrólise do ADP a AMP e  $\text{Pi}$ <sup>60</sup>.

É improvável que o seleneto livre seja o substrato da SPS *in vivo*, já que a constante de Michaelis-Menten ( $K_M$ ) da enzima SELD de *E. coli* na presença de seleneto é igual a 7,3  $\mu\text{M}$ , valor que se encontra no limiar de toxicidade para muitos organismos<sup>61</sup>. Uma proteína *Nifs-like*, como a CsdB de *E. coli*, homóloga a cisteína desulfurases e selenocisteína liases, deve entregar o seleneto derivado de selenocisteína para a SPS *in vivo*. Proteínas *Nifs-like*, em bactérias, e selenocisteína liases, em eucariotos, apresentam o mesmo enovelamento das selenocisteína sintases, sendo enzimas piridoxal-5'-fosfato dependentes e catalisam a abstração do selênio/enxofre de selenocisteína/cisteína livre. A interação direta entre a Selenofosfato Sintetase e uma proteína *Nifs-like* em bactéria foi postulada e investigada por Docagem Molecular por Itoh e colaboradores em 2009<sup>60</sup>. Neste trabalho, eles mostraram que a volta N-terminal rica em glicina conservada entre Selenofosfato Sintetases pode alcançar o

grupo selenossulfeto formado na cisteína 364 (Cys364) presente no sítio ativo da enzima CsdB de *E. coli* (código PDB 1KMK).

Em animais, existem dois homólogos de Selenofosfato Sintetase, designados por SPS1 e SPS2, com identidade sequencial de 72% em humano. Enquanto observa-se a conservação da cisteína/selenocisteína catalítica para a SPS2, a SPS1 não a possui e sua função é desconhecida. Especula-se que a SPS1 esteja envolvida na via de salvação de selênio que recicla selenocisteína livre <sup>62</sup>.

Wang e colaboradores, em 2009, a partir da análise das estruturas cristalográficas da proteína SPS1 humana em complexo com ADP e seu análogo AMPCP (código PDB 3FD5 e 3FD6), além de estudos bioquímicos da SPS1 na presença de íons monovalentes, mostraram que o íon  $K^+$  deve catalisar a hidrólise do ADP no sítio ativo da enzima, posicionando o grupamento  $\beta$ -fosfato para ataque nucleofílico por uma molécula de água. Este trabalho, ainda, indica que a SPS deve consumir a segunda ligação fosfoéster para proteger o produto selenofosfato lábil durante a reação catalítica. Deste modo, a proposta para o mecanismo de reação é dividida em duas etapas:

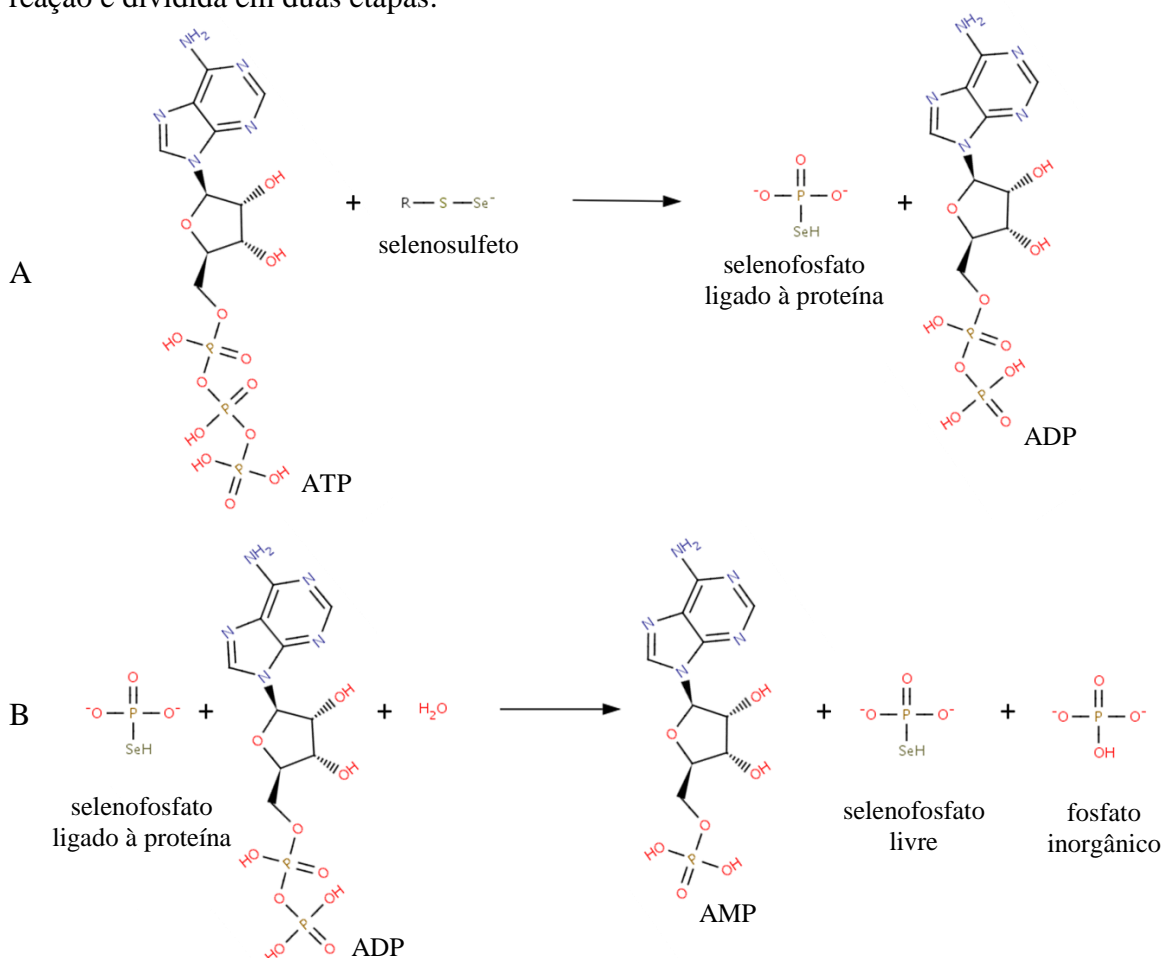


FIGURA 8 – Primeira (A) e segunda (B) etapas da reação catalisada pelas Selenofosfato Sintetases.

Small-Howard e colaboradores <sup>48</sup>, em 2006, mostraram que a proteína SPS1 interage *in vitro* e *in vivo* com a selenofosfato sintase (SepSecS) humana, que catalisa a formação do selenocisteil-tRNA<sup>Sec</sup>, a partir do monoselenofosfato produzido na reação catalisada pela SPS2. Tanto a SepSecS quanto a selenocisteína liase possuem um enovelamento do tipo I de proteínas ligadoras de piridoxal-5'-fosfato. Desta forma, se a SPS2 também é capaz de se ligar a SepSecS, então o monoselenofosfato sintetizado pela SPS2 pode ser entregue à SepSecS, como postulado por Itoh e colaboradores <sup>60</sup>. A transmissão sequencial de selênio entre *Nifs*-SELD-SELA, em bactéria, e Sec-liase-SPS2-SepSecS, em eucariotos, é razoável para evitar a difusão de compostos reativos de selênio, levando à eficiente biossíntese de selenocisteína.

No presente trabalho, a proteína recombinante SELD de *E. coli* foi caracterizada em solução quanto à sua forma oligomérica. Dada a proposta de transmissão sequencial do composto biologicamente ativo de selênio da SELD à SELA <sup>60</sup>, a interação SELD-SELA-SELC foi investigada *in vitro* e possíveis regiões de interação da SELD com o complexo SELA-SELC foram analisadas.



## 2 Objetivos específicos

- Analisar a sequência de aminoácidos da proteína SELD de *Escherichia coli* quanto às suas características físico-químicas e construir um modelo de sua estrutura tridimensional.
- Amplificar a fase aberta de leitura do gene *selD* a partir do DNA genômico de *Escherichia coli* e cloná-la em vetor de expressão pET28a(+) (Novagen).
- Expressar a proteína recombinante SELD e purificá-la.
- Caracterizar a proteína recombinante SELD por meio das técnicas: Espectroscopia de dicroísmo circular, Eletroforese em gel nativo, Espalhamento dinâmico de luz, Espalhamento de raios-X a baixos ângulos e Ultracentrifugação analítica.
- Realizar testes de cristalização com a proteína recombinante SELD.
- Investigar a interação da proteína SELD com o complexo formado entre a proteína SELA e o tRNA<sup>Sec</sup> (SELC) por meio das técnicas de Anisotropia de fluorescência e Espectrometria de massas com troca hidrogênio/deutério.





## 3 Materiais e Métodos

### 3.1 Análise da proteína SELD de *E. coli* por meio de ferramentas de Bioinformática

A predição *in silico* de propriedades das proteínas tem-se mostrado bastante útil na sua investigação funcional e também como guia para o trabalho experimental com uma dada proteína. Dada sua sequência primária, ou seja, sua sequência de aminoácidos, é possível fazer predições acerca de estrutura secundária, terciária e quaternária, possíveis modificações pós-traducionais e até mesmo de interações com outras proteínas. Deste modo, ferramentas de Bioinformática foram empregadas para o estudo da proteína SELD de *E. coli*.

#### 3.1.1 Análise da sequência de aminoácidos

A análise teórica dos parâmetros físico-químicos da proteína SELD de *E. coli* foi feita através da ferramenta *PROTPARAM* do servidor *ExPASy*<sup>63</sup>, a partir de sua sequência de aminoácidos. O servidor *DISULFIND*<sup>64</sup> foi então utilizado para prever se as cisteínas da proteína poderiam fazer ligações dissulfeto para manutenção da estrutura proteica.

Uma análise de domínios para esta sequência de aminoácidos foi feita através do servidor *SMART*<sup>65</sup> e também na Base de Dados de Domínios Conservados<sup>66</sup> (*CDD*, sigla em inglês), onde também foi possível identificar a conservação de aminoácidos importantes para manutenção de sua estrutura e função. Um alinhamento múltiplo com sequências homólogas foi gerado por meio do programa *ClustalW*<sup>67</sup> e a predição de estrutura secundária foi feita por meio do servidor *PSIPRED*<sup>68</sup>.

### 3.1.2 Modelagem das estruturas terciária e quaternária

A estrutura tridimensional de uma proteína e seu arranjo em estrutura quaternária estão diretamente relacionados à sua função biológica. Métodos experimentais como cristalografia de proteínas ou ressonância magnética nuclear (RMN) são bastante empregados para resolução da sua estrutura tridimensional, porém apresentam diversas limitações experimentais, como necessidade de alta concentração proteica e condições experimentais adequadas, muitas vezes, difíceis de serem alcançadas com uma amostra proteica. Por isso, métodos computacionais têm sido desenvolvidos para a predição da estrutura tridimensional de proteínas <sup>69</sup>.

Tem-se evidência de que a estrutura tridimensional de proteínas é mais conservada evolutivamente que a sequência de aminoácidos, sendo que, nos últimos anos, menos de 15% das estruturas depositadas no PDB (do inglês, *Protein Data Bank*) apresentam um tipo de enovelamento novo <sup>70</sup>. Assim, os métodos computacionais de predição de estrutura proteica são, em geral, baseados em estruturas conhecidas obtidas por cristalografia ou RMN, tomadas como molde para a Modelagem Molecular de proteínas.

Dentre as técnicas de Modelagem Molecular de proteínas, destacam-se a modelagem *Ab Initio*, o *Threading* e a Modelagem Comparativa ou Modelagem por Homologia, sendo que a primeira não utiliza de nenhuma informação já conhecida acerca da estrutura tridimensional de proteínas homólogas, partindo de um campo de força molecular que define a interação entre os átomos que compõem a sequência primária e de condições iniciais do sistema. Para tal, é necessário definir o ambiente químico em que se considera o sistema, isto é, se a modelagem considera a presença de água, por exemplo. Porém, esta técnica é limitada à capacidade de processamento do computador utilizado, dificilmente sendo utilizada para predição de estrutura de sequências de aminoácidos relativamente grandes (>100 aminoácidos).

Proteínas com identidade da sequência de aminoácidos acima de 20 % com proteínas cuja estrutura já é conhecida por cristalografia de raios-X ou RMN podem ser usadas para modelagem por *Threading* (de 20 – 30 % de identidade sequencial) ou mesmo modelagem comparativa (acima de 30 % de identidade sequencial). Porém, para valores de identidade entre sequências de aminoácidos próximos a 30%, a escolha entre uma técnica ou outra depende bastante da qualidade do alinhamento entre as sequências de aminoácidos. Além

disso, a combinação de técnicas de modelagem molecular tem se mostrado bastante útil na modelagem de proteínas que não apresentam uma boa cobertura no alinhamento de sua sequência de aminoácido com homólogas <sup>69</sup>. Na busca por um modelo confiável da estrutura tridimensional da SELD de *E. coli*, ambas as técnicas foram empregadas para a modelagem do monômero da proteína, sendo que a Modelagem por *Threading* foi feita pelo servidor *I-Tasser* <sup>69</sup> e a Modelagem Comparativa foi feita utilizando-se o programa *Modeller 9.10* <sup>73</sup>.

No caso do *Threading* (que significa alinhamento), inicia-se com a sequência de aminoácidos da proteína em estudo e procuram-se, em um banco de dados de estruturas de proteínas, fragmentos de sequências de aminoácidos com estrutura tridimensional conhecida que se alinham com a sequência em estudo. Esta etapa é guiada pela similaridade local entre a estrutura secundária predita para a proteína em estudo e a apresentada pelo fragmento molde daquela região. A qualidade do alinhamento é medida baseando-se na significância estatística do melhor alinhamento, no caso do servidor *I-Tasser* <sup>69</sup> utilizado neste trabalho. O resultado são fragmentos de estruturas proteicas do banco de dados cuja sequência de aminoácidos corresponde a regiões da estrutura primária da proteína em estudo. A organização dos fragmentos em uma estrutura tridimensional é feita através de simulação de Monte Carlo por *replica-exchange* neste servidor, até que se encontre um modelo otimizado. Caso existam regiões da sequência de interesse que não alinhem com nenhuma sequência com estrutura conhecida, este servidor utiliza modelagem *Ab initio* para modelagem das mesmas. A avaliação do melhor modelo resultado do *threading* foi feita por meio dos servidores de validação de estrutura proteica SAVES <sup>71</sup> e ProSa <sup>72</sup>.

No caso da Modelagem Comparativa, para a qual foi utilizado o programa *Modeller 9.10* <sup>73</sup>, a construção do modelo tridimensional para a proteína alvo se inicia com a busca de sequências homólogas à proteína em estudo em banco de dados de sequências de aminoácidos de proteínas cuja estrutura já está resolvida. Neste trabalho foi utilizada a ferramenta *BLASTp* do servidor *BLAST* <sup>74</sup> para esta etapa. A construção do modelo parte do alinhamento com as sequências de aminoácidos das estruturas-molde, feita por meio do programa *ClustalW* <sup>75</sup>, e busca satisfazer restrições espaciais, obtidas a partir da estrutura tridimensional do molde. A avaliação dos modelos gerados foi feita através do valor de energia dos modelos (DOPE) e dos servidores SAVES <sup>71</sup> e ProSa <sup>72</sup>.

Os modelos de estrutura terciária gerados foram visualizados e analisados utilizando os programas *PYMOL* (*DeLano Scientific LLC*) e *UCSF Chimera* <sup>30</sup>. A modelagem de um homodímero foi feita a partir do alinhamento 3D de dois monômeros com as cadeias A e B da estrutura de código PDB 3FD5, por meio do programa *UCSF Chimera* <sup>30</sup>, seguida de

minimização de energia pelo servidor *ROSETTA DOCK*<sup>76</sup>. Este servidor identifica conformações de baixa energia do complexo de interação proteína-proteína por meio de otimização de orientação de corpo rígido e conformação de cadeias laterais. A análise do resultado foi feita pelo servidor *PISA*<sup>77</sup>, que é uma ferramenta para a exploração de interfaces entre macromoléculas e predição de possíveis estruturas quaternárias. A obtenção de propriedades hidrodinâmicas do modelo estrutural para a SELD de *E. coli* foi feita através do programa *HYDROPRO*<sup>78</sup>.

## **3.2 Preparação da SELD de *E. coli***

A preparação da proteína recombinante SELD de *E. coli* foi feita a partir da amplificação da fase aberta de leitura (ORF) do gene *selD* utilizando DNA genômico de *E. coli*, seguida de sua clonagem em vetor de expressão pET28a(+), expressão da proteína SELD recombinante e purificação por cromatografia de afinidade e cromatografia de exclusão molecular. As etapas envolvidas na preparação da SELD de *E. coli* foram acompanhadas pela técnica Dra. Susana A. Sculaccio e encontram-se detalhadas nas subseções seguintes.

### **3.2.1 Extração de DNA genômico de *Escherichia coli* DH5a**

Células da cepa de *Escherichia coli* DH5a foram inoculadas em 5 mL de meio de cultura Luria-Bertani (LB) e incubadas por 16 horas sob agitação de 250 rpm a 37 °C. Um volume de 1 ml desta amostra foi utilizado para extração de DNA genômico por meio do kit *Wizard® Genomic DNA Purification* (PROMEGA), de acordo com as condições do fabricante.

### 3.2.2 Amplificação da fase aberta de leitura do gene *selD* de *Escherichia coli*

Com base na sequência de nucleotídeos do gene *selD* de *Escherichia coli*, com número de acesso NC\_004431 no banco de dados do NCBI <sup>79</sup>, foram sintetizados oligonucleotídeos com os sítios de clivagem para enzimas de restrição *NdeI* e *XhoI* (TABELA 1), de modo que fosse possível a amplificação da fase aberta de leitura (ORF) do gene *selD* de *E. coli*. Nesta etapa, foram utilizados os programas *BioEdit* <sup>80</sup> e *Gene Runner* <sup>81</sup> para alinhamento dos fragmentos de nucleotídeos, verificação dos sítios de clivagem e cálculo das temperaturas de anelamento.

TABELA 1 - Oligonucleotídeos utilizados para amplificação da ORF do gene de interesse por meio de PCR.

SÍTIO RESTRIÇÃO	OLIGONUCLEOTÍDEOS
<i>NdeI</i> - CATATG	ACTGTATCATATGAGCGAGAACTCGATTTCGTTTGACCCAATAC
<i>XhoI</i> - TCTGAG	TGCACTCGAGTCATTAACGAATCTCAACCATGGCACGACCGAC

A amplificação do fragmento de interesse foi feita por meio do método de reação em cadeia da polimerase (PCR), a partir do DNA genômico extraído de *E. coli* e dos oligonucleotídeos descritos na TABELA 1, nas condições da TABELA 2, empregando-se o ciclo esquematizado na FIG. 9.

TABELA 2 - Condições da reação de amplificação dos genes.

Reação de amplificação	Volume (μL)
Tampão (20mM Tris-HCl pH 8,4, 50mM KCl)	5
dNTP [10mM] (N= A, G, C, T)	1
MgCl <sub>2</sub> [25mM]	1
Oligonucleotídeo [150 ng/μL] (5' do gene)	1
Oligonucleotídeo [150 ng/μL] (3' do gene)	1
DNA genômico	2
Taq DNA Polimerase (5 U/μL)	1
H <sub>2</sub> O milli-Q	38

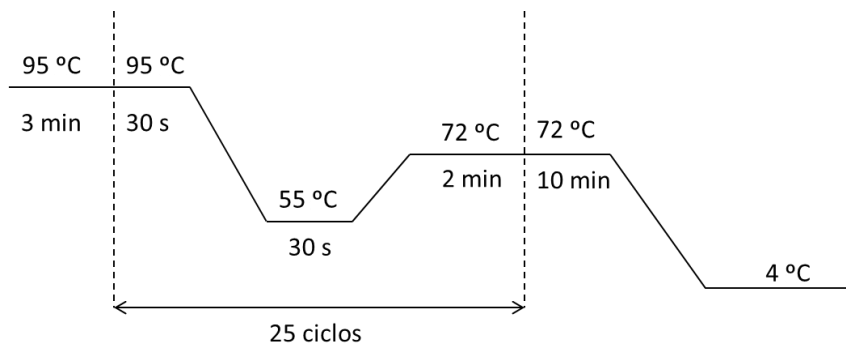


FIGURA 9 - Representação esquemática da condição da reação de amplificação do gene *selD* de *E. coli*.

O resultado foi analisado em gel de agarose preparativo 1 % com brometo de etídio (0,5 µg/mL), em tampão TAE 1X (8 mM Tris-acetato pH 8,0, 0,2 mM EDTA). O fragmento do gene *selD* de *Escherichia coli* amplificado foi então purificado a partir do gel de agarose preparativo utilizando-se o kit *Wizard® SV Gel Clean-Up System* (PROMEGA), de acordo com as condições do fabricante.

### 3.2.3 Clonagem do gene *selD* de *Escherichia coli* em vetor pET28a(+) (Novagen)

O vetor de expressão pET28a+ (Novagen) (FIG. 10) possui uma região promotora T7 localizada anteriormente ao inserto do gene em estudo, além de um sistema de controle de indução através da variação da concentração do indutor isopropil-tio-B-D-galactopiranosídeo (IPTG). Também apresenta uma sequência codificante para um grupo de 6 histidinas (His), de modo que o DNA clonado neste vetor é expresso produzindo uma proteína recombinante em fusão N-terminal com esses aminoácidos. Além disso, entre a sequência codificada pelo gene e a cauda de histidinas, há uma região de reconhecimento da protease Trombina, para posterior clivagem desta cauda.

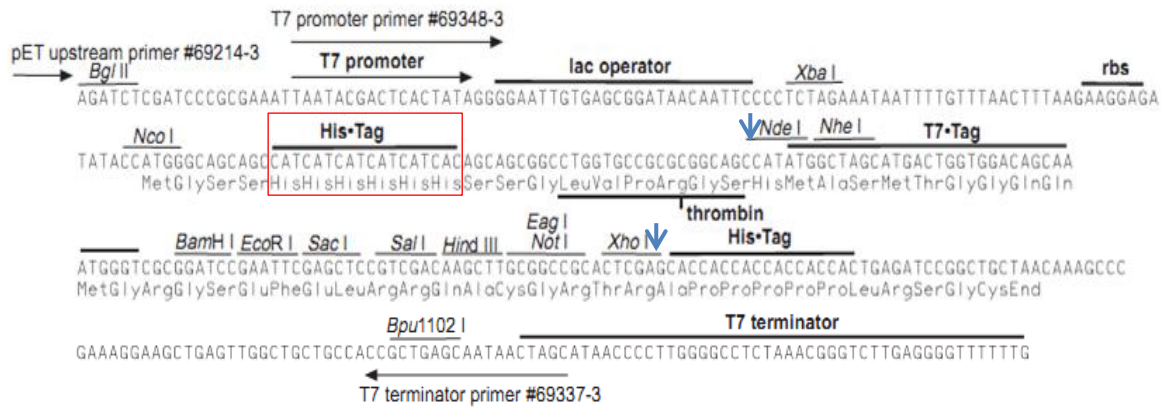


FIGURA 10 - Vetor de expressão pET28a(+). A cauda N-terminal de 6xHis está marcada em vermelho. As setas indicam as regiões reconhecidas pelas enzimas de restrição *NdeI* e *XhoI*, entre as quais foi inserida a ORF *selD* de *E. coli*.

Para a clonagem do gene *selD* de *Escherichia coli*, o produto da reação de PCR purificado do gel de agarose 1%, como descrito na seção 3.2.2, e o vetor pET28a(+) (Novagen) foram digeridos utilizando as endonucleases de restrição *Nde I* e *Xho I* (BioLabs) de acordo com a reação descrita na TABELA 3.

TABELA 3 - Reagentes usados nas reações de digestão do vetor de clonagem e do gene de interesse.

Reação de digestão	Volume (µL)
Produto de PCR	20,0
Enzima <i>Nde I</i> (1 U)	1,0
Enzima <i>Xho I</i> (1 U)	1,0
NE buffer 2 (10X)	5,0
BSA (100 X)	2,0
H <sub>2</sub> O q.s.p	21,0

A reação de digestão foi incubada a 37 °C por 4 horas. Em seguida, os produtos de digestão foram submetidos à eletroforese em gel de agarose e então purificados como descrito na seção 3.2.2. O produto de digestão purificado foi submetido à ligação em vetor pET28a(+) digerido e purificado, seguindo o protocolo de reação descrito na TABELA 4. A reação foi incubada por 16 horas a 16°C. O resultado é designado construção *selD E. coli/pET28a(+)*.

TABELA 4 - Reagentes usados na reação de ligação do fragmento dos genes de interesse em vetor pET28a(+).

Reação de ligação	Volume (µL)
Inserto digerido e purificado	6,0
Vetor pET28a(+) digerido e purificado	2,0
Enzima T4 DNA ligase (20U/µl) (BioLabs)	1,0
Tampão (10x)	1,0

O volume total da mistura de ligação (10 µL) foi utilizado na transformação de células competentes *E. coli* DH5α, preparadas através do método do cloreto de cálcio. As células foram, primeiramente, mantidas em repouso no gelo por 30 minutos. Em seguida foi feito um choque térmico a 42°C por 1 minuto, seguido de incubação no gelo por 2 minutos. Após esta etapa, foram adicionados 250 µL de meio SOC (2 % triptona, 0,5 % de extrato de levedura, 10 mM de NaCl, 2,5 mM de KCl, 10 mM de MgCl<sub>2</sub>, 10 mM de MgSO<sub>4</sub>, 20 mM de glicose, pH 7,4), previamente aquecido a 37 °C. A amostra foi mantida sob agitação por 1 hora a 200 rpm a 37°C. Em seguida, 100 µl de produto foram adicionados a placa de Petri contendo 20 ml de meio LB sólido e antibiótico canamicina em concentração de 30 µg/mL. A placa foi incubada durante 16 horas a 37°C.

Após transformação, foram selecionadas colônias para serem inoculadas a 37 °C por 16 horas em 5 mL de meio LB contendo canamicina (30 µg/mL). Em seguida, procedeu-se com extração do DNA plasmidial através do kit *GeneJET™ Plasmid Miniprep* (Fermentas *Life Sciences*), de acordo com as condições do fabricante.

A confirmação da ligação foi feita pelo teste de digestão, por meio da reação descrita na TABELA 3, mantida por 4 horas a 37°C. Os resultados foram verificados a partir de gel de agarose 1 %.

Os plasmídeos recombinantes também foram submetidos ao sequenciamento dos nucleotídeos pelo método de terminação de cadeia por di-deoxinucleotídeos<sup>82</sup>, utilizando o equipamento *ETDye Terminator* (GE Healthcare), processo executado pela técnica Dra. Susana A. Sculaccio. O resultado do sequenciamento foi analisado pelo programa *SeqMan* pertencente ao software *DNAStar* (Lasergene Sequence Analysis Software – DNAStar).



### 3.2.4 Expressão da proteína SELD de *E. coli*

Células da cepa de *E. coli*, BL21(DE3), transformadas com a construção *selD E. coli/pET28a(+)*, foram inoculadas, inicialmente, em 5 mL de meio de cultura LB contendo canamicina (30 µg/mL) e incubadas por 16 horas sob agitação de 250 rpm a 37°C. Após crescimento das células, inocularam-se 10 mL dessa amostra em 1 L de meio de cultura LB líquido, mantendo sob agitação a 37°C até atingir um valor de densidade óptica de 0,5 absorbância, medido a 600 nm em espectrofotômetro. Após esta etapa, a expressão da proteína foi induzida com 1 mM de IPTG por um período de 3 horas.

Em seguida, as células foram centrifugadas a 12000 g por 15 minutos a 4°C. O *pellet* foi ressuspensionado em 50 mL de Tampão de Lise (50 mM TRIS-HCl pH 8,0, 300 mM NaCl e 10 mM Imidazol) e submetido a ultrassonicação por 6 ciclos, no equipamento *550 Sonic Dismembrator (Fisher Scientific)*. Cada ciclo corresponde a uma sonicação de 30 segundos com intervalo de 1 minuto. As células lisadas foram centrifugadas por 5 minutos a 20000 g a 4°C, de modo a separar a fração solúvel (sobrenadante) e insolúvel (precipitado). Durante a expressão, foram separadas alíquotas de cada etapa, que foram analisadas a SDS-PAGE 15%. O gel foi corado com 0.010% de *Comassie brilliant blue* e descorado com ácido acético 10%. A fração solúvel foi submetida à purificação.

### 3.2.5 Purificação por Cromatografia de Afinidade

A clonagem do gene que expressa a proteína SELD em vetor de expressão pET28a(+) (*Novagen*), que levou à expressão de uma proteína em fusão com uma cauda N-terminal de seis histidinas, possibilitou a utilização da técnica de cromatografia por afinidade a um íon  $\text{Co}^{2+}$  imobilizado (FIG. 11). Assim, em pH 8,0, que corresponde à condição experimental, ocorre a interação reversível entre resíduos do aminoácido histidina da cadeia polipeptídica e o íon metálico  $\text{Co}^{2+}$  imobilizado em coluna *TALON™ Metal Affinity Resin (BD Biosciences)*, sendo possível separar a proteína recombinante em solução.

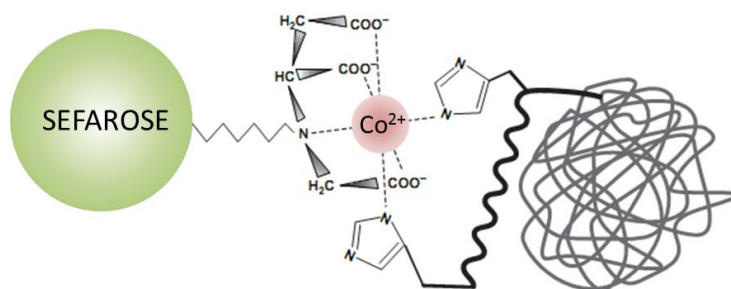


FIGURA 11 - Purificação por afinidade a metal imobilizado. Adaptado do manual da coluna *TALON™ Metal Affinity Resin* (BD Biosciences).

O composto químico imidazol foi utilizado para realizar a eluição da proteína que se liga à resina contendo os íons  $\text{Co}^{2+}$ , já que o anel imidazólico tem maior afinidade química para ligação a esses íons em pH 8,0 e pode competir com os resíduos de histidina pela ligação. No caso da proteína SELD de *E. coli*, a coluna empacotada com 1 mL de resina de cobalto foi previamente lavada com água milli-Q e, em seguida, equilibrada com 10 ml de tampão I contendo 50 mM TRIS-HCl pH 8,0, 300 mM NaCl e 10 mM de Imidazol. O sobrenadante (50 mL) proveniente da lise celular de 1 L de cultura bacteriana descrito na seção 3.2.4 foi aplicado à coluna em alíquotas de 25 mL. Em seguida, fez-se uma lavagem com 30 ml do tampão I adicionado de 10 mM de Imidazol, para retirada das proteínas que não se ligaram à resina. A proteína foi eluída da coluna através de aplicação de 30 ml do tampão I adicionado de 20 mM de Imidazol, seguida de aplicação de 10 mL de tampão I com 200mM de Imidazol e, finalmente, 10 mL do mesmo tampão com 500 mM de Imidazol.

### 3.2.6 Diálise

As amostras eluídas na purificação por afinidade de ambas as proteínas recombinantes (~ 40 ml) foram submetidas ao processo de diálise contra um tampão contendo 50 mM TRIS-HCl pH 8,0, 300 mM NaCl e 5 mM DTT. O composto químico DTT atua impedindo a formação de ligações dissulfeto, evitando agregação das proteínas. A diálise foi realizada a 4 °C e foram feitas 3 trocas do mesmo tampão (~ 1 L), de 3 em 3 horas. A amostra de proteína foi submetida à concentração por ultra-filtração em membrana *Millipore*, sob rotação de 2000 g a 4 °C. A concentração de proteína das amostras resultantes foi medida utilizando-se o

espectrofotômetro *Nanodrop<sup>TM</sup> 1000 (Thermo Scientific)*, no qual é possível aplicar apenas 2  $\mu\text{L}$  de amostra de proteína para leitura de absorbância,  $A$ , a 280 nm e cálculo da concentração proteica,  $C$ , dado o coeficiente de extinção molar teórico,  $\epsilon$ , para a proteína (seção 3.1.1), por meio da Lei de Beer-Lambert <sup>83</sup>:

$$A(\lambda) = \log(I/I_0) = C\epsilon(\lambda)l \quad (1)$$

onde  $l$  é o caminho óptico percorrido pela luz, que no caso é igual a 1 mm.

### 3.2.7 Clivagem da cauda de histidinas com Trombina

As amostras concentradas foram submetidas à clivagem da cauda de histidinas utilizando a protease Trombina (*Amersham Biosciences*), que foi adicionada na proporção de 1 unidade de protease para 100  $\mu\text{g}$  de proteína recombinante. A reação foi mantida por 14 horas a 20°C e interrompida com a adição de 1 mM de PMSF, que atua como inibidor de proteases. O resultado foi verificado por SDS-PAGE 15%.

### 3.2.8 Purificação por Cromatografia de Exclusão Molecular

Na cromatografia por exclusão molecular, as moléculas são separadas de acordo com diferenças em seus tamanhos e formas moleculares. A matriz da coluna é um polímero com poros de tamanhos selecionados. Proteínas com maior massa molecular migram mais rápido, pois tendem a passar direto pela matriz sem penetrar nos poros da resina, enquanto que moléculas com menor massa molecular tendem a penetrar nos poros da resina e migram mais vagarosamente através da coluna <sup>84</sup>.

Nesta etapa de purificação, foi utilizada uma coluna *Superdex 200 HR 10/300 (GE)* acoplada ao sistema de cromatografia *AKTA Explorer10 (GE)*. A amostra de proteína pura, após diálise e clivagem da cauda de histidinas, com uma concentração aproximada de 2,0 mg/mL (54,5  $\mu\text{M}$ , em 500  $\mu\text{L}$ ), foi aplicada na coluna previamente equilibrada com

tampão 50 mM TRIS-HCl pH 8,0, 300 mM NaCl e 5 mM DTT. O fluxo de corrida foi de 0,5 mL/min e o andamento da purificação foi monitorado a 280 nm por espectrofotometria. As frações eluídas foram analisadas em gel 15% SDS-PAGE e as amostras resultantes foram concentradas por ultrafiltração.

### 3.3 Caracterização biofísica da proteína recombinante SELD de *E. coli*

A caracterização biofísica da proteína recombinante SELD de *E. coli* foi feita empregando-se espectroscopia de dicroísmo circular para investigação de sua estrutura secundária, além de eletroforese em gel nativo, espalhamento dinâmico de luz, espalhamento de raios-X a baixo ângulo e ultracentrifugação analítica para estudo de suas formas oligoméricas em solução e suas propriedades hidrodinâmicas. Os experimentos empregados foram realizados em colaboração com aluna de doutorado Livia Maria Faim, paralelamente ao seu trabalho com as Selenofosfato Sintetases de *Trypanosoma brucei* e *Leishmania major*, e encontram-se detalhados nas subseções seguintes.

#### 3.3.1 Espectroscopia de Dicroísmo Circular

A Espectroscopia de Dicroísmo Circular (CD) é uma técnica que se baseia na interação diferenciada de moléculas assimétricas com feixes de luz circularmente polarizada à esquerda e à direita. Estes feixes de luz circularmente polarizada, quando atingem moléculas opticamente ativas, são absorvidos diferentemente, devido aos diferentes coeficientes de extinção da molécula para cada polarização. Após atravessar a molécula, cada componente do feixe de luz continua circularmente polarizado, porém os raios das circunferências traçadas pelo vetor campo elétrico de cada componente são diferentes. Combinando-as, tem-se como resultado um feixe de luz elipticamente polarizada, já que as duas componentes têm amplitudes diferentes<sup>83</sup>, como se observa na FIG. 12.

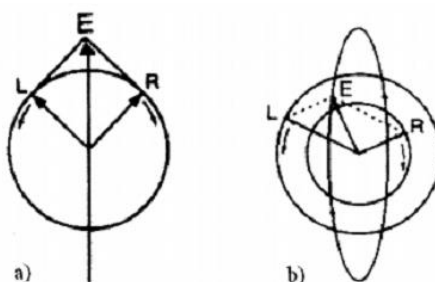


FIGURA 12 - Decomposição do feixe de luz plano polarizado em duas componentes circulares polarizadas à esquerda (L) e à direita (R), a) antes e b) depois de passar por amostra opticamente ativa. Adaptado <sup>83</sup>.

O Dicroísmo Circular (CD) pode ser definido como a diferença entre a absorção das componentes circularmente polarizadas à esquerda e à direita ( $\Delta A$ ) <sup>83</sup>, com A dada pela equação 1. A elipticidade ( $\theta$ ) corresponde ao ângulo cujo arco é igual à razão entre os semi-eixos maior e menor da elipse. Este valor está relacionado com  $\Delta A$  da seguinte forma <sup>83</sup>:

$$\theta(\text{graus}) = (180)(\ln 10)(\Delta A/4\pi) \quad (2)$$

Pode-se definir, ainda, a elipticidade molar, como sendo <sup>83</sup>

$$[\theta] = \frac{\theta}{c l} \quad (3)$$

dada a concentração C da amostra e o caminho óptico l.

A maioria das macromoléculas biológicas são opticamente ativas e a atividade óptica está intimamente relacionada à sua conformação. Como as proteínas estão repletas de centros quirais (como o carbono  $\alpha$ , de todos os aminoácidos exceto a glicina, e o carbono  $\beta$  da treonina e isoleucina) distribuídos ao longo das suas estruturas secundárias ( $\alpha$ -hélices, folhas- $\beta$ , voltas- $\beta$ ), estas estruturas podem ser identificadas e quantificadas por meio da técnica de CD. Espectros de CD de proteínas exibem bandas na região de 250-260 nm e 180-250 nm que correspondem a ligações dissulfeto e ligações peptídicas, respectivamente <sup>83</sup>. A FIG. 13 mostra os diferentes espectros de CD descritos para as diferentes estruturas secundárias de uma proteína.

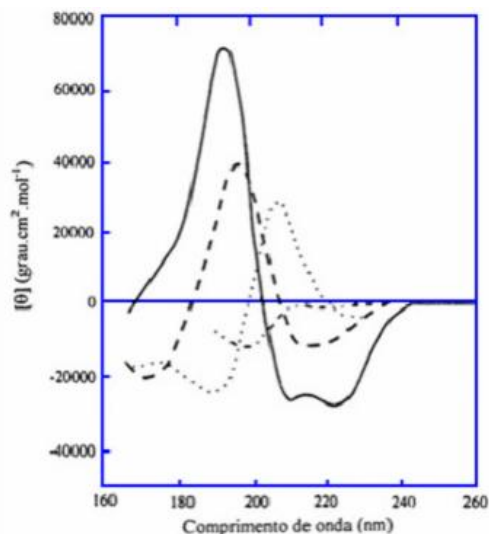


FIGURA 13 - Espectro de CD no UV distante. ( — )  $\alpha$ -hélice; ( - - - ) folha- $\beta$  antiparalela; ( . . . ) volta- $\beta$ , ( - . - . ) estrutura irregular. Adaptado <sup>83</sup>.

A partir do espectro de CD da proteína em estudo é possível estimar a porcentagem de cada estrutura secundária presente, já que estas possuem espectro de CD característico. Considerando-se o espectro de uma proteína,  $[\theta(\lambda)]$ , como combinação linear dos espectros de estruturas secundárias tem-se <sup>83</sup>:

$$[\theta(\lambda)] = \chi_{\alpha}[\theta_{\alpha}(\lambda)] + \chi_{\beta}[\theta_{\beta}(\lambda)] + \chi_t[\theta_t(\lambda)] + \chi_r[\theta_r(\lambda)] \quad (4)$$

onde  $\theta_i$ , com  $i = \alpha, \beta, t$  e  $r$ , designam os espectros característicos de estruturas de hélices-  $\alpha$ , folhas-  $\beta$ , voltas e estruturas desordenadas, respectivamente. A deconvolução do espectro de CD de uma proteína consiste em se obter os coeficientes  $\chi_{\alpha}, \chi_{\beta}, \chi_t$  e  $\chi_r$ , que correspondem às porcentagens de estrutura secundária na estrutura, a partir da combinação linear de espectros de estruturas secundárias de base de dados de espectros de CD para biomoléculas <sup>83</sup>.

O espectro de Dicroísmo Circular da proteína SELD de *E. coli* foi obtido utilizando um espectropolarímetro *JASCO J-810* (JASCO International Co. Ltd), equipado com um sistema de controle de temperatura do tipo *PELTIER PTC 423S/15*, do Laboratório de Espectrometria e Calorimetria (LEC) do LNBio. O espectro foi coletado na região do UV distante, no intervalo de comprimento de onda de 190 a 250 nm, com resolução de 0,2 nm, utilizando uma cubeta de quartzo de caminho óptico de 0,1 cm para amostra de proteína em concentração de 0,20 mg/ml (5,45  $\mu$ M). A sensibilidade foi mantida em 100 miligraus e o tempo de resposta em 0,5 segundo. O programa *Spectra Manager* (JASCO) foi usado para registro dos dados. A deconvolução do espectro de Dicroísmo Circular foi feita usando o programa *CDNN* <sup>85</sup>.

### 3.3.2 Estimativa da massa molecular e raio hidrodinâmico por Cromatografia de Exclusão Molecular

Para se determinar o valor da massa molecular e raio hidrodinâmico (raio de Stokes) da proteína em seu estado oligomérico predominante em solução e, possivelmente, confirmar seu valor teórico, proteínas-padrão com diferentes massas moleculares foram analisadas nas mesmas condições da cromatografia por exclusão molecular realizada com a proteína SELD de *E. coli*. A partir do volume de eluição de cada padrão e conhecendo-se sua massa molecular e raio hidrodinâmico, foi possível obter curvas de calibração através do programa *ORIGIN 8.0 (ORIGINLab)*.

### 3.3.3 Eletroforese em Gel Nativo

A eletroforese em gel nativo é uma técnica que permite determinar a homogeneidade da amostra de proteína, bem como seu estado oligomérico em solução. O experimento com a proteína recombinante SELD de *E. coli* foi realizado sob condições padrão, utilizando um gel com gradiente *PhastGel 8-25 (GE)* em tampão 0,112 M Tris e 0,112M de acetato pH 6,4. Foram aplicados 4 µL de amostra de proteína a 2,0 mg/ml (54,5 µM). O gel foi corado com *Coomassie Blue G-50* 0,1% em solução aquosa de metanol 25% e ácido acético 5%. Um conjunto de proteínas de massa molecular e raio hidrodinâmico conhecido foi usado como padrão. A partir da distância de migração de bandas no gel, foi possível estimar a massa molecular e o raio hidrodinâmico correspondente a cada banda a partir de curvas de calibração.

### 3.3.4 Espalhamento Dinâmico de Luz (DLS)

O Espalhamento Dinâmico de Luz é uma técnica capaz de analisar biomoléculas de interesse em solução. Na prática, as macromoléculas encontram-se hidratadas e o raio

calculado a partir do coeficiente de difusão oferece uma estimativa do tamanho aparente da partícula com sua camada de solvatação, denominado assim, raio hidrodinâmico,  $R_H$ , a partir do qual também é possível estimar sua massa molecular <sup>83</sup>.

Como as partículas estão em Movimento Browniano em solução, é possível associar um coeficiente de difusão translacional para seu movimento em solução, dado pela equação de Einstein <sup>83</sup>:

$$D = \frac{RT}{N_A f} \quad (5)$$

que depende do coeficiente friccional,  $f$ , da partícula com a solução, da temperatura  $T$ , da constante universal dos gases  $R$  e do número de Avogadro  $N_A$ . O coeficiente friccional, por sua vez, está relacionado com a viscosidade da solução,  $\eta$ , através da relação de Stokes <sup>83</sup>:

$$f = 6\pi\eta R_H \quad (6)$$

o que indica que quanto maior for a partícula, maior o atrito sentido para seu movimento translacional, já que também depende do seu raio hidrodinâmico,  $R_H$ .

Uma correção do valor experimental  $D$  para o estado padrão da água à 20°C é necessária para análise comparativa de dados obtidos em diferentes condições experimentais em diferentes laboratórios e é dada por:

$$D_{20,w} = D_{obs} \frac{293,2}{T} \frac{\eta_{T,w}}{\eta_{20,w}} \frac{\eta_s}{\eta_w} \quad (7)$$

onde  $T$  designa as condições experimentais de temperatura,  $\eta_s$  e  $\eta_w$  são as viscosidades do solvente e da água, respectivamente, calculadas utilizando o programa SEDNTERP <sup>95</sup>, e o índice  $20, w$  indica as condições padrões  $T = 20$  °C em água.

A razão do valor máximo de  $D$  ( $D_{esfera}$ , coeficiente de difusão para uma esfera de raio igual a  $R_H$ ) para o valor de  $D$  observado,  $D_{esfera} / D_{20,w}$ , é igual à razão do coeficiente friccional experimental para o coeficiente friccional mínimo ( $f_0$ , calculado para uma esfera de raio  $R_H$ ). Esta razão,  $f / f_0$ , é conhecida como fator de Perrin, a qual mede a forma assimétrica máxima da partícula a partir de uma esfera <sup>83</sup>.

O princípio da técnica de DLS se baseia na incidência de luz monocromática em solução de partículas e monitoramento das flutuações da intensidade de luz espalhada com o tempo. Como as partículas estão em movimento em solução, há uma forte correlação entre o padrão de espalhamento de luz em um dado instante e em um instante muito próximo, da



ordem de nano ou microssegundos. A correlação é reduzida comparando-se intervalos de tempo maiores <sup>83</sup>. A função de autocorrelação temporal é dada por <sup>86</sup>

$$G(t) = \langle I(t_0) \rangle \langle I(t_0 + \delta t) \rangle \quad (8)$$

Sendo  $I(t_0)$  a intensidade de espalhamento em um instante e  $I(t_0 + \delta t)$  a intensidade de espalhamento em um instante imediatamente depois. Para partículas esféricas e monodispersas, a função  $G(t)$  é dada por <sup>86</sup>

$$G(t) = A e^{-\Gamma t} + B \quad (9)$$

com A e B constantes e  $\Gamma$  a constante de decaimento da curva exponencial, sendo

$$\Gamma = Dq^2 \quad (10)$$

onde  $q$  é o vetor de onda de luz espalhada dada por

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad (11)$$

onde  $\lambda$  é o comprimento de onda da luz espalhada e  $\theta$  é o ângulo de espalhamento de luz.

O resultado do experimento, portanto, é a obtenção do valor de  $\Gamma$  a partir da função de autocorrelação e, então, o cálculo de D e, conseqüentemente, de  $R_H$ . A estimação de massa molecular da proteína considerando-a como esférica é possível, deste modo, conhecendo-se também sua densidade. Assim, podemos estimar seu estado oligomérico em solução <sup>86</sup>.

Para tanto, utilizou-se o equipamento *ProteinSolutions DynaPro*<sup>TM</sup> do LEC-LNBio. A amostra resultante da purificação por cromatografia por exclusão molecular, posteriormente submetida à concentração por ultrafiltração, foi centrifugada por 10 minutos a 10000 g e, então, aplicada em cubeta de quartzo para realização de várias medidas DLS, a diferentes concentrações. O software *DYNAMICS*<sup>TM</sup> seleciona as medidas mais representativas e calcula seu valor médio.

### 3.3.5 Espalhamento de Raios-X a Baixo Ângulo (SAXS)

Em um experimento de SAXS, a amostra de proteínas em solução é iluminada com feixe monocromático de raios-X e os átomos que compõem a amostra interagem com a radiação incidente. Como a energia dos fótons incidentes é bem menor que a de ligação dos elétrons nos átomos, os elétrons se comportam como se estivessem livres. Partículas carregadas livres e oscilantes produzem ondas eletromagnéticas em todas as direções, coerentes com a onda incidente e com o mesmo comprimento de onda, no caso de espalhamento elástico. A intensidade da radiação espalhada depende do ângulo de espalhamento ( $2\theta$ )<sup>87</sup>.

A principal diferença entre a técnica de cristalografia de raios-X e SAXS ocorre devido ao fato da estrutura cristalina apresentar um arranjo periódico nas três dimensões com um grande número de repetições, enquanto que em solução as partículas não estão ordenadas, mas sim em movimento browniano<sup>87</sup>. Há, portanto, uma perda de informação em experimentos de espalhamento de raios-X a baixo ângulo, devido à necessidade de se tomar a média sobre todas as orientações no espaço.

Mesmo assim, em sistemas monodispersos e polidispersos é possível a determinação da dimensão ( $D_{\text{máx}}$ ) das proteínas analisadas, bem como do seu raio de giro ( $R_g$ ), que corresponde à distância quadrática média dos elétrons em relação ao centro de massa da molécula, e da sua forma em solução.

Os dados coletados do espalhamento de raios-X em função do ângulo incidente podem ser usados para o cálculo da função de distribuição de distâncias,  $p(r)$ , que inclui toda a informação estrutural possível de ser obtida. Com esta função, é possível obter um *modelo ab initio* do envelope (forma) da proteína. Deste modo, é possível compará-lo ao espalhamento que seria obtido com uma proteína homóloga, cuja estrutura cristalográfica é conhecida<sup>87</sup>.

Neste trabalho, os resultados experimentais de espalhamento de raios-X a baixo ângulo nas amostras de SELD de *E. coli* em concentração de 1,5 mg/ml (40,9  $\mu\text{M}$ ) em tampão 25 mM TRIS-HCl pH 8,0 e 50 mM NaCl, na presença de 5 mM DTT, foram coletados em colaboração com o doutor Mario de Oliveira Neto do Grupo de Cristalografia de Proteínas do IFSC-USP. Os dados de SAXS foram coletados na linha SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP), utilizando um comprimento de onda  $\lambda = 1,488 \text{ \AA}$ , em detector bidimensional disposto a 1068,97 mm da amostra.

A análise sistemática dos resultados experimentais foi feita, inicialmente, através do programa FIT2D<sup>88</sup>, normalizando-se os dados pela intensidade do feixe e multiplicando-se pela absorção da amostra. Subtraiu-se o espalhamento da solução tampão foi subtraído da curva de espalhamento da solução de proteína. O raio de giro da proteína foi obtido utilizando-se a aproximação de Guinier<sup>87</sup>:

$$I(q) = I(0)\exp\left(-\frac{R_g^2 q^2}{2}\right) \quad (12)$$

onde  $q$  é o módulo do vetor de espalhamento ( $q = 4 \pi \sin(\theta/\lambda)$ ),  $R_g$  é o raio de giro e  $I(0)$  é a intensidade no ângulo zero. Esta expressão é válida tipicamente para  $qR_g < 1,3$ . O valor de  $I(0)$  normalizado para a concentração de proteína é proporcional à massa molecular da proteína e foi utilizado pelo programa *SAXS MoW*<sup>89</sup> para estimação da massa molecular da proteína em solução. Adicionalmente, a distribuição de distâncias foi calculada pelo método de Transformada Inversa de Fourier, onde<sup>87</sup>

$$p(r) = \frac{r}{2\pi^2} \int_0^\infty q I(q) \sin(rq) dq \quad (13)$$

Este método está implementado no programa *GNOM*<sup>90</sup>, onde foi possível obter a curva de distribuição de distâncias  $p(r)$  e a  $D_{\text{máx}}$ .

O programa *CRYSOL*<sup>91</sup> foi utilizado para a comparação deste resultado com estruturas de SELD homólogas já conhecidas e também com o modelo da estrutura tridimensional da SELD de *E. coli* obtido como descrito na seção 3.1.2.

Modelos de baixa resolução foram gerados pelo método *Ab Initio* através do programa *DAMMIN*<sup>92</sup>. Arrefecimento simulado é usado na procura por um modelo compacto que concorde bem com os dados experimentais, através de minimização de uma função do parâmetro discrepância,  $\chi$ , que leva em consideração a intensidade de luz espalhada e aquela calculada do modelo gerado. Foram realizadas 10 reconstruções *ab initio* independentes com o programa *DAMMIN*<sup>92</sup>, que foram promediadas através do programa *DAMVER*<sup>93</sup>.

### 3.3.6 Ultracentrifugação Analítica

A Ultracentrifugação Analítica (AUC) é uma técnica biofísica que permite o monitoramento da sedimentação de macromoléculas em solução em função do tempo, quando submetidas a um campo centrífugo uniforme. Deste modo, permite a determinação de parâmetros hidrodinâmicos e termodinâmicos de proteínas em solução, através de dois métodos: Velocidade de Sedimentação (SV, em inglês) e Sedimentação em Equilíbrio (SE, em inglês) <sup>94</sup>.

Quando uma partícula é submetida à centrifugação sob velocidade de rotação constante,  $\omega$ , esta é forçada para o fundo do tubo, sob certa velocidade de sedimentação,  $u$ . A relação entre a velocidade de rotação e a velocidade de sedimentação é conhecida como constante de sedimentação ou constante de Svedberg <sup>83</sup> e é dada por:

$$S = \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N_A f} = \frac{MD(1-\bar{v}\rho)}{RT} \quad (14)$$

onde  $r$  é a posição radial,  $M$  é a massa molar,  $\bar{v}$  é o volume parcial específico,  $\rho$  é a densidade do solvente,  $N_A$  é o número de Avogadro,  $f$  é o coeficiente friccional,  $D$  é o coeficiente de difusão translacional e  $R$  é a constante universal dos gases. O máximo valor de coeficiente de sedimentação é conhecido por  $s_{esfera}$  e é dado por <sup>83</sup>:

$$s_{esfera} = \frac{M(1-\bar{v}\rho)}{N_A 6\pi\eta \left( \frac{3M\bar{v}}{4\pi N_A} \right)^{1/3}} \quad (15)$$

com  $R_0 = \left( \frac{3M_w\bar{v}}{4\pi N_A} \right)^{1/3}$ , obtidos rearranjando-se as equações 5, 6 e 14.

No caso da SELD de *E. coli*, o coeficiente de difusão foi obtido utilizando-se a técnica de DLS, enquanto o volume parcial específico da proteína e a densidade do solvente foram calculados através do programa SEDNTERP <sup>95</sup>. Para todas as amostras o tampão utilizado foi TRIS-HCl 25 mM pH 8,0, NaCl 50 mM,  $\beta$ -Mercaptoetanol 1mM, sendo  $\bar{v} = 0,7397$  ml/g,  $\rho = 1,00101$  g/mL e  $\eta = 0,01014$  poise. O coeficiente de sedimentação máximo ( $s_{esfera}$ ) e o coeficiente de difusão máximo ( $D_{esfera}$ ) para diferentes formas oligoméricas da SELD de *E.*

*coli* considerando-a como globular foram calculados por meio do mesmo programa e são apresentados na TABELA 5.

TABELA 5 - Parâmetros hidrodinâmicos da SELD de *E. coli*.

SELD de <i>E. coli</i>	MM (kDa)	$s_{\text{esfera}} (10^{-13} \text{ S})$	$D_{\text{esfera}} (\times 10^{-7} \text{ cm}^2/\text{s})$
Monômero	36,97	3,8977	9,6948
Dímero	73,94	6,1713	7,6947
Trímero	110,91	8,0866	6,7220
Tetrâmero	147,87	9,7963	6,1073
Octâmero	295,75	15,5510	4,8474

\* Neste caso, foram incluídos os aminoácidos resultantes da clivagem com a protease Trombina.

Uma correção do valor experimental de  $s$  para o estado padrão da água à 20°C é necessária para análise comparativa de dados obtidos em diferentes condições experimentais e diferentes laboratórios e é dada por :

$$s_{20,w} = s_{T,B} \left( \frac{\eta_{T,B}}{\eta_{20,w}} \right) \frac{(1 - \bar{v}\rho)_{20,w}}{(1 - \bar{v}\rho)_{T,B}} \quad (16)$$

onde T e B designam as condições experimentais de temperatura e tampão, e o índice 20, w indica condições padrões.

Deve-se mencionar que a razão do valor máximo de  $s$  para o valor de  $s$  observado,  $s_{\text{esfera}} / s_{20,w}$ , é igual ao fator de Perrin, o qual mede a forma assimétrica máxima a partir de uma esfera.

Rearranjando a equação 14, temos a relação entre a massa molecular da partícula, o coeficiente de sedimentação e o coeficiente de difusão translacional<sup>83</sup>:

$$M = \frac{sRT}{D(1 - \bar{v}\rho)} \quad (17)$$

através da qual é possível se obter a massa da partícula, dados  $s$  e  $D$  experimentais.

Todos os processos que ocorrem durante um experimento de ultracentrifugação analítica podem ser descritos pela equação de Lamm<sup>83</sup>:

$$\frac{\partial c}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left[ rD \frac{\partial c}{\partial t} - s\omega^2 r^2 c(r,t) \right] \quad (18)$$

que descreve a evolução no tempo da distribuição de concentração,  $c(r)$ , da partícula em estudo, dados  $s$  e  $D$ , sob uma velocidade de rotação constante.

O experimento de velocidade de sedimentação (SV) (FIG. 12A) é baseado em propriedades termodinâmicas e usa alta velocidade de rotação para formar um gradiente de concentração de partículas em uma célula de amostra, que é medido por absorção de luz UV. À medida que a proteína sedimenta com o tempo, a migração da partícula em estudo é medida através da comparação com uma célula de referência, em que se encontra somente o tampão da amostra<sup>96</sup>. Deste modo, a partir das curvas de absorbância versus posição (raio) na célula, obtidos em diferentes instantes de tempo, é possível calcular uma distribuição de coeficiente de sedimentação  $s$ ,  $c(s)$ , utilizando o programa SEDFIT<sup>97</sup>. Neste trabalho, os experimentos de SV foram realizados no LEC/LNBio, para amostras de SELD de *E. coli* nas concentrações 0,15, 0,30, 0,45, 0,60, 0,80 e 1,00 mg/ml (4,09, 9,80, 12,27, 16,36, 21,83 e 27,27  $\mu\text{M}$ , respectivamente), sob velocidade de rotação de 30000 rpm, à 20 °C em ultracentrifuga analítica *Beckman Optima XL*, equipada com rotor *AN-60Ti* (*Beckman Coulter*), em colaboração com o professor Dr. Júlio Cesar Borges do Grupo de Biologia Molecular e Bioquímica do IQSC-USP.

O experimento de Sedimentação em Equilíbrio (SE) (FIG. 12B), por sua vez, baseia-se em princípios termodinâmicos, sendo sensível à massa da partícula, mas não ao seu formato. O experimento é feito em velocidades de rotação baixas o suficiente para que as forças de sedimentação e difusão alcancem um equilíbrio<sup>96</sup>. No estado estacionário, é possível derivar a constante de associação entre diferentes formas da proteína em solução, analisando diferentes modelos de associação da mesma, pelo método *Self-Association*, do programa SEDPHAT<sup>98</sup>. Neste trabalho, os experimentos de SE foram realizados, também, no LEC/LNBio, com amostras de SELD de *E. coli* em concentrações de 0,50, 0,75 e 1,00 mg/ml (13,63, 20,45 e 27,27  $\mu\text{M}$ , respectivamente), sob velocidades de rotação de 8000, 10000 e 12000 rpm, também em colaboração com o professor Dr. Júlio Cesar Borges.

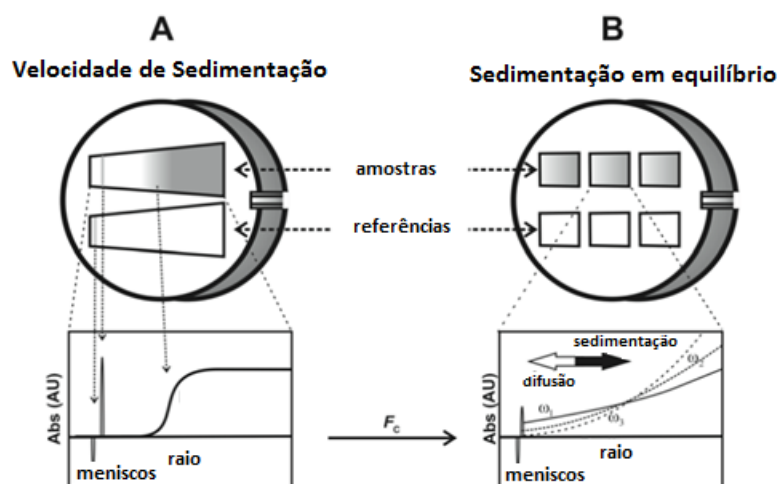


FIGURA 14 - Experimentos de Ultracentrifugação Analítica. A) SV e B) SE. Adaptado <sup>96</sup>.

### 3.3.7 Ensaios de Cristalização

A elucidação da estrutura tridimensional de proteínas pelo método cristalográfico está condicionada à obtenção de cristais de tamanho e qualidade que possibilitem a coleta de dados de difração de raios-X. A área de cristalografia de proteínas tem tido um grande avanço ao longo das últimas décadas, porém a taxa de sucesso na cristalização de proteínas ainda continua baixa. Na média, apenas de 15 a 20 % das proteínas alvo de estudo que podem ser purificadas cristalizam de forma que possam ter sua estrutura resolvida por difração de raios-X <sup>99</sup>.

O crescimento de um cristal de proteína inicia-se a partir de uma solução supersaturada e prossegue até um estado termodinamicamente estável no qual a proteína está em fase sólida na forma de cristal. O tempo para se atingir o equilíbrio tem grande influência no resultado final, que pode ser desde um precipitado amorfo ou micro-cristais até um cristal único e grande. Condições de supersaturação podem ser obtidas através da adição de agentes precipitantes, como sais, solventes orgânicos e polietilenoglicóis (PEG), ou modificando parâmetros físico-químicos da solução, tais como pH e temperatura <sup>83</sup>.

O procedimento mais comum utilizado para cristalização de proteínas é o método de difusão de vapor, que se baseia em um processo de equilíbrio entre duas soluções (uma gota e um reservatório) através da fase de vapor, em ambiente fechado. A gota possui uma solução menos concentrada do que o reservatório e tende a equilibrar a sua pressão de vapor com a

pressão de vapor da solução. Ocorre, então, a difusão de compostos voláteis da gota para o reservatório, tornando a gota concentrada até um nível ótimo para cristalização da proteína. Desde que o sistema entre em equilíbrio, esta condição é mantida até que a cristalização seja completada. Entre os métodos que utilizam este princípio estão o método da gota pendurada (*hanging drop*) e o método da gota apoiada (*sitting drop*), que se diferem na orientação da gota de solução proteica no sistema<sup>100</sup>, como mostrado na FIG. 15.

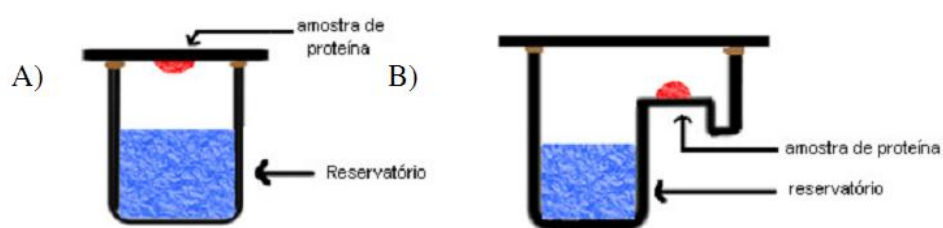


FIGURA 15 - Método de Difusão de Vapor: A) gota pendurada e B) gota apoiada.

A triagem inicial de condições de cristalização da SELD de *E. coli* em concentrações de 5,0, 6,0 e 7,0 mg/ml (136,4, 163,6 e 190,9  $\mu$ M, respectivamente) foi feita em placas de cristalização com 96 poços a 20°C, através do método de difusão de vapor. Foram utilizados os kits comerciais *Classics Suite*, *Classics II Suite*, *PEG Suite*, *PEG II Suite* (QIAGEN), *Index HT* e *Salt RX* (Hampton Research), em um total de 576 condições de cristalização diferentes. As gotas de cristalização foram feitas em robô de cristalização *Honeybee 939* (Molecular Dynamics), contendo 0,7  $\mu$ l de solução de proteína (25 mM TRIS/HCl pH 8.0, 25 mM NaCl, 2 mM DTT, 5 mM  $MgCl_2$ , na presença e na ausência de 5 mM ADP, 5 mM ATP ou 5 mM AMPcPP) e 0,7  $\mu$ l de solução precipitante, utilizando o método da gota apoiada. Considerando-se as diferentes condições da amostra de proteína, foram analisadas mais de 2000 condições de cristalização diferentes.

As técnicas de cristalização de proteína, que em geral focam na otimização de estratégias de cristalização, estão se voltando para o melhoramento das propriedades das proteínas sujeitas à cristalização, tendência que começou na década de 1990, com o desenvolvimento de técnicas de Biologia Molecular e Espectrometria de Massas. Deste modo, cientistas puderam focar os esforços de cristalização em domínios estáveis das proteínas-alvo, identificados, por exemplo, por proteólise limitada<sup>99</sup>.

Ensaio de proteólise *in-situ* foram realizados para otimizar a cristalização da SELD de *E. coli*, considerando-se que a proteína apresenta um N-terminal bastante flexível, o que



pode dificultar a formação de contatos cristalinos e, conseqüentemente, levar à não cristalização. Primeiramente, testes de proteólise com uma variedade de proteases foram realizados para identificar uma protease capaz de gerar um padrão de degradação estável, detectado por SDS-PAGE. Como a protease Quimiotripsina apresentou um produto estável, esta foi adicionada aos ensaios de cristalização na concentração de 250 ug/ml e mantida por 20 minutos. Após este tempo, foi adicionado inibidor de protease PMSF (1 mM) na reação.

O produto de proteólise com Quimiotripsina também foi purificado por cromatografia de exclusão molecular em coluna *SUPERDEX 75 10/300* (GE) e a amostra de proteína purificada foi submetida às mesmas condições de cristalização. As condições que levaram à formação de cristais foram otimizadas por variação de pH e concentração de precipitante. Cristais selecionados foram testados para difração de raios-X na linha MX2 do Laboratório Nacional de Luz Síncrotron em Campinas-SP.

### **3.4 Estudo da interação da SELD de *E. coli* com o complexo SELA-SELC**

#### **3.4.1 Anisotropia de Fluorescência**

A medida de anisotropia de fluorescência de uma amostra é ideal para determinação da associação de proteínas e formação de complexos de macromoléculas. Isso porque a anisotropia varia em resposta à mudança na taxa de rotação de uma molécula em solução, que depende do tamanho e da forma da partícula, além da viscosidade do solvente. Além disso, é uma medida independente da concentração de proteína em solução<sup>101</sup>.

Deste modo, a amostra de SELD de *E. coli* foi titulada em solução contendo 5  $\mu$ M do complexo SELA-SELC, na proporção 1:1 em mols de monômeros, sendo 10 nM de tRNA<sup>Sec</sup> marcado com fluoresceína. Esse experimento foi realizado em colaboração com o aluno de mestrado Vitor Hugo Balasco Serrão e a aluna de doutorado Lívia Regina Manzine, ambos do Grupo de Cristalografia de Proteínas do IFSC-USP. O procedimento de marcação do SELC com fluoresceína e sua interação com SELA foi padronizado pela doutoranda Lívia Regina Manzine (dados não publicados).

As medidas de anisotropia de fluorescência foram realizadas a 25 °C em espectrofluorímetro *ISS-PCI* (*ISS Champaign, IL*), em geometria L. O comprimento de onda de excitação foi de 495 nm e um filtro de *cutoff* de 515 nm (corte de 50% da intensidade em 515 nm). Os valores de anisotropia ( $r$ ) foram calculados pelo programa *Vinci-ISS* (*ISS Champaign, IL, EUA*) obedecendo à relação <sup>101</sup>:

$$r = \frac{[I_{VV} - GI_{VH}]}{[I_{VV} + 2GI_{VH}]} \quad (19)$$

em que  $I$  refere-se à intensidade de fluorescência medida e os índices indicam a orientação dos polarizadores durante a excitação e emissão, nesta ordem, sendo  $V$  correspondente à orientação vertical e  $H$  à horizontal, e  $G$  é um fator de correção que considera diferenças da sensibilidade nas duas direções e é dado por <sup>101</sup>:

$$G = \frac{I_{HV}}{I_{HH}} \quad (20)$$

Para cada concentração de proteína titulada, a medida de anisotropia corresponde à média de pelo menos cinco medidas após estabilização de 10 minutos à temperatura do experimento. A mudança na anisotropia de fluorescência resultante foi usada para estimar a estequiometria de ligação de SELD no complexo SELA-SELC de *E. coli*.

### 3.4.2 Espectrometria de massas com troca H/D

A troca isotópica por deutério do hidrogênio das amidas da cadeia principal de uma proteína que estão expostas ao solvente (troca H/D), combinada com a espectrometria de massas (MS), é uma técnica promissora para determinação de mudanças na conformação de uma proteína após ligação a um ligante, sendo capaz de fornecer informações locais da interação entre as estruturas envolvidas, mostrando-se vantajosa em comparação com a cristalografia de raios-X ou RMN por ser rápida, bastante sensível e não requerer muita amostra <sup>102</sup>.

A análise das regiões da SELD de *E. coli* que interagem com o complexo SELA-SELC foi investigada por meio desta técnica, em colaboração com o aluno de mestrado Vitor Hugo Balasco Serrão do Grupo de Cristalografia de Proteínas do IFSC-USP, com o Dr. Daniel Menezes Saidemberg e o professor Dr. Mario Sergio Palma, ambos do Instituto de Biociências da UNESP - Rio Claro. A preparação da amostra foi iniciada com a diluição de 3-5 vezes da proteína em tampão com D<sub>2</sub>O (pD 7.0) a 25 °C (~ 70% D<sub>2</sub>O) para sua deuteração. As amostras foram incubadas em vários intervalos de tempo (5, 10, 15 e 30 minutos) e a cada ponto, uma alíquota de 70 µl era adicionada a 60 µl de tampão 20mM fosfato de sódio pH 2,5 para inibir a troca H/D. As amostras eram imediatamente digeridas com pepsina (1 mol de enzima por 10 mol de proteína) a 0 °C por 5 minutos e então aplicadas em espectrômetro de massas de triplo quadrupolo *Quattro II* (Micromass, Altrincham, UK), equipado com fonte de ionização *electrospray* (ESI) padrão.

No espectrômetro de massas, após ionização por ESI, os íons pépticos foram separados de acordo com a sua razão carga/massa ( $z/m$ ) no analisador de massas do tipo quadrupolo e, então, detectados, resultando em um espectro de massas. A deconvolução do espectro de massas da proteína intacta, sem ter sido submetida à digestão com pepsina, foi feita por meio do *software MaxEnt Electrospray* (Masslynx 3.2, Micromass, Altrincham, UK).

O deutério incorporado em um peptídeo leva ao deslocamento do centroide do pico correspondente a esse peptídeo no espectro de massas. Sendo assim, o nível de deuteração de cada peptídeo gerado por proteólise foi determinado a partir das diferenças nos centroides de massa entre os fragmentos deuterados e não-deuterados. Este procedimento foi aplicado tanto para a SELD ligada ao complexo SELA-SELC, quanto para a não-ligada. O *software MS-Digest*<sup>103</sup> foi usado para identificar a sequência selecionada de íons pépticos. O total de troca H/D foi calculado como o número total de peptídeos ligados mais um hidrogênio N-terminal por peptídeo, desconsiderando-se o número de resíduos de prolina.

Com base nos resultados, é possível analisar regiões com maior ou menor mobilidade em solução, bem como aquelas que estão expostas ou interiorizadas ao solvente e também interfaces de oligomerização e interação com ligante.



## 4 Resultados e Discussão

### 4.1 Análise da proteína SELD de *E. coli* por meio de ferramentas de Bioinformática

#### 4.1.1 Análise da sequência de aminoácidos

O monômero da proteína SELD de *E. coli* possui 347 aminoácidos e massa molecular de 36,67 kDa, com ponto isoelétrico teórico igual a 5,29, conforme calculado pela ferramenta *PROTPARAM*<sup>63</sup>. A TABELA 6 mostra a porcentagem de aminoácidos na estrutura primária da proteína.

TABELA 6 - Porcentagem de aminoácidos na estrutura primária da SELD de *E. coli*.

Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)
Ala (A)	11,0	Gly (G)	11,5	Pro (P)	5,5
Arg (R)	4,3	His (H)	1,7	Ser (S)	4,6
Asn (N)	3,5	Ile (I)	7,2	Thr (T)	6,1
Asp (D)	4,0	Leu (L)	8,6	Trp (W)	0,3
Cys (C)	2,0	Lys (K)	4,6	Tyr (Y)	1,7
Gln (Q)	2,3	Met (M)	3,5	Val (V)	6,6
Glu (E)	7,5	Phe (F)	3,5	Sec (U)	0

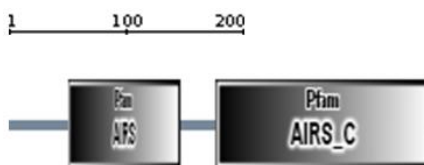
Note a presença de um número considerável de cisteínas na sequência. Porém, o servidor *DISULFIND*<sup>64</sup> prediz com alta confiabilidade que todas as cisteínas estão em estado não ligado, ou seja, provavelmente não devem formar ligações dissulfeto para manutenção da estrutura da proteína em solução.

O coeficiente de extinção molar calculado é de  $14440 \text{ M}^{-1}\text{cm}^{-1}$ , considerando-se todos os resíduos de Cys reduzidos, valor que foi utilizado para medida de concentração das amostras de proteína, descrita na seção 3.2.6.

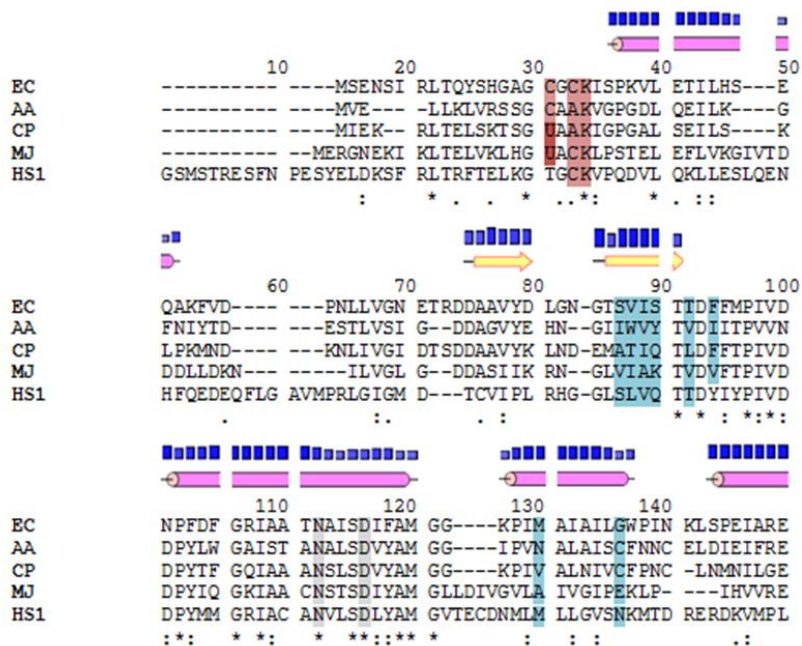
A análise de domínios da sequência de SELD de *E. coli* através do servidor SMART<sup>65</sup> revela dois domínios: AIRS (PFAM PF00586) e AIRS\_C (PFAM PF02769), presentes em proteínas da família AIRS, descrita anteriormente (seção 1.3). A TABELA 7 e a FIG. 16 mostram os resultados obtidos.

TABELA 7 - Domínios da SELD de *E. coli* obtidos a partir de análise da sequência de aminoácidos pelo servidor SMART<sup>65</sup>.

Domínio	Início	Fim	<i>E-value</i>
AIRS	49	141	2,50 e-16
AIRS_C	169	345	1,70 e-25

FIGURA 16 - Representação dos domínios da SELD de *E. coli*.

Uma análise mais detalhada através do servidor CDD<sup>66</sup> revela a conservação de aminoácidos catalíticos importantes dentro dessa família de proteínas, como os cinco aminoácidos responsáveis pela ligação de ATP e os 15 aminoácidos que formam a interface de dimerização, que são mostrados no alinhamento múltiplo (FIG. 17), obtido através do programa CLUSTALX<sup>67</sup>. Na FIG. 17, também se observa a predição de estrutura secundária para a sequência de aminoácidos, obtida por meio do programa PSIPRED<sup>68</sup>.



Continua

continuação

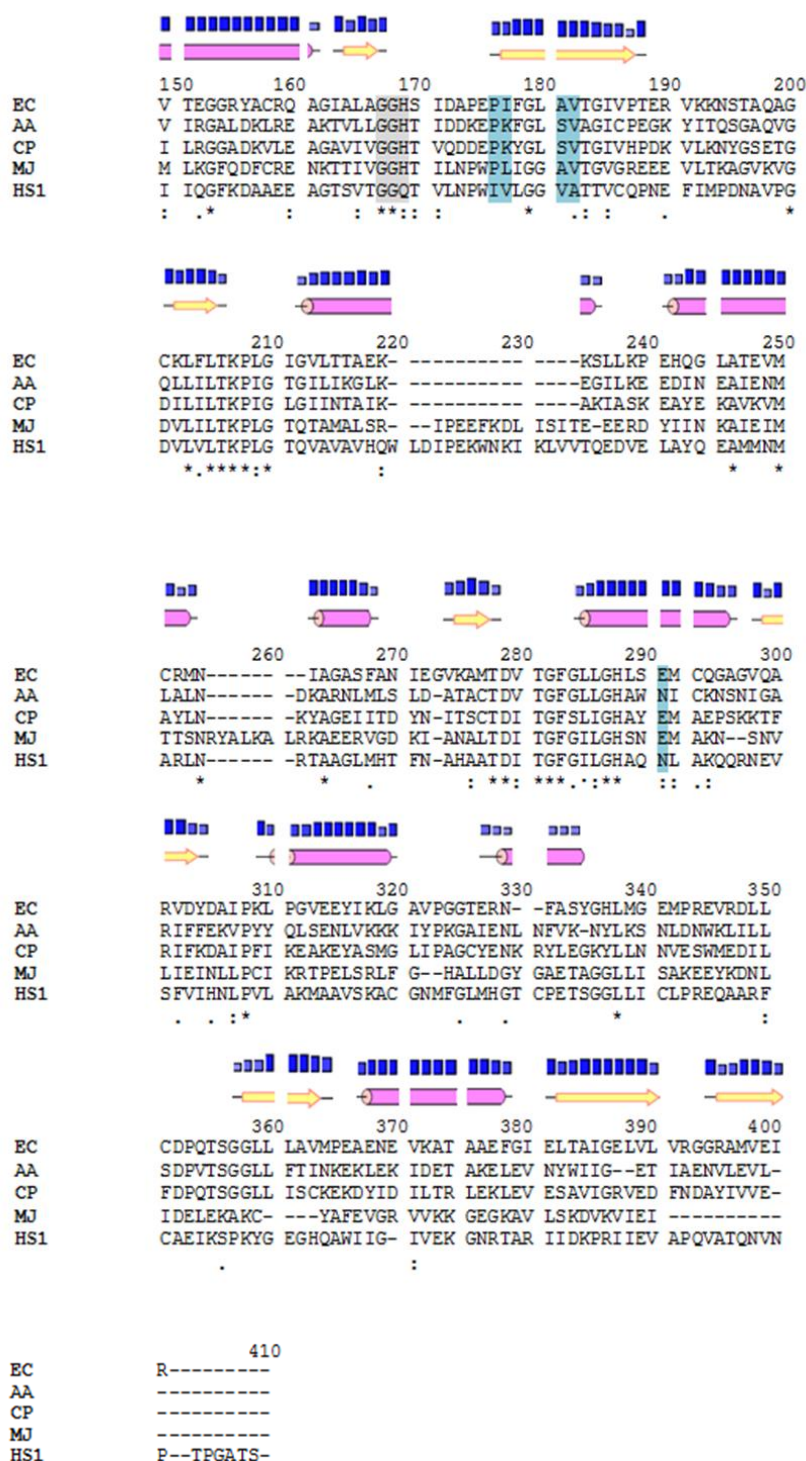


FIGURA 17 - Alinhamento múltiplo das seqüências de aminoácidos das Selenofosfatos Sintetases. Em rosa estão assinaladas a Cys (17) catalítica e a Cys (19) não catalítica, além da Lys (20) catalítica. Em cinza estão assinalados os aminoácidos que participam da ligação de ATP e, em azul, os aminoácidos que fazem parte da interface de dimerização. Alguns aminoácidos acumulam as funções de ligação de ATP e formação da interface de dimerização. São mostradas 5 seqüências representativas de Selenofosfato Sintetases: EC, *E. coli*; AA, *Aquifex aeolicus*; CP, *Clostridium perfringens*; MJ, *Methanocaldococcus jannaschii*; HS1, SPS1 de *Homo sapiens*.

#### 4.1.2 Modelagem da estrutura terciária e quaternária

O modelo inicial da estrutura tridimensional da SELD de *E. coli* foi obtido independentemente a partir das técnicas de modelagem molecular por *Threading* (servidor *I-Tasser*<sup>69</sup>) e por homologia (programa *Modeller 9.10*<sup>73</sup>).

O servidor *I-Tasser* executa a predição de estrutura terciária de uma proteína em estudo pela técnica de *Threading* de forma automatizada, como descrito na seção 3.1.2. O melhor modelo foi escolhido baseado no parâmetro de confiança *C-score* e na pontuação *TM-score* na pontuação, respectivamente iguais a 1,47, valor esperado para proteínas com enovelamento correto e 0,92, que é um valor maior que o resultado de 0,17 esperado para conformação aleatória. O *C-score* mede a convergência dos parâmetros durante a simulação computacional da organização dos fragmentos de *threading* enquanto o *TM-score* calcula a média das distâncias entre os resíduos da estrutura de *threading* com os presentes nos moldes usados para cada região. Adicionalmente, o resultado foi avaliado quanto a sua estereoquímica por meio do servidor *SAVES*<sup>71</sup>.

A disposição dos aminoácidos no gráfico de Ramachandran (FIG. 18) revela 94,8 % dos aminoácidos em regiões favoráveis, enquanto 4,9 % dos aminoácidos estão em regiões adicionalmente permitidas e apenas 0,3 % está em região generosamente permitida (apenas o resíduo Thr(229), que se encontra em uma volta), sendo que nenhum aminoácido encontra-se em região não permitida. Este resultado indica um modelo confiável. Além disso, todas as pontuações relativas à estereoquímica da estrutura também apontam a um modelo confiável.

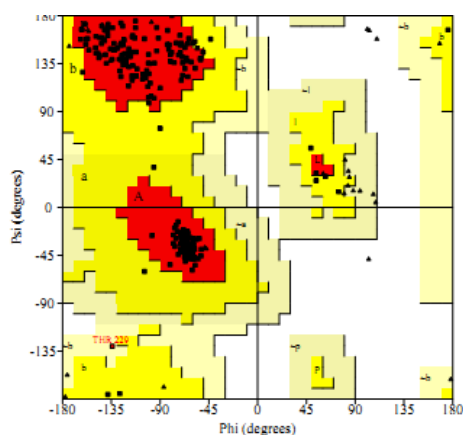


FIGURA 18 - Gráfico de Ramachandran para o modelo de *Threading* da SELD de *E. coli*.



O servidor *ProSa*<sup>72</sup> indica uma pontuação, conhecida com *Z-score*, que mede a qualidade de um modelo de estrutura tridimensional de proteína por meio da medida da energia total da estrutura com respeito a uma função distribuição de energia de conformações aleatórias, fornecendo uma faixa de valores característicos para proteínas nativas, comparadas às proteínas com estrutura conhecida por cristalografia ou RMN. No caso do melhor modelo de *Threading*, o *Z-score* foi de -7,83, valor este que se encontra dentro do limite esperado para proteínas cuja estrutura é conhecida e que tem mesmo tamanho de sequência de aminoácidos (FIG. 19-A).

O cálculo da energia média para cada resíduo de aminoácido da estrutura (FIG. 19-B), feito considerando-se janelas de 10 e 40 resíduos em torno de cada posição na estrutura primária, revelou algumas regiões de alta energia considerando-se uma janela de 10 aminoácidos, que correspondem basicamente a regiões de voltas no modelo tridimensional gerado. Porém, para a janela de 40 aminoácidos, este valor apresenta-se negativo, o que evidencia um modelo energeticamente favorável.

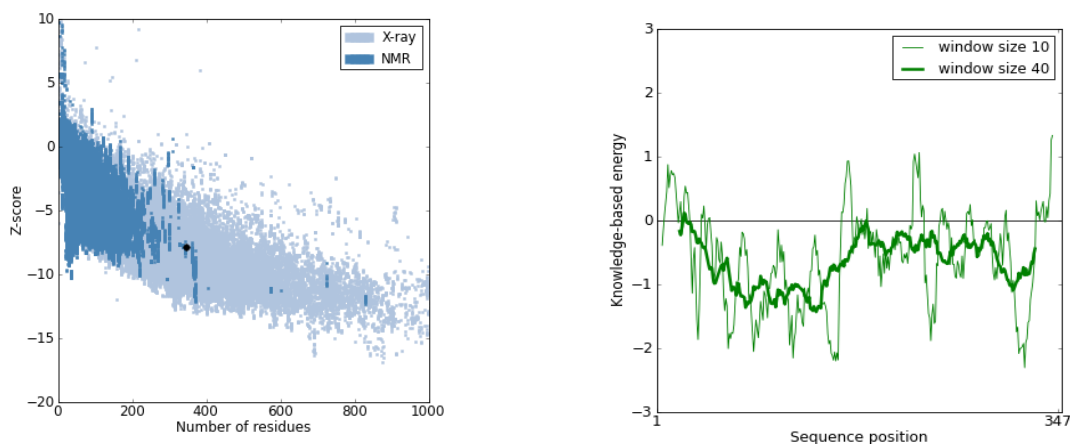


FIGURA 19 – A- Avaliação do *Z-score* e B- avaliação da energia por resíduo de aminoácido do modelo de *Threading* da SELD de *E. coli*.

Como estruturas cristalográficas estão disponíveis para Selenofosfato Sintetases com identidade sequencial próxima de 30% com a SELD de *E. coli*, a obtenção de um modelo tridimensional para a estrutura da SELD de *E. coli* também foi possível através de Modelagem Molecular por Homologia.

A busca por um modelo confiável por modelagem comparativa foi feita analisando-se as estruturas já resolvidas por cristalografia para Selenofosfatos Sintetases, cujas informações cristalográficas encontram-se na TABELA 8.

TABELA 8 - Informações cristalográficas de Selenofosfato Sintetases com estrutura resolvida e de alinhamento local com a SELD de *E. coli* (BLASTp)<sup>74</sup>.

PDB	Resolução (Å)	R-value	R-free	Grupo Espacial	Identidade com a SELD de <i>E. coli</i> (%)	Positividade (%)	Cobertura (%)
PDB 3FD5 SPS 1 humana	1.90	0.140	0.195	P4 <sub>3</sub>	30	53	88
PDB 3FD6 SPS 1 humana	1.95	0.133	0.196	P4 <sub>3</sub>	30	53	88
PDB 2ZAU SPS <i>A. aeolicus</i>	2.00	0.212	0.239	C222 <sub>1</sub>	37	53	82
PDB 2ZOD SPS <i>A. aeolicus</i>	1.98	0.204	0.240	P2 <sub>1</sub>	37	51	95
PDB 2YYE SPS <i>A. aeolicus</i>	2.10	0.173	0.225	P3 <sub>1</sub> 21	37	51	95

Adicionalmente, deve-se mencionar que as estruturas de código PDB 2ZAU e 2ZOD não apresentam ligantes no sítio ativo, enquanto que na estrutura de código PDB 3FD6 há uma molécula de ADP no sítio ativo e nas estruturas de código PDB 3FD6 e 2YYE há um análogo de ADP (AMPcP) e um análogo de ATP (AMPcPP), respectivamente. A TABELA 9 mostra a distribuição de aminoácidos em regiões do gráfico de Ramachandran para estas estruturas cristalográficas:

TABELA 9 - Porcentagem de aminoácidos nas regiões do gráfico de Ramachandran\*.

Região	PDB 3FD5 SPS 1 humana	PDB 3FD6 SPS 1 humana	PDB 2ZAU SPS <i>A.</i> <i>aeolicus</i>	PDB 2ZOD SPS <i>A.</i> <i>aeolicus</i>	PDB 2YYE SPS <i>A.</i> <i>aeolicus</i>
mais favorecida	90,5	89,1	90,7	90,4	90,4
adicionalmente permitida	8,8	10,0	8,3	8,7	8,9
generosamente permitida	0,7	0,8	0,9	0,9	0,7
não permitida	0,0	0,0	0,1	0,0	0,0

\* O gráfico de Ramachandran foi obtido através do servidor SAVES<sup>71</sup>.

Os resultados de alinhamento local da sequência de aminoácidos de SELD de *E. coli* contra as sequências do banco de dados *PDB* e os parâmetros cristalográficos das próprias estruturas depositadas sugerem que a estrutura de código *PDB* 3FD5 é adequada como molde para modelagem comparativa. Os parâmetros cristalográficos têm maior qualidade e maior resolução, o que sugere um modelo confiável, além de apresentar todos os aminoácidos em regiões permitidas do gráfico de Ramachandran. Note que a estrutura de código *PDB* 2YYE, apesar de apresentar maior identidade sequencial, tem uma qualidade inferior e, por isso, não foi escolhida como molde. Além disso, a diferença de cobertura de alinhamento deste com a SELD de *E. coli* em comparação com a estrutura de código *PDB* 3FD5 é observada em apenas em regiões preditas como alças.

Deste modo, iniciou-se o processo de modelagem comparativa a partir de um alinhamento global da sequência primária de SELD de *E. coli* com o molde, através do programa *CLUSTALW*<sup>67</sup>. Como a estrutura de código *PDB* 3FD5 apresenta-se como homodímero, com duas cadeias, designadas A e B, estas foram separadas e o processo foi repetido em paralelo para ambas. O resultado do alinhamento global das sequências é mostrado a seguir, juntamente com a estrutura secundária do molde e a predita para a SELD de *E. coli* usando o programa *PSIPRED*<sup>68</sup>:



FIGURA 20 - Alinhamento global da sequência de aminoácidos da SELD de *E. coli* com as cadeias A e B, respectivamente, do *PDB* 3FD5. Em vermelho estão coloridas as regiões de estrutura em hélice- $\alpha$  e em azul, as regiões de fitas- $\beta$ . A estrutura secundária representada para a SELD de *E. coli* foi predita utilizando o servidor *PSIPRED*<sup>68</sup>.

Nota-se, claramente, a ausência de aminoácidos assinalados na estrutura de código *PDB* 3FD5 para ambas as cadeias, que representam porções da estrutura cuja densidade eletrônica não permitia assinalamento de posição de aminoácidos. Além disso, observa-se a

ausência de alinhamento no C-terminal, em região com estrutura secundária predita como fita- $\beta$  para a SELD de *E. coli*

A partir do alinhamento global das sequências de aminoácido foram gerados 100 modelos tridimensionais para cada uma das cadeias utilizadas como molde por meio do programa *MODELLER 9.10*. Os modelos resultantes foram avaliados segundo sua Energia Proteica Otimizada Discreta (DOPE, em inglês), um potencial estatístico que mede a energia de cada modelo gerado. O resultado é mostrado a seguir (FIG. 21):

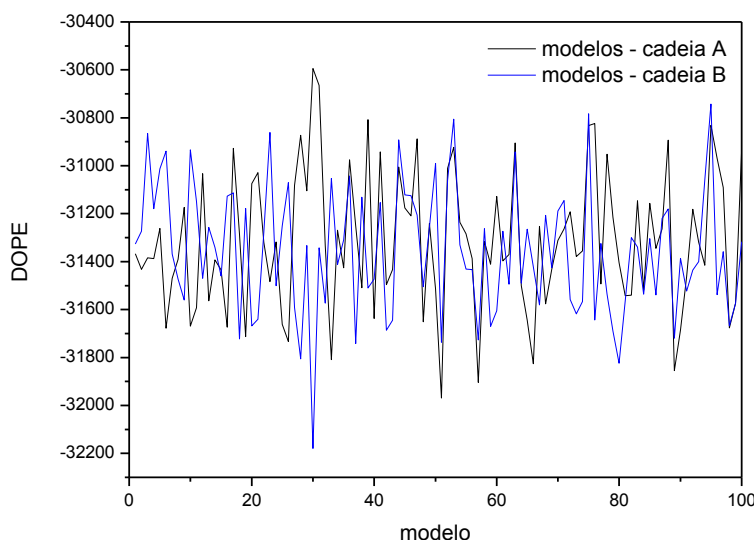


FIGURA 21 - Avaliação de DOPE para os modelos gerados por Homologia.

Dentre os 100 modelos gerados, observa-se que aqueles que apresentam menor energia são os modelos de número 51 e 30, para as cadeias A e B, respectivamente. Estes foram usados na etapa de otimização de alças (*loops*) e através dos métodos *Variable Target Function* (VTFM) e Dinâmica Molecular, implementados no programa *MODELLER 9.10*. Também foi necessária a imposição de formação de estrutura secundária em algumas regiões, de acordo com uma análise do alinhamento tridimensional do molde com o modelo final e a comparação com a predição de estrutura secundária. O valor de *DOPE score* para o modelo cujo molde é a cadeia A é de - 32760,75, enquanto para o modelo cujo molde é a cadeia B é de - 32415,04, o que mostra que a otimização diminuiu a energia dos modelos.

A avaliação da estereoquímica de ambos os modelos otimizados foi feita por meio do servidor *SAVES*<sup>71</sup>. O gráfico de Ramachandran para ambos os modelos é mostrado a seguir (FIG. 22):

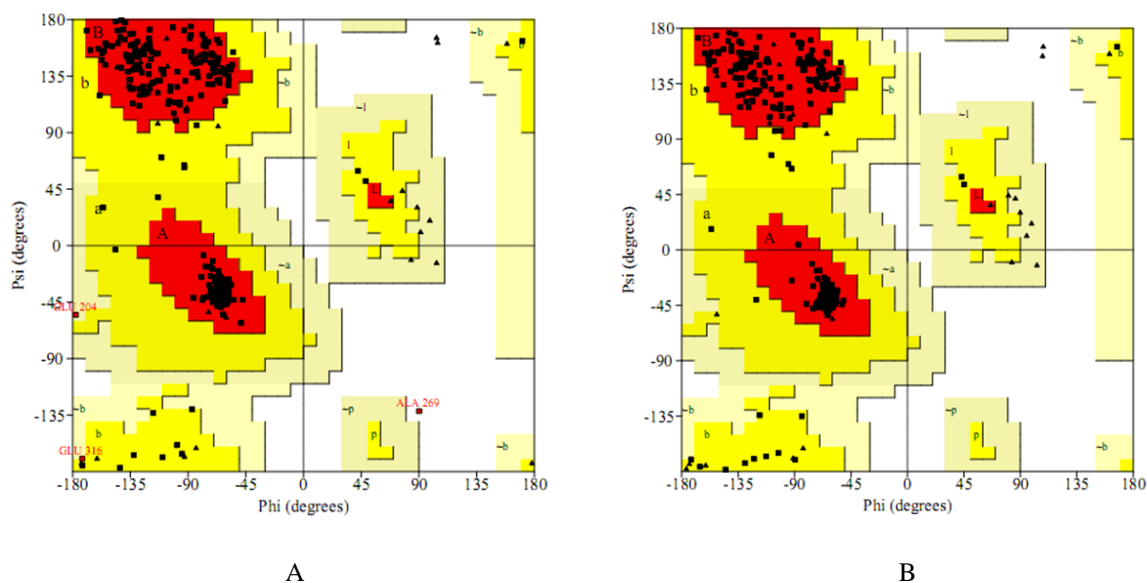


FIGURA 22 - Gráfico de Ramachandran para os modelos otimizados, obtidos das cadeias A e B, respectivamente.

TABELA 10 – Distribuição de aminoácidos no gráfico de Ramachandran para o modelo baseado na cadeia A da estrutura de código PDB 3FD5 como molde.

Regiões	Porcentagem para PDB 3FD5-A (%)	Porcentagem para PDB 3FD5-B (%)
Mais favoráveis	92,3	91,3
Adicionalmente permitidas	6,6	7,7
Generosamente permitidas	0,7	1,0
Não permitidas	0,3	0,0

Nota-se a presença de três aminoácidos em regiões generosamente permitidas do gráfico de Ramachandran para o modelo obtido a partir da cadeia A. Os resíduos de aminoácidos Ala(269) e Glu(316) encontram-se em alças, enquanto Ala(204) encontra-se no meio de uma hélice- $\alpha$ , o que indica que o modelo, mesmo otimizado, ainda não se mostra adequado para estudos estruturais. Porém, o modelo gerado a partir da cadeia B não apresenta problemas com relação à estereoquímica de aminoácidos.

O resultado gerado pelo servidor ProSa<sup>72</sup> para avaliação dos modelos é mostrado na FIG. 23.

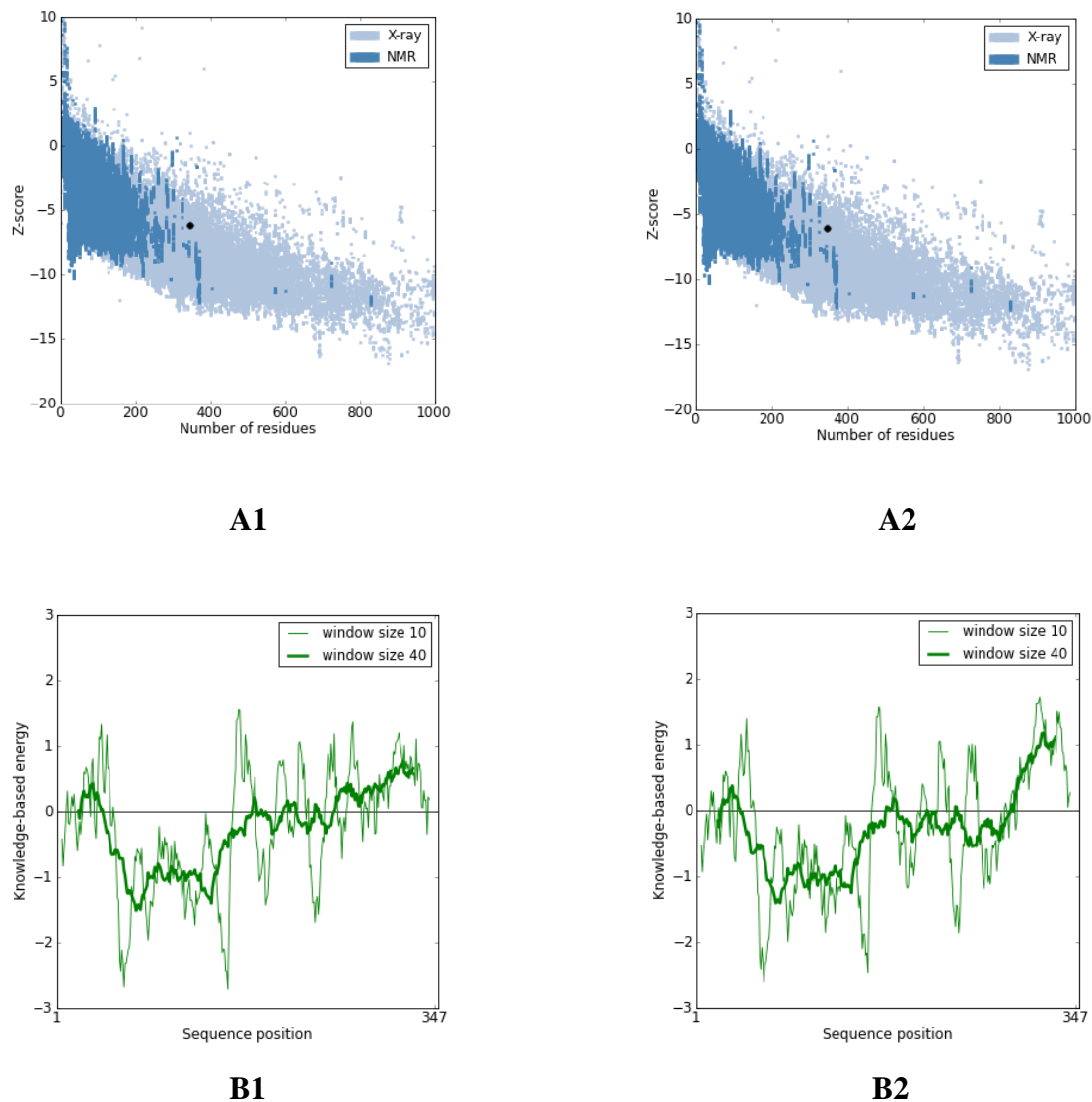


FIGURA 23 - Avaliação de *Z-score* obtido para os modelos de Modelagem por Homologia da SELD de *E. coli* (A1 – cadeia A; A2 – cadeia B) e avaliação da energia por resíduo de aminoácido da estrutura dos modelos (B1 – cadeia A e B2 – cadeia B).

O *Z-score* foi de - 6,10 para o modelo que teve como molde a cadeia A e - 6,14 para o outro, valores estes que se encontram dentro do limite esperado para proteínas cuja estrutura é conhecida e que têm mesmo tamanho de sequência de aminoácidos (FIG. 23- A1 e A2). A análise da energia por resíduo de aminoácido (FIG. 23- B1 e B2), por sua vez, revela várias regiões de alta energia, que em geral correspondem a resíduos que estão em alças (*loops*), que tendem a apresentar maior desordem. Assim, a densidade eletrônica do modelo cristalográfico é menos definida nessas regiões, o que leva a uma dificuldade de interpretação do resultado de forma confiável. Além disso, nota-se que a ausência de um molde para o C-terminal, em ambos os casos, devido à não cobertura do alinhamento entre as sequências nessa região, levou a um modelo com alta energia nessa região.

O modelo tridimensional obtido a partir da cadeia B foi comparado com aquele obtido por *Threading*, por meio de alinhamento tridimensional no programa *UCSF Chimera* e o resultado é mostrado a seguir (FIG. 24):

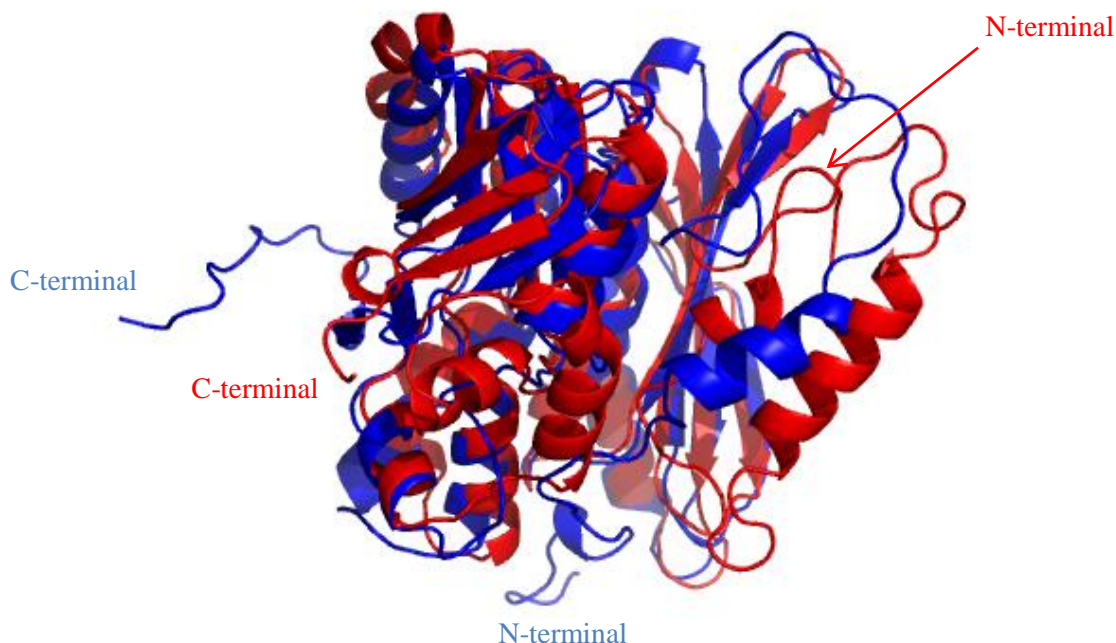


FIGURA 24 – Alinhamento tridimensional dos modelos do monômero de SELD de *E. coli* gerados por meio de Modelagem por Homologia (azul – PDB 3FD5-B) e *Threading* (vermelho).

Nota-se que a região C-terminal do modelo obtido por modelagem comparativa está totalmente desordenada, o que não condiz com a predição de estrutura secundária para região, onde deveria haver duas fitas- $\beta$  (FIG. 20). Por outro lado, observa-se que o N-terminal do modelo por homologia encontra-se em conformação “aberta”, enquanto o modelo obtido por *Threading* encontra-se em conformação “fechada”, o que é evidente quando se compara as posições das hélices de ambos os modelos na região próxima ao N-terminal.

Deste modo, o modelo obtido por *Threading* foi escolhido para a continuação dos estudos estruturais por apresentar C-terminal estruturado, além de sua boa qualidade estereoquímica, apesar de se encontrar na conformação “fechada” na ausência de ligante.

Com exceção da estrutura cristalográfica de código PDB 2ZAU, todas as outras foram cristalizadas como dímero na unidade assimétrica. Além disso, a SELD de *E. coli* mostra conservação dos aminoácidos responsáveis por dimerização na família AIRS. Assim, a estrutura dimérica da proteína em estudo foi modelada. Um alinhamento de dois monômeros idênticos com a estrutura homodimérica de código PDB 3FD5 por meio do programa *UCSF*

*Chimera*<sup>30</sup> possibilitou a modelagem do dímero de SELD de *E. coli*, que foi submetido ao servidor *ROSETTA DOCK*<sup>76</sup> para uma minimização de energia de modo a se encontrar a melhor orientação para interação. O resultado é mostrado na FIG. 25.

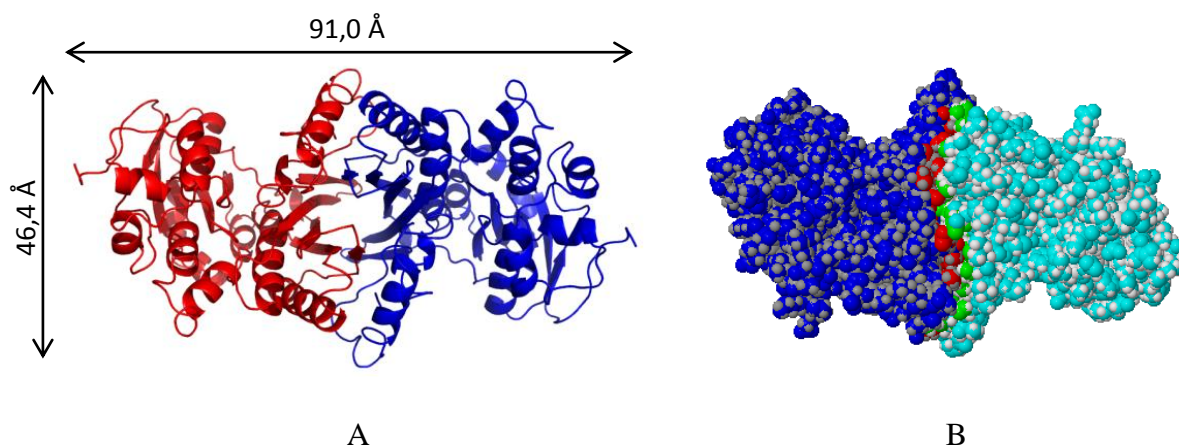


FIGURA 25 – A - Modelo da estrutura dimérica da SELD de *E. coli* obtido por meio do servidor *RosettaDock*. B – Superfície acessível à solvatação do modelo do dímero de SELD de *E. coli* (em azul claro e azul escuro) e interface de interação entre monômeros (em verde e vermelho), obtida a partir de análise do modelo no servidor *PISA*<sup>77</sup>. As dimensões do modelo da estrutura dimérica foram obtidas através do programa *UCSF Chimera*<sup>30</sup>.

Observa-se que após a minimização de energia para manutenção do dímero, a região C-terminal ainda apresenta uma ligação de hidrogênio mantendo sua interação na estrutura (dados não mostrados), diferentemente do observado para a estrutura obtida por modelagem comparativa, em que a região C-terminal não interagia com o restante da estrutura do monômero e parecia bastante flexível, o que não condiz com a predição de estrutura secundária. Deve-se mencionar, ainda, que esta região não faz parte da interface de dimerização.

Uma análise da qualidade do modelo do dímero no servidor *Prosa*<sup>72</sup> revela um *Z-score* de -7,83, valor bastante baixo comparado ao modelo do monômero obtido por *Threading* (FIG. 26-A), o que indica uma melhora na qualidade do modelo após dimerização e minimização de energia do sistema. Nota-se ainda que não há região de alta energia no modelo do dímero (FIG. 26-B).



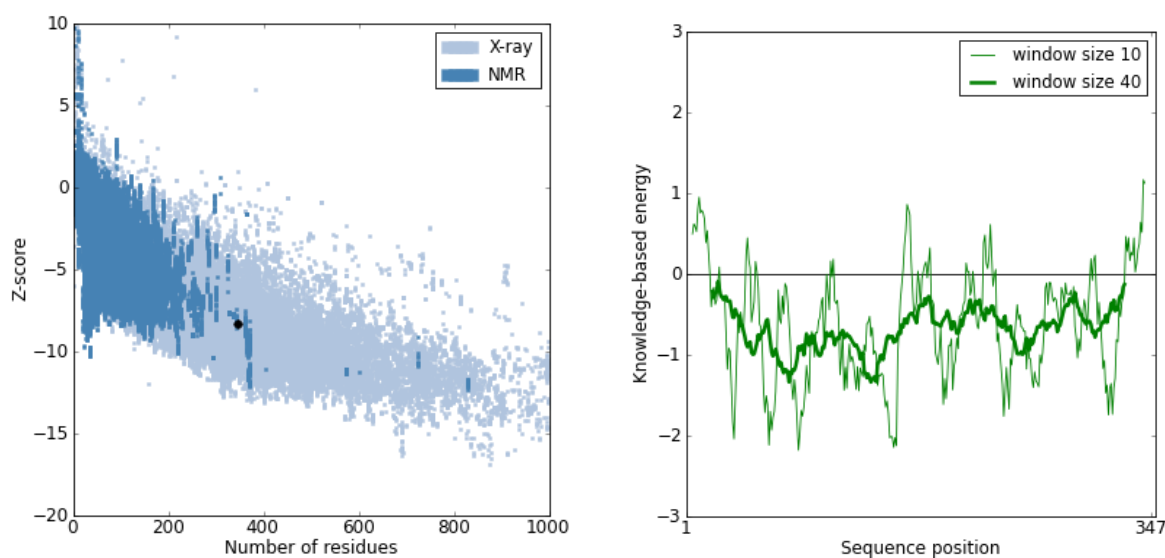


FIGURA 26 – A - Avaliação do Z-score obtido para o modelo do dímero de SELD de *E. coli* e B - Avaliação da energia por resíduo de aminoácido da estrutura do modelo.

A análise da estrutura quaternária da SELD de *E. coli* foi feita por meio do servidor PISA<sup>77</sup>. Cada monômero apresenta 347 aminoácidos, sendo 326 resíduos de aminoácidos em superfície exposta ao solvente de área total igual a 17705,4 Å<sup>2</sup>. A energia de solvatação para cada monômero é predita como sendo -328,6 kcal/mol. Os resíduos de aminoácidos inacessíveis ao solvente são Gly(81), Ala(85), Ile(89), Ile(92), Ala(102), Gly(124), Gly(125), Gly(155), Cys(172), Leu(174), Gly(181), Val(184), Ala(188), Ala(224), Ala(308), sendo, em sua maioria, resíduos de aminoácidos apolares. Entre os dois monômeros, são preditas duas interações iônicas e três ligações de hidrogênio (TABELAS 11 e 12), sendo a energia de interação calculada de -27,4 kcal/mol. A área da interface de dimerização é predita com 1854,3 Å<sup>2</sup>, formando um barril-β, como observado para estruturas cristalográficas de proteínas da família AIRS. Interações hidrofóbicas que atuam na manutenção da estrutura do dímero não são preditas pelo servidor e, portanto, os aminoácidos conservados na família de proteínas AIRS essenciais para a dimerização foram identificados somente como presentes em áreas não expostas ao solvente, mas não como responsáveis por interações fundamentais para a manutenção do dímero.

TABELA 11 – Ligações de hidrogênio preditas na interface de dimerização (PISA<sup>77</sup>).

Ligações de Hidrogênio	Distância (Å)
ASN(46) – GLY(133)	2,22
GLU(116) – LYS(36)	2,45
LEU(136) – THR(48)	2,64

TABELA 12 – Interações iônicas preditas na interface de dimerização (PISA<sup>77</sup>).

Interações iônicas	Distância (Å)
LYS(46) – GLU(123)	3,69
GLU(116) – LYS(36)	3,17

Para uma completa análise teórica da proteína SELD de *E. coli*, o modelo do homodímero foi submetido à predição de parâmetros hidrodinâmicos no programa *HYDROPRO*<sup>78</sup> e os resultados encontram-se na TABELA 13:

TABELA 13 - Propriedades hidrodinâmicas do melhor modelo de SELD de *E. coli*.

	MM (kDa)	$s_{20,w}^0$ (S)	$s_{esfera}/s_{20,w}^0$	$D_{20,w}$ ( $10^{-7}$ cm <sup>2</sup> /s)	$D_{esfera}/D_{20,w}$	$D_{máx}$ (Å)	Rg (Å)
<b>Modelo SelD de <i>E. coli</i></b>	73,94	4,49	1,37	5,70	1,35	91	28,3

MM - massa molecular

$s_{20,w}^0$  - coeficiente de sedimentação no estado padrão

$s_{esfera}$  - coeficiente de sedimentação para uma esfera de mesma massa

$D_{20,w}$  - coeficiente de difusão no estado padrão

$D_{esfera}$  - coeficiente de difusão para uma esfera de mesma massa

$D_{máx}$  - máximo comprimento do envelope molecular da proteína

Rg - raio de giro

As razões  $s_{esfera}/s_{20,w}^0$  e  $D_{esfera}/D_{20,w}$ , acima de 1,4, indicam uma proteína um pouco alongada, como observado visualmente (FIG. 23).

Os parâmetros teóricos apresentados nesta seção foram obtidos experimentalmente por meio de diferentes técnicas biofísicas e são apresentados nas seções seguintes.

## 4.2 Preparação da SELD de *E. coli*

A partir de oligonucleotídeos específicos, foi possível amplificar a fase aberta de leitura do gene *selD* do DNA genômico de *E. coli* por PCR, utilizando a enzima Taq DNA polimerase. O resultado da amplificação foi analisado em gel de agarose 1% e é mostrado na FIG. 27.

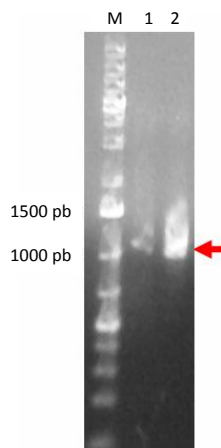


FIGURA 27 - Amplificação da ORF do gene *selD* de *E. coli*: Resultado analisado em gel de agarose 1%, TAE 1X. M: marcador de massa molecular *Gene Ruler 1Kb Ladder Plus* (Fermentas Life Sciences). Os números 1 e 2 representam o fragmento amplificado por PCR, referente ao gene *selD* *E. coli*, indicado pela seta, sendo que a amostra 2 apresenta mais DNA amplificado (dada a intensidade da banda indicada pela seta).

A análise do resultado em gel de agarose mostra que o produto de PCR possui entre 1000 e 1500 pares de bases (FIG. 27), o que é consistente com o tamanho esperado do fragmento de DNA produto da amplificação, igual a 1041 pb.

A ligação em vetor de expressão pET28a(+) foi verificada por meio da reação de digestão com as enzimas de restrição *NdeI* e *XhoI* e o resultado é mostrado na FIG. 28.

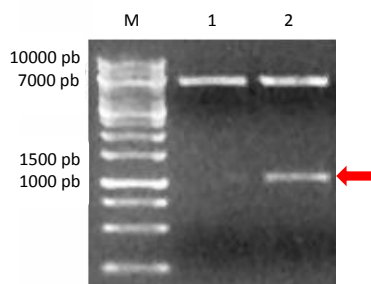


FIGURA 28 - Verificação de clonagem da ORF do gene *selD* de *E. coli* em vetor de expressão pET28a(+). Resultado analisado em gel de agarose 1% TAE 1X. Representa a digestão da construção *selD E. coli/pET28a(+)* com as enzimas *NdeI* e *XhoI*. A seta indica uma banda entre 1000pb e 1500pb, correspondente ao gene *selD E. coli*. M: marcador de massa molecular de 1Kb plus DNA Ladder (Fermentas Life Sciences). 1: Controle negativo. 2: Amostra resultante da reação de digestão.

Este resultado da digestão, obtido a partir de análise em gel de agarose, mostra uma banda entre 1000 e 1500 pares de bases, tamanho esperado do fragmento de DNA correspondente à ORF de *selD* de *E. coli*, igual a 1041 pb. A banda com maior número de pares de bases corresponde ao vetor pET28a(+) linearizado. Deste modo, confirma-se a ligação do inserto no vetor de expressão.

O sequenciamento da ORF do gene *selD* *E.coli* em vetor pET28 a(+) foi realizado pelo método de Sanger e as sequências de nucleotídeos obtidas foram analisadas e alinhadas com a sequência do banco de dados através do programa *SeqMan (Lasergene Sequence Analysis Software – DNASTa)* e o resultado do alinhamento das extremidades 5' e 3' do gene *selD* de *E.coli* é mostrado na FIG. 29.

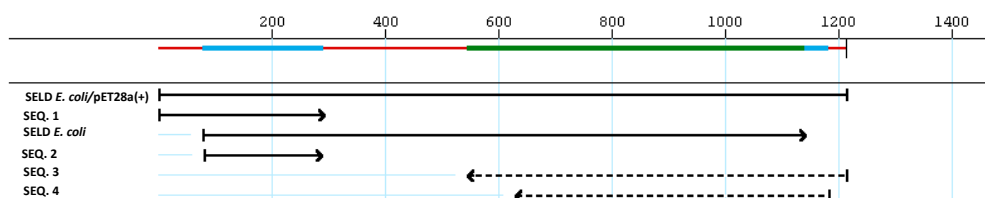


FIGURA 29 - Alinhamento das extremidades 5' e 3' do gene *selD* *E.coli*. Os alinhamentos indicados pelas setas no sentido esquerda para direita (SEQ. 1 e SEQ. 2) representam o alinhamento da extremidade 5' e os alinhamentos indicados pelas setas no sentido direita para esquerda (SEQ. 3 e SEQ. 4) representam o alinhamento 3'. *SEL D E. coli* representa a sequência original.

Nota-se que a ORF do gene não foi sequenciada por completo, mas o sequenciamento das extremidades 3' e 5' já é o suficiente para supor que o gene *selD* *E.coli* foi clonado em vetor de expressão pET28a(+).

Células de *E. coli* da cepa BL21(DE3) transformadas com a construção recombinante *selD E. coli* / pET28a(+), como descrito na seção 3.2.3, foram inoculadas em 1 L de meio de cultura LB e induzidas a expressar a proteína SELD de *E. coli* a 37 °C com adição de 1 mM de IPTG por período de 3h. O resultado pôde ser visualizado em gel SDS-PAGE 15%, como mostra a FIG. 30, onde se nota, na coluna 3, que a maior parte da proteína se encontra na fração solúvel.

As amostras contendo proteínas SELD de *E. coli* expressas em sua fração solúvel foram submetidas à purificação por cromatografia de afinidade, realizada em resina de cobalto, como descrito na seção 3.2.5. A proteína SELD de *E. coli* foi eluída da resina pela aplicação de tampão I com diferentes concentrações de Imidazol, o que pôde ser visualizado em gel SDS-PAGE 15%, como mostra a FIG. 30.

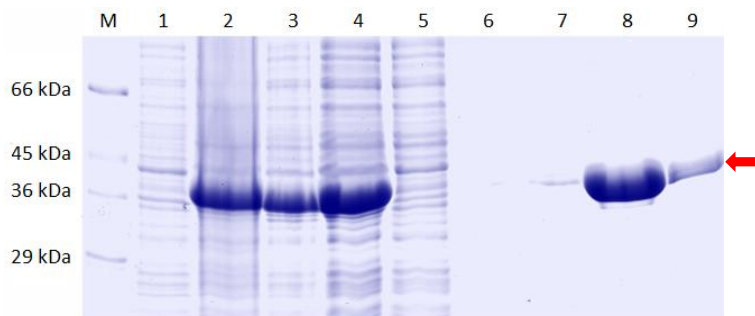


FIGURA 30 - Expressão e purificação por cromatografia de afinidade. M: padrão de massa molecular; 1: antes e 2: depois de indução com IPTG; 3: pellet e 4: sobrenadante após lise; 5: eluato; eluição com 6: 10 mM, 7: 20mM, 8: 200 mM e 9: 500 mM Imidazol.

A concentração de proteína da amostra eluída a 200 mM de Imidazol, medida através do equipamento *Nanodrop<sup>TM</sup> 1000* (*Thermo Scientific*), foi de 2,0 mg/ml (54,5  $\mu$ M) em 10 ml de amostra, a qual foi submetida à diálise e posterior concentração por ultrafiltração até uma concentração de 5,0 mg/ml (136,4  $\mu$ M). Em seguida, a amostra foi submetida à clivagem da cauda de histidinas com a protease Trombina, como descrito na seção 3.2.6, durante 14 horas. O resultado foi verificado em SDS-PAGE 15% (FIG. 31).

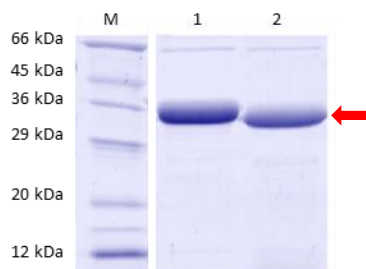


FIGURA 31 - Clivagem com Trombina da proteína SELD de *E. coli*. Analisado em SDS-PAGE 15%. M: padrão de massa molecular; amostras 1: não submetida e 2: submetida à reação de clivagem. A seta indica a banda correspondente à amostra clivada.

Observa-se que, após 14 horas de reação com Trombina, já se pode supor clivagem completa das proteínas da amostra de SELD de *E. coli*, pela diminuição da massa molecular da proteína.

A etapa seguinte, de purificação por cromatografia de exclusão molecular, foi realizada em coluna Superdex<sup>TM</sup> 200 HR 10/300 (GE). O resultado é mostrado a seguir (FIG. 32):

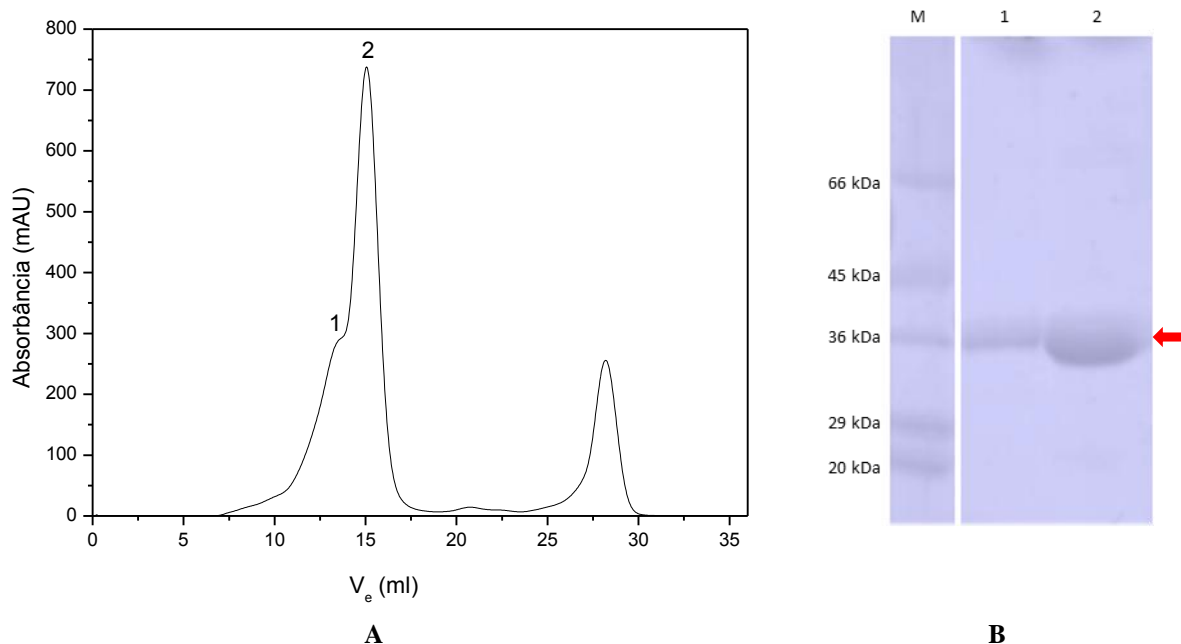


FIGURA 32 - Resultado da purificação da proteína SELD *E. coli* através de cromatografia de exclusão molecular. A - Perfil de eluição de SELD *E.coli*. B - SDS-PAGE 15% dos picos P1 ( $V_e = 13,50$  ml) e P2 ( $V_e = 15,06$  ml).

São notáveis dois picos de eluição da proteína SELD de *E. coli* no cromatograma, o que é um indício da presença de mais de uma forma oligomérica em solução, já que ambos apresentam a proteína SELD de *E. coli*, como observado no SDS-PAGE. O cálculo da massa molecular e raio de Stokes correspondentes às partículas separadas nos picos 1 e 2 é mostrado na seção 4.3.2.

### 4.3 Caracterização biofísica da proteína recombinante SELD de *E. coli*

#### 4.3.1 Espectroscopia de Dicroísmo Circular

Utilizando a amostra de SELD de *E. coli* a uma concentração de 0,2 mg/ml (5,5  $\mu$ M), foi realizada espectroscopia de Dicroísmo Circular, cujo espectro é mostrado a seguir (FIG. 33):

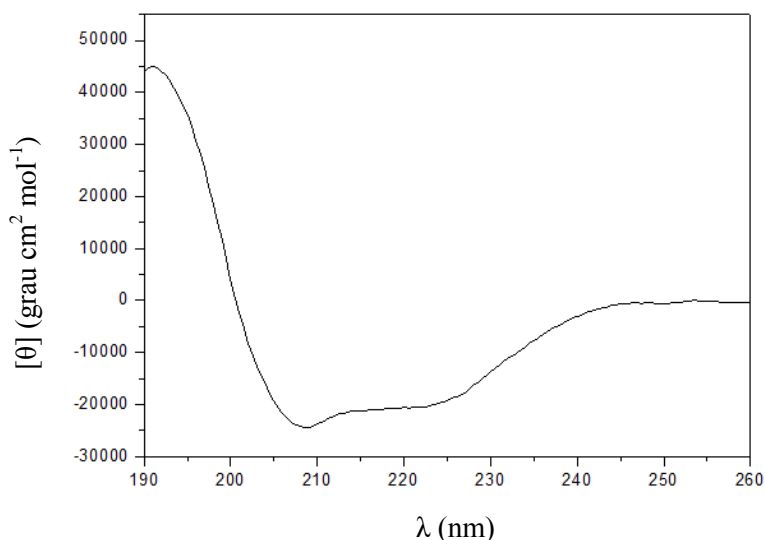


FIGURA 33 - Espectro de Dicroísmo Circular da SELD de *E. coli*.

Nota-se um pico negativo em 222 nm, devido a transições do tipo  $n-\pi^*$  na ligação peptídica, além de outro pico negativo em 208 nm e uma banda positiva intensa em 192 nm, devido a transições  $\pi-\pi^*$ . Este tipo de espectro é característico de proteínas com predominância de hélices- $\alpha$ .

A deconvolução do espectro experimental, por meio do programa *CDNN*<sup>85</sup> (base de dados referência com 13 proteínas), levou ao resultado apresentado a seguir, em comparação com a porcentagem de estrutura secundária observada no modelo tridimensional da SELD de *E. coli* (seção 4.1.2):

TABELA 14 - Deconvolução do espectro de CD da SELD de *E. coli*, apresentando a porcentagem de estruturas secundárias da proteína obtida por CD em comparação com a do modelo tridimensional da mesma, obtido por Modelagem Molecular e com a estrutura de código PDB 3FD5.

	Hélices- $\alpha$ (%)	Fitas- $\beta$ (%)	Voltas (%)	Região Desordenada (%)
<b>CD</b>	26,0	27,7	18,5	35,7
<b>Modelo SELD <i>E. coli</i></b>	30,3	22,5		47,2
<b>PDB 3FD5</b>	32,4	25,1	3,8	38,6

Uma comparação entre os resultados experimental e teórico indica um resultado de deconvolução confiável, comparável, inclusive, à porcentagem de estrutura secundárias observada para as Selenofosfato Sintetases com estrutura cristalográfica depositada no Banco de Dados *PDB*.

### 4.3.2 Estimativa da massa molecular e raio hidrodinâmico por Cromatografia de Exclusão Molecular

Para verificar o estado oligomérico das proteínas em estudo, foi realizada uma cromatografia de exclusão molecular, nas mesmas condições descritas em 2.2.8, com padrões de massa molecular (MM) descritos na TABELA 15, adquiridos da *SIGMA-ALDRICH (USA)*. Os resultados experimentais de volumes de eluição ( $V_e$ ) na coluna para cada padrão também são apresentados a seguir:

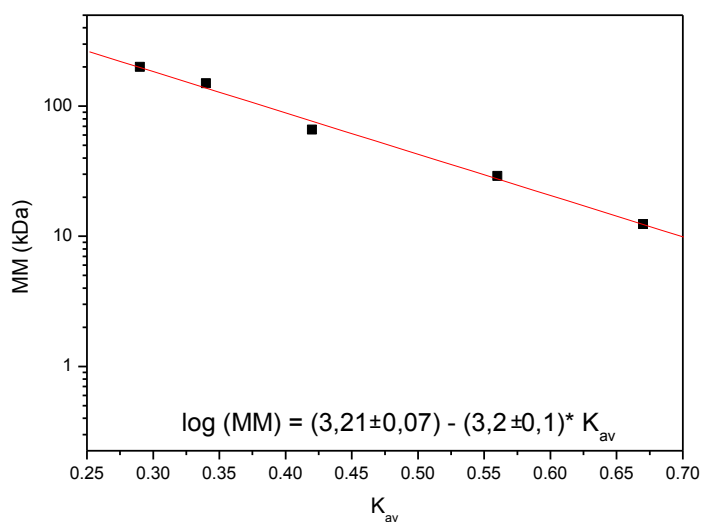
TABELA 15 - Padrões para calibração da SUPERDEX™ 200 10/300 (GE) e respectivos volumes de eluição experimentais.

Padrões	MM (kDa)	$R_H$ (Å)	$V_e$ (mL)	$K_{av}$ *
Citocromo c equino	12,4	17,0	18,44	0,66
Anidrase Carbônica bovina	29,0	23,6	16,82	0,56
BSA	66,0	33,9	14,48	0,42
Álcool Desidrogenase (levedura)	150,0	46,0	13,19	0,34
$\beta$ -amilase (batata doce)	200,0	50,4	12,47	0,29
Dextran Azul	2000,0	-	7,74	-

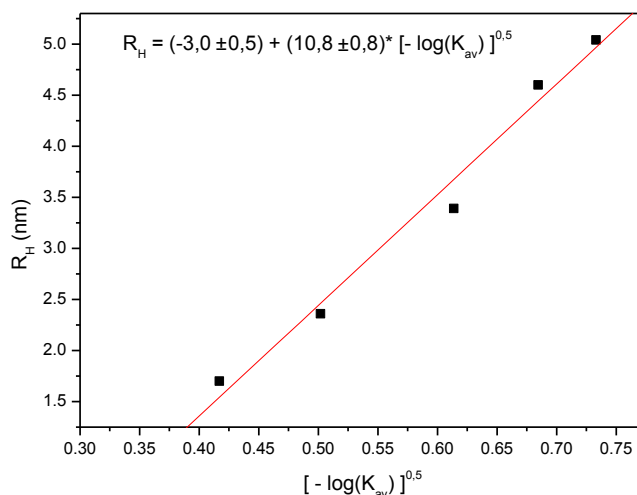
$$* K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

O volume total da coluna é  $V_t = 24$  ml e o volume morto ( $V_0$ ) igual ao volume de eluição do dextran azul. As curvas de calibração para Massa Molecular (MM) e raio hidrodinâmico ( $R_H$ ) são apresentadas a seguir:





A



B

FIGURA 34 - Curvas de calibração da SUPERDEX 200™ 10/300 para estimativa de A – massa molecular e B – raio hidrodinâmico.

No caso da SELD de *E. coli*, foram observados dois picos de eluição (FIG. 32), sendo um deles com volume de eluição igual a  $V_e = 13,50$  ml e outro com  $V_e = 15,06$  ml, que correspondem a massas moleculares de  $(123 \pm 29)$  kDa e  $(62 \pm 15)$  kDa, e raios hidrodinâmicos de  $(40 \pm 10)$  Å e  $(30 \pm 10)$  Å, respectivamente. Estes resultados sugerem a presença de uma mistura de dímeros ( $MM_{teórica} = 73,34$  kDa) e tetrâmeros ( $MM_{teórica} = 146,68$  kDa) em solução, com erros relativos de 16,1% e 15,5%, respectivamente.

### 4.3.3 Eletroforese em Gel Nativo

O experimento de Eletroforese em Gel Nativo foi realizado com a amostra do pico 2 observado no cromatograma da segunda etapa de purificação (FIG. 32), que se mostrava mais concentrada. Como padrão de massa molecular foi utilizado o conjunto de proteínas mostrado na TABELA 16, junto com o valor do raio hidrodinâmico conhecido para cada proteína e sua distância de migração (D) no gel com 3,90 cm de extensão.

TABELA 16 - Padrões para determinação de massa molecular e raio hidrodinâmico a partir de Eletroforese em Gel Nativo e respectivas distâncias de migração.

Padrões	MM (kDa)	$R_H$ (Å)	D (cm)
Tiroglobulina bovina	669,0	85,0	0,20
Ferritina (cavalo)	440,0	61,0	0,60
Catalase	232,0	52,2	1,15
Aldolase	140,0	43,0	1,75
BSA	66,0	33,9	2,60

O resultado da migração da SELD de *E. coli* no gel nativo é mostrado na FIG. 35, e as as curvas de calibração na FIG. 36:

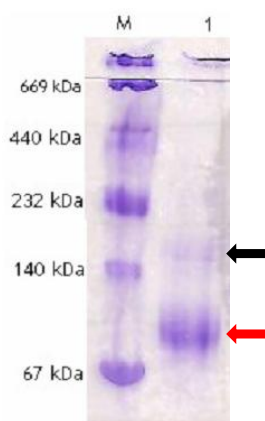


FIGURA 35 – Eletroforese em gel nativo. M: amostra com proteínas-padrão. 1: SDEC *E. coli* (pico 2 da cromatografia de exclusão molecular). A seta preta indica a forma tetramérica e a seta vermelha indica a forma dimérica da SELD de *E. coli*

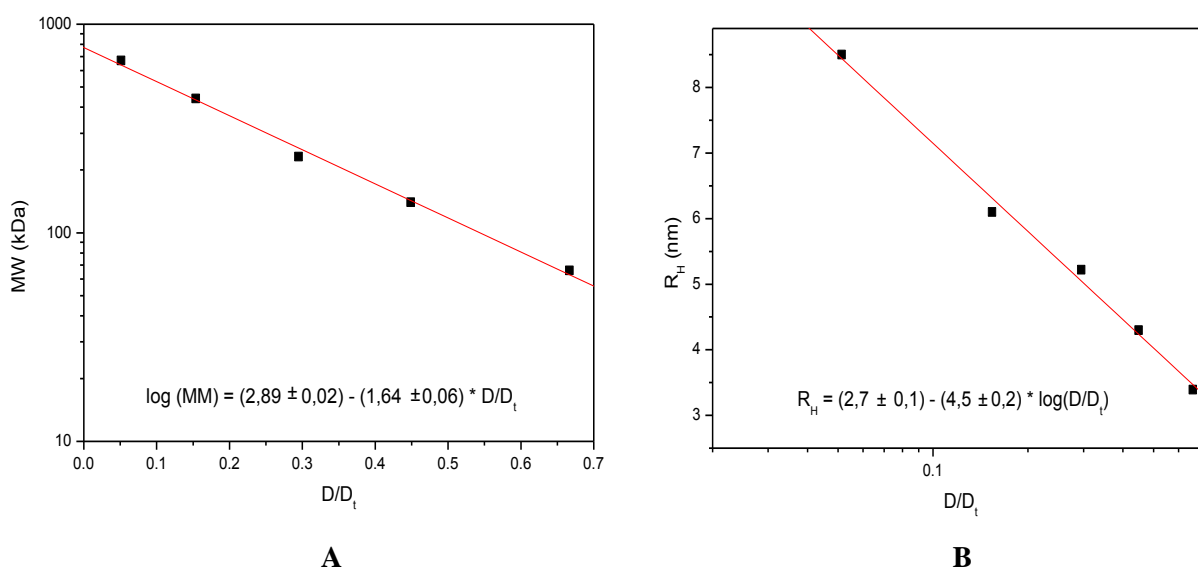


FIGURA 36 - A - Curva de Calibração para Massas Moleculares a partir de Eletroforese em Gel Nativo e B - Curva de Calibração para Raio Hidrodinâmico a partir de Eletroforese em Gel Nativo.

Para a amostra de SELD de *E. coli* concentrada do pico 2 (FIG. 32), observam-se duas bandas (FIG. 35), uma com maior intensidade e distância de migração no gel igual 2,7 cm e outra menos intensa com distância de migração igual a 2,0 cm. Estes resultados correspondem a uma massa molecular estimada de  $(57 \pm 8)$  kDa e  $(112 \pm 13)$  kDa e um raio hidrodinâmico de  $(34 \pm 1)$  Å e  $(40 \pm 2)$  Å, respectivamente. Estes resultados concordam com aquele obtido por cromatografia de exclusão molecular, revelando uma mistura entre dímeros e tetrâmeros em solução. Note que o tetrâmero aparece em menor quantidade (FIG. 35).

É notável que, mesmo após ter sido isolada, a amostra do pico 2, na forma dimérica, ainda continha uma pequena quantidade de proteína na forma tetramérica, como observado em Gel Nativo, o que pode ser devido à dificuldade da separação dos picos com a coluna cromatográfica utilizada ou devido a um possível equilíbrio dinâmico entre as formas em solução. Este comportamento foi mais bem investigado através das técnicas de DLS, SAXS e AUC, descritas nas seções seguintes.

Um cuidado que se deve tomar na análise do resultado desse experimento é com respeito ao fato de que em gel não desnaturante, a proteína é separada de acordo com sua massa, mas também de acordo com sua carga e forma. Distribuição irregular de cargas na superfície proteica podem levar a um resultado duvidoso.

#### 4.3.4 Espalhamento Dinâmico de Luz (DLS)

A técnica de DLS foi utilizada para verificação da homogeneidade das soluções com as proteínas em estudo. A seguir, encontram-se tabelados os resultados medidos para coeficiente de difusão translacional das partículas em solução em diferentes concentrações de amostra:

TABELA 17 - Parâmetros hidrodinâmicos da SELD de *E. coli* obtidos por DLS.

C (mg/ml)	C (μM)	D (10 <sup>-7</sup> cm <sup>2</sup> /s)	D <sub>20,w</sub> (10 <sup>-7</sup> cm <sup>2</sup> /s)*
0,6	16,4	4,97	5,03
0,8	21,8	5,03	5,09
1,0	27,3	5,03	5,09
1,2	32,7	4,79	4,85
1,5	40,9	4,76	4,82

$$* D_{20,w} = \frac{(293,2 \text{ K})}{T} \frac{\eta_{T,w}}{\eta_{20,w}} \frac{\eta_s}{\eta_w} D_{obs}$$

O coeficiente de difusão translacional corrigido para o estado padrão da água, a 20°C, igual a  $D_{20,w}^0 = (5,28 \pm 0,14) \times 10^{-7} \text{ cm}^2/\text{s}$ , representa o coeficiente de difusão translacional sem influência da concentração de proteína na amostra e foi obtido a partir do coeficiente linear da reta mostrada a seguir:

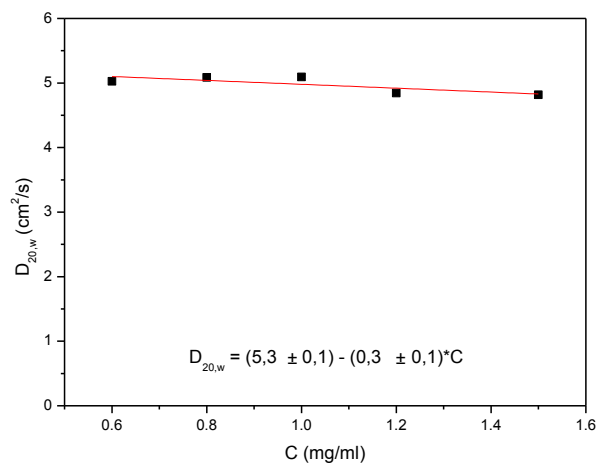


FIGURA 37 - Comportamento do coeficiente de difusão translacional com a concentração da amostra de proteína, medido por DLS.

O resultado obtido experimentalmente para o coeficiente de difusão translacional concorda com o teórico, calculado para o modelo obtido por modelagem molecular (TABELA 17), com erro relativo de 7,3%. Note que, considerando o resultado experimental igual a  $D_{20,w}^0 = (5,28 \pm 0,14) \times 10^{-7} \text{ cm}^2/\text{s}$ , tem-se que a razão friccional é igual a  $D_{\text{esfera}} / D_{20,w}^0 = 1,46$ , o que indica uma proteína um pouco alongada, como observado no modelo tridimensional da SELD de *E. coli*.

O objetivo deste experimento foi obter medidas de DLS a baixas concentrações para que o efeito de aumento de concentração não influenciasse na formação de outras formas oligoméricas em solução, caso fosse verdadeira a hipótese de um equilíbrio entre os dímeros e tetrâmeros em solução. Note, pelos resultados, que os resultados mostram valores de coeficiente de difusão e raio de Stokes para uma partícula dimérica em solução, o que concorda com o observado por Cromatografia de Exclusão Molecular e Gel Nativo. O raio de Stokes (raio hidrodinâmico), obtido para esta amostra por meio de DLS, foi de 34,1 Å e a massa molecular média foi igual a 59,8 kDa calculada como se a proteína fosse uma esfera com o raio de Stokes, indicam a predominância de um dímero em solução, apesar de a solução apresentar-se polidispersa, com coeficiente de polidispersão em torno de 20%.

#### **4.3.5 Espalhamento de raios-X a baixo ângulo (SAXS)**

Como o experimento de DLS não identificou outra forma além da dimérica para a SELD de *E. coli*, optou-se por empregar a técnica de SAXS, que tem maior precisão, para investigar as formas oligoméricas observadas por Eletroforese em Gel Nativo. A curva experimental de espalhamento de raios-X a baixo ângulo na amostra de proteína nativa a 1,5 mg/ml (40,9 µM) é mostrada a seguir, na FIG. 38, assim como o gráfico de Guiner, inserido.

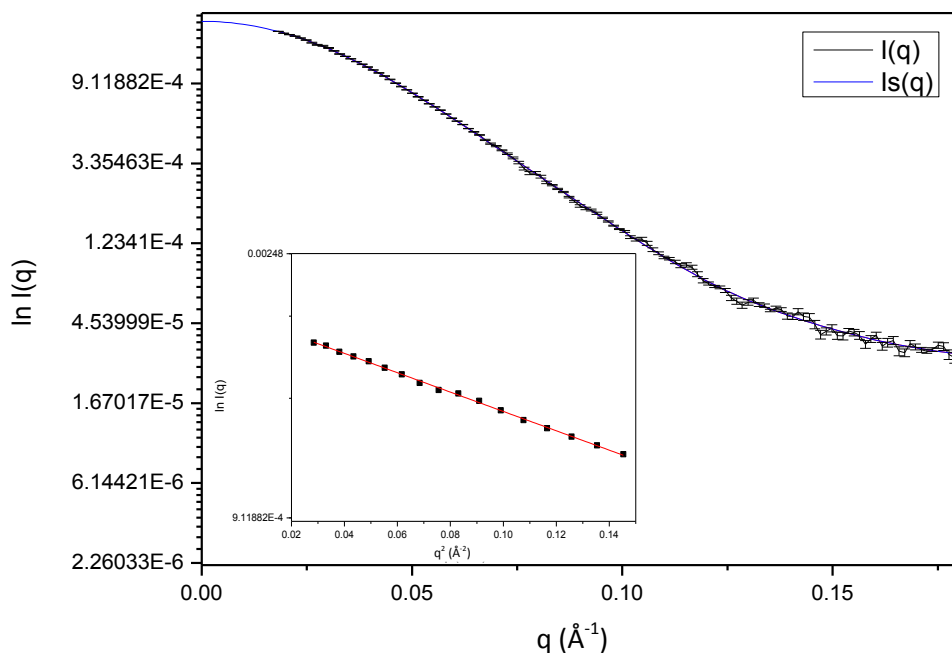


FIGURA 38 - Curva de espalhamento de raios a baixo ângulo para a SELD de *E. coli* e Gráfico de Guinier inserido.  $I(q)$  é a curva experimental e  $I_s(q)$  é a curva extrapolada para  $q = 0$  (programa GNOM<sup>90</sup>).

Os dados de SAXS apresentados têm resolução na faixa de 34,3 Å a 345,6 Å, baseado nos valores medidos de ângulo de espalhamento ( $q$ ). A aproximação de Guinier é válida para os primeiros 17 pontos da curva de espalhamento ( $qR_g < 1,3$ ) e resulta em um raio de giro igual a  $(33,1 \pm 0,3)$  Å. A partir do valor de  $I(0)$ , que corresponde ao intercepto do gráfico de Guinier no eixo das ordenadas, por extrapolação, foi possível estimar a massa molecular da partícula em solução como sendo 85,4 kDa, através do servidor *SAXS MoW*. Este valor indica a predominância de um dímero em solução e o erro relativo para a massa molecular teórica do dímero é de 16,4%. Novamente, o resultado concorda com o observado por DLS. O alto erro relativo é um indicativo da existência de outra forma oligomérica em solução, além do dímero. Porém, dado o resultado, pode-se afirmar que o dímero está em grande porcentagem em solução, como observado por Gel Nativo.

Uma vantagem da técnica de SAXS é que se pode calcular a distribuição de distâncias,  $p(r)$ , entre centros espalhadores de raios-X na molécula, o que pode ser usado para se reconstruir a superfície da molécula. Utilizando-se o programa GNOM<sup>90</sup>, obteve-se a distribuição de distâncias,  $p(r)$ , mostrada a seguir:

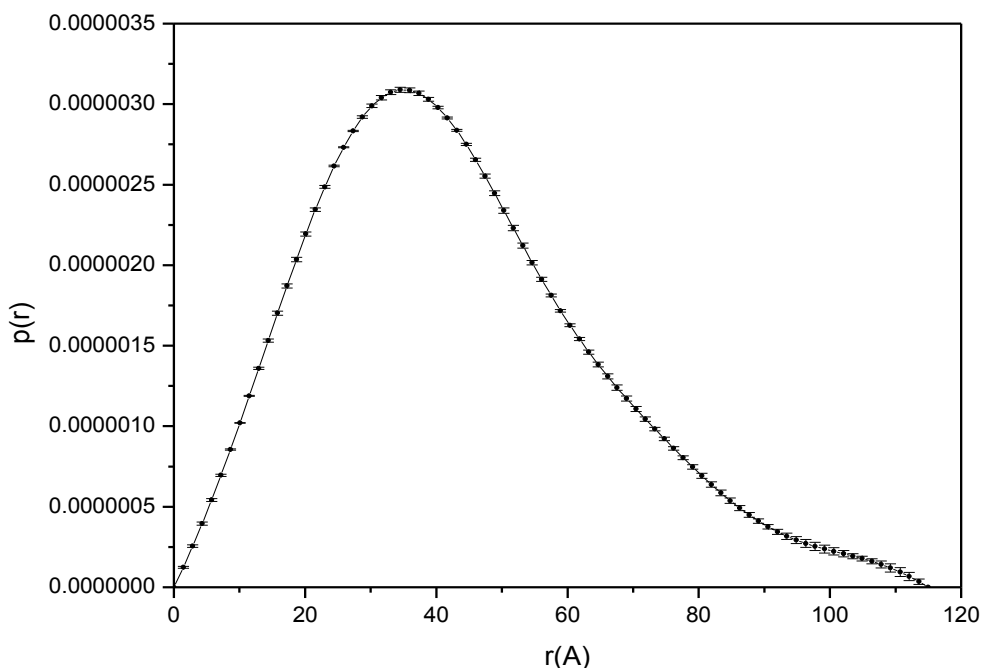


FIGURA 39 - Distribuição de distâncias,  $p(r)$ .

A distância máxima dentro da estrutura da SELD de *E. coli* foi estimada como sendo 113 Å e seu raio de giro como sendo igual a  $(34,16 \pm 0,06)$  Å em concordância com o valor do raio de giro obtido no espaço recíproco, através da aproximação de Guinier. Porém, o resultado não está próximo ao calculado para o modelo tridimensional obtido por Modelagem Molecular. Este é outro indício de que a presença de tetrâmeros em solução afeta os resultados de SAXS. Portanto, os valores calculados para  $D_{\text{máx}}$  e raio de giro através de SAXS não correspondem a uma partícula em solução, mas a uma mistura entre formas oligoméricas da SELD de *E. coli* com predominância de dímeros.

Através do programa *CRY SOL*<sup>91</sup>, foi possível comparar as curvas de espalhamento preditas para as estruturas disponíveis no banco de dados PDB, além do modelo obtido por homologia para a SELD de *E. coli* com o resultado experimental de SAXS. O melhor ajuste foi obtido para o PDB 3FD6, com  $\chi = 3,849 (> 1)$ , como é mostrado a seguir:

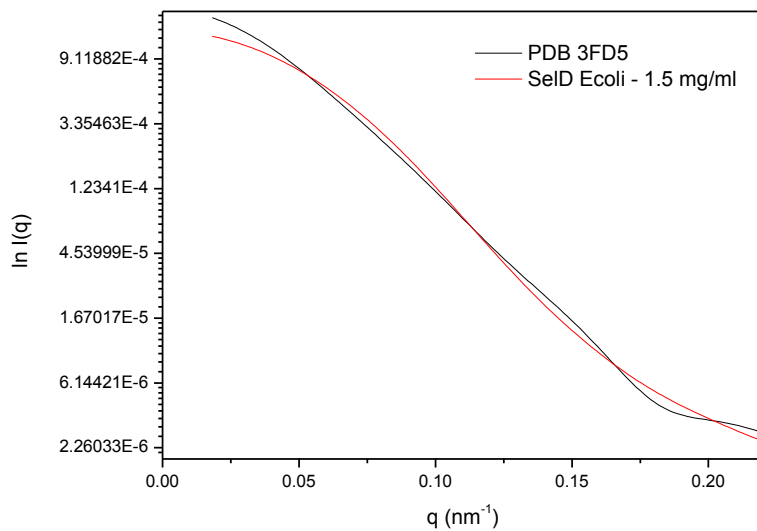


FIGURA 40 - Comparação das curvas de espalhamento de raios-X a baixo ângulo experimental e calculada para o PDB 3FD5.

No caso do modelo por homologia da SELD de *E.coli*, na forma dimérica, o ajuste resultou em  $\chi = 8,196 (> 1)$ , como é mostrado a seguir:

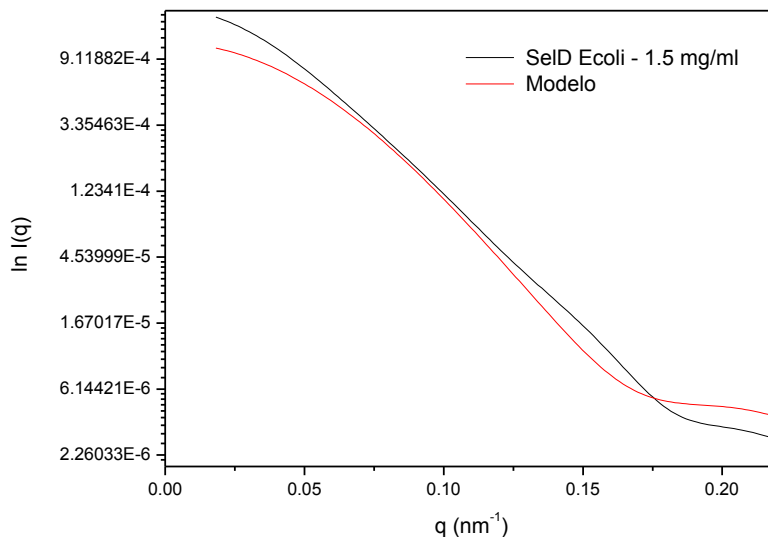


FIGURA 41 - Comparação das curvas de espalhamento de raios-X a baixo ângulo experimental e calculada para o modelo da SELD de *E. coli* obtido através de modelagem molecular.

Os resultados indicam, portanto, que o ajuste com o dímero não está bom. Os dados de SAXS, em geral, apresentam bom ajuste para amostra mais que 90% homogênea, o que pode



não ser o caso desta amostra de proteína a 1,5 mg/ml (40,9  $\mu$ M). Como não se tem nenhum modelo do tetrâmero disponível até o momento, não é possível calcular uma combinação de dímeros e tetrâmeros em solução para se obter uma curva de espalhamento que ajuste à curva experimental. Esta estratégia seria interessante na estimação da porcentagem de cada forma em oligomérica solução, nesta concentração de amostra.

A partir da curva de espalhamento foram gerados modelos *ab initio* do envelope da proteína pelo programa DAMMIN<sup>92</sup>. O resultado obtido após média de 10 modelos é mostrado a seguir, alinhado pelo programa SUPCOMB<sup>92</sup> com o modelo tridimensional obtido por Modelagem por Homologia:

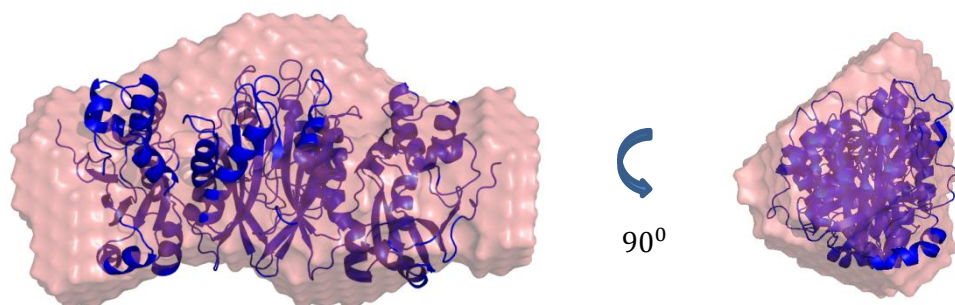


FIGURA 42 - Modelo *Ab Initio* para o envelope molecular da proteína SELD de *E. coli* alinhado com o modelo tridimensional obtido por Modelagem Molecular por meio do programa SUPCOMB<sup>92</sup>.

O valor de  $\chi$ , em média, foi cerca de 1,3 para os 10 modelos, o que se mostra um bom ajuste dos dados no programa DAMMIN<sup>92</sup>. O alinhamento tridimensional dos modelos (FIG. 42) revela um bom ajuste.

#### 4.3.6 Ultracentrifugação Analítica (AUC)

Para uma análise completa das formas oligoméricas da SELD de *E. coli* em solução e determinação de seus parâmetros hidrodinâmicos, realizaram-se os experimentos de AUC, como descrito na seção 3.3.6. O software SEDFIT<sup>97</sup> foi utilizado para resolução da equação de Lamm. O modelo utilizado para ajuste dos dados foi a distribuição  $c(s)$  padrão, que ajusta os dados a partir da razão friccional  $f / f_0$ , também conhecida como fator de Perrin. Deste modo, foram obtidas distribuições de coeficiente de sedimentação aparente,  $s$ , para diferentes

concentrações de amostra de SELD de *E. coli* (FIG. 43). Para evitar erros relacionados à não idealidade da amostra, como variação de temperatura ou mesmo não homogeneidade do tampão, os valores de  $s$  foram transformados para  $s_{20,w}$ , como explicado na seção 2.3.6. Mudanças no  $s_{20,w}$  induzidas por pH, força iônica, ligantes ou temperatura devem ser devidas a mudanças conformacionais.

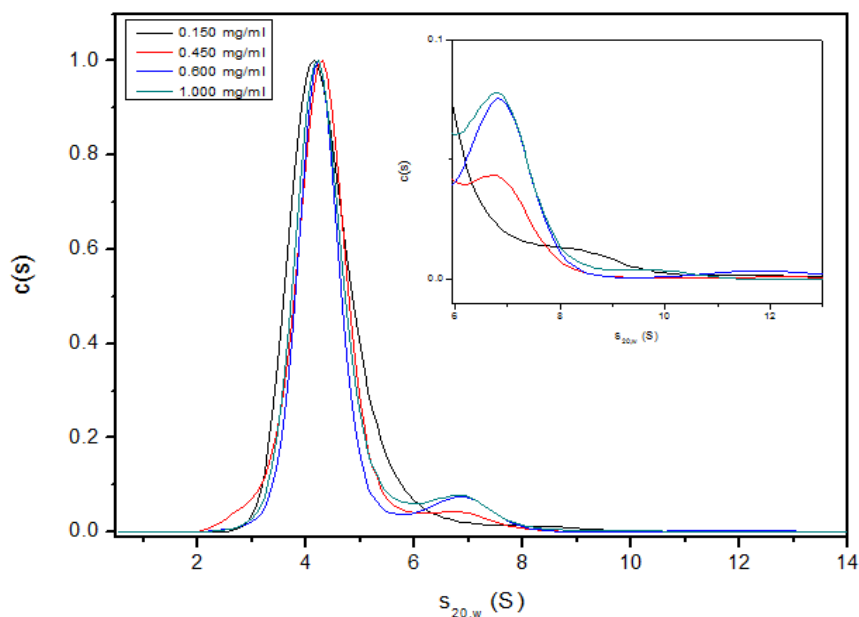


FIGURA 43 - Distribuição de coeficientes de sedimentação,  $c(s)$ . É mostrado um *zoom* na região de 6 S a 13 S (inserido).

Os gráficos da FIG. 43 apresentam três picos, que correspondem a três espécies com coeficientes de sedimentação diferentes em solução. Note que as curvas foram normalizadas para o primeiro pico, que é predominante, de modo a se mostrar o comportamento dos outros dois picos com a concentração da amostra de SELD. Os valores de coeficiente de sedimentação  $s_{20,w}$  para cada partícula nas dadas concentrações são mostrados na FIG. 44, onde também é mostrada uma regressão linear para cada espécie, usada para calcular o valor  $s_{20,w}^0$  (extrapolação para 0), observado para cada partícula. Este procedimento minimiza erros causados por temperatura, viscosidade da solução ou mesmo efeitos de alta concentração.

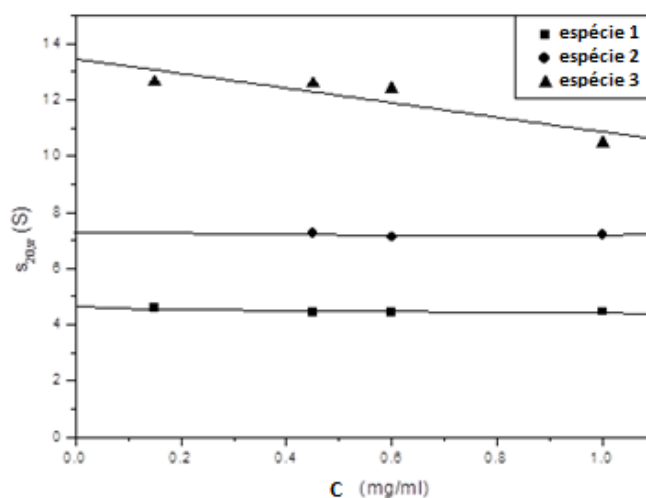


FIGURA 44 - Regressão linear dos coeficientes de sedimentação para as 3 espécies identificadas na distribuição de coeficientes de sedimentação (FIG. 43).

Os valores de  $s_{20,w}^0$ ,  $s_{esfera} / s_{20,w}^0$ , razão friccional e massa molecular calculada a partir da razão friccional para cada partícula são mostrados na TABELA 18:

TABELA 18 - Parâmetros hidrodinâmicos para a SELD de *E. coli* obtidos por SV.

SELD de <i>E. coli</i>	Espécie 1	Espécie 2	Espécie 3
$s_{20,w}^0$ (S)	$4,53 \pm 0,07$	$7,1 \pm 0,4$	$13,6 \pm 0,5$
$s_{esfera} / s_{20,w}^0$	$1,36 \pm 0,01$	$1,38 \pm 0,05$	$1,14 \pm 0,02$
$f / f_0$	$1,43 \pm 0,04$	$1,43 \pm 0,04$	$1,43 \pm 0,04$
$\bar{M}_w$ (kDa) *	$77 \pm 7$	$148 \pm 5$	$328 \pm 40$
$M_w$ (kDa) **	$74 \pm 1$	-	-

\*Massa molecular calculada a partir do coeficiente de Perrin, pelo programa *SEDFIT*<sup>97</sup>.

\*\* Massa molecular calculada usando o coeficiente de difusão translacional obtido por DLS e o coeficiente de sedimentação obtido por AUC.

Portanto, as partículas observadas em solução através desta técnica apresentam-se como dímero, predominantemente, além de tetrâmero e octâmero. As duas primeiras já haviam sido observadas por meio de outras técnicas, porém, curiosamente, observa-se a presença de octâmero em solução através do experimento de SV. Este resultado é um forte

indício de que há um equilíbrio de formas oligoméricas em solução, já que a amostra analisada corresponde ao pico 2 separado por Cromatografia de Exclusão Molecular.

Um resultado importante a ser observado é que o valor do coeficiente de Perrin obtido através do ajuste dos dados de SV e os calculados a partir dos coeficientes de sedimentação experimental para cada espécie ( $s_{esfera} / s_{20,w}^0$ ) mostram uma proteína um pouco alongada, exceto pela espécie 3, cujo resultado indica uma forma bastante globular. Porém, percebe-se, na FIG. 44, que a espécie 3 apresenta um coeficiente de sedimentação bastante baixo para a concentração de 1,0 mg/ml (27,3  $\mu$ M), quando comparado às outras duas concentrações. Este fato pode ter levado a uma subestimação do fator de Perrin.

Os parâmetros hidrodinâmicos e de massa molecular calculados para os estados oligoméricos da SELD de *E. coli* em solução por meio das diferentes técnicas biofísicas empregadas neste trabalho estão resumidos no ANEXO 1, possibilitando sua análise comparativa. Os resultados revelam, portanto, uma mistura de dímeros, tetrâmeros e octâmeros em solução, com parâmetros hidrodinâmicos e massas moleculares comparáveis às esperadas para estes estados oligoméricos. Adicionalmente, suas formas se mostraram alongadas, dado o coeficiente de Perrin calculado a partir dos resultados das diferentes técnicas biofísicas empregadas.

Para se analisar um possível equilíbrio dinâmico entre as três formas oligoméricas da SELD de *E. coli* em solução, foi utilizado o método de Sedimentação em Equilíbrio (SE). Os gráficos de absorvância contra raio da célula para a amostra na concentração de 0,50 mg/ml (13,6  $\mu$ M) e velocidades de rotação de 8000, 10000 e 12000 rpm são mostrados na FIG. 43. A partir deste resultado, foi possível ajustar aos dados um modelo de equilíbrio entre três formas em solução, usando o programa *SEDPHAT*<sup>98</sup>. Os residuais, distribuídos aleatoriamente, são mostrados na FIG. 45, indicando um modelo confiável.

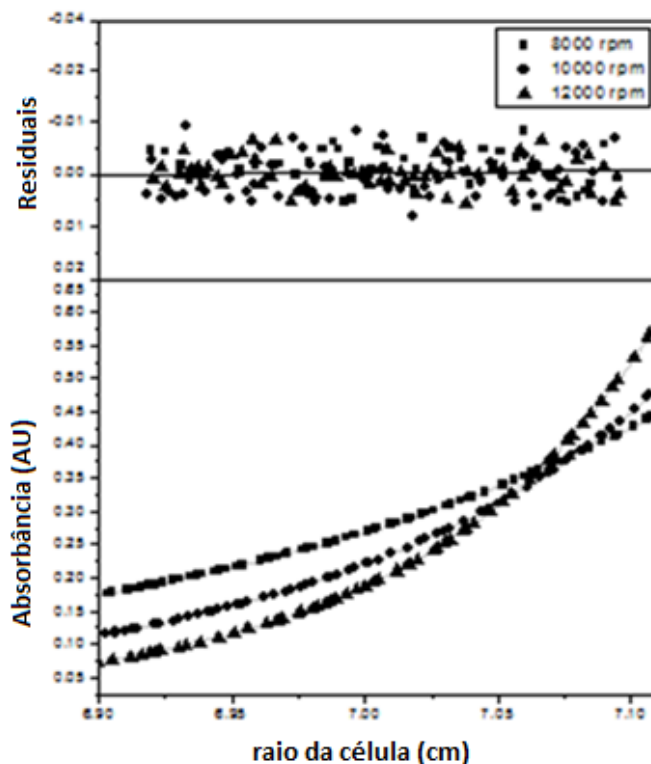


FIGURA 45 – Sedimentação em equilíbrio (SE). São mostradas as curvas de Absorbância da amostra contra o raio da célula e o comportamento dos residuais do modelo utilizado no programa SEDPHAT<sup>98</sup> para ajuste dos dados a um sistema de equilíbrio entre dímero-tetrâmero e tetrâmero-octâmero.

O ajuste ao modelo de equilíbrio termodinâmico entre as espécies resultou em constantes de associação dímero-tetrâmero e tetrâmero-octâmero iguais a  $(14.2 \pm 0.4)$  e  $(1.8 \pm 0.1) \text{ mM}^{-1}$  e, conseqüentemente, constantes de dissociação de  $(70 \pm 2)$  e  $(560 \pm 40) \mu\text{M}$ , respectivamente. O resultado indica um equilíbrio fraco entre as diferentes espécies. Note que a constante de associação dímero-tetrâmero é bastante alta, o que condiz com o fato de observarmos a predominância de dímeros em solução.

### 4.3.7 Ensaios de Cristalização

Com o objetivo de se obter informações experimentais de alta resolução acerca da estrutura tridimensional da SELD de *E. coli*, foram realizadas triagens iniciais de condições de cristalização, como descrito na seção 3.3.7. Cristais na forma de pequenas agulhas (FIG. 44) foram observados em uma semana em uma condição contendo 0,2M AmSO<sub>4</sub>, 0,1M HEPES pH 7,5, 16% (w/v) PEG 4000 e 10% Isopropanol para a amostra de SELD de *E. coli* a 5,0 mg/ml (136,4 μM) na presença de 5 mM ADP. Como os cristais são muito pequenos, não foi possível testá-los para difração de raios-X. Os cristais obtidos não foram reproduzidos a partir de variação de concentração das substâncias da condição de cristalização.

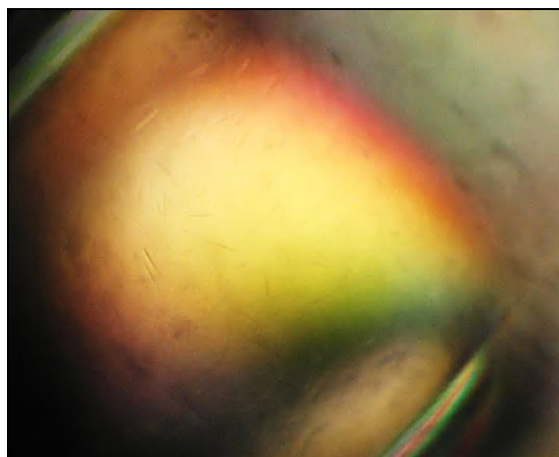


FIGURA 46 - Pequenos cristais na forma de agulhas obtidos em condição de cristalização com 0,2M AmSO<sub>4</sub>, 0,1M HEPES pH 7,5, 16% (w/v) PEG 4000 e 10% Isopropanol para a SELD de *E. coli* na presença de 5 mM ADP.

O resultado da triagem inicial mostra, portanto, que dentre mais de 2000 condições de cristalização diferentes, considerando-se a adição de ligantes da enzima, apenas uma das condições levou à cristalização da proteína, o que é uma porcentagem bastante baixa. Além disso, não foi possível reproduzir tal condição em maior escala. Levando em conta esses resultados, foi utilizada a técnica de proteólise limitada para otimização pré-cristalização.

A protease Quimiotripsina, em uma concentração de 250  $\mu\text{g/ml}$ , levou a um padrão de proteólise estável da SELD de *E. coli*, como se observa pela FIG. 47.



FIGURA 47 - Proteólise da SELD de *E. coli*. 1: Marcador de massa molecular. 2: Quimiotripsina 250  $\mu\text{g/ml}$ . 3: SELD de *E. coli* 5 mg/ml. 4: Produto de proteólise com Quimiotripsina (250  $\mu\text{g/ml}$ ) por 20 minutos. SDS-PAGE 15%. A seta indica o produto de proteólise.

Baseando-se neste resultado, a Quimiotripsina foi adicionada na amostra de SDEC de *E. coli* a 5,0 mg/ml (136,4  $\mu\text{M}$ ), que foi submetida a ensaios de cristalização usando os *kits* comerciais *Crystal Screen*, *Index HT*, *Salt RX* (HAMPTON Research), *PEG Suite* e *PEG II Suite* (Qiagen), varrendo, assim, 480 condições de cristalização. Porém, o experimento de proteólise *in situ* não resultou em cristais. Deste modo, optou-se por purificar o produto de proteólise com Quimiotripsina para posterior ensaio de cristalização.

O cromatograma de purificação do resultado da proteólise por cromatografia de exclusão molecular é mostrado na FIG. 48, em comparação com o cromatograma referente à proteína não-clivada.

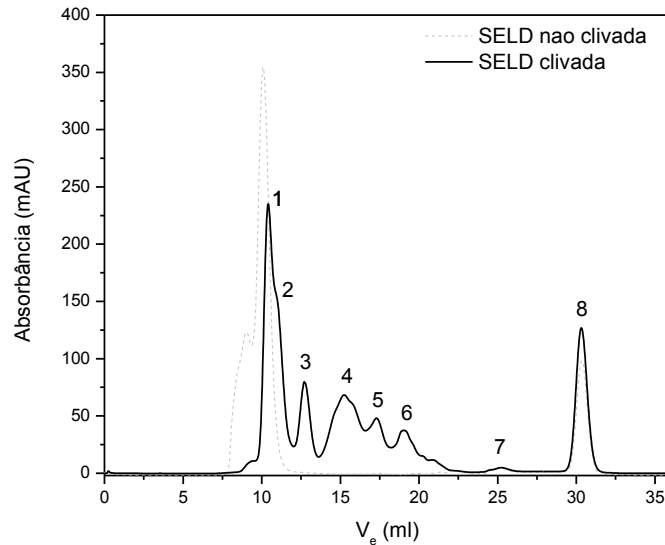


FIGURA 48 - Resultado da purificação do produto de proteólise da proteína SELD *E. coli* através de cromatografia de exclusão molecular (preto) e perfil de eluição de SELD *E. coli* (cinza).

O perfil cromatográfico apresenta dois picos predominantes (picos 1 e 2) no cromatograma, correspondentes à fragmentos de massas moleculares estimadas de 61,7 e 48,0 kDa, respectivamente. Para verificar a homogeneidade da amostra de cada pico, procedeu-se com eletroforese em gel nativo (FIG. 49).

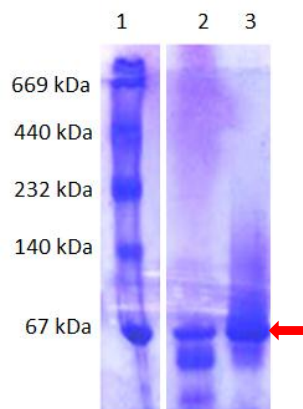


FIGURA 49 - Eletroforese em gel nativo dos picos 1 (coluna 2) e 2 (coluna 3) resultantes da cromatografia de exclusão molecular do produto de proteólise da SELD de *E. coli* com Quimiotripsina. A seta indica o produto de proteólise.

O resultado da eletroforese em gel nativo mostra maior homogeneidade para a amostra do pico 2. Ambas as amostras foram concentradas até 7,0 mg/ml (190,9  $\mu$ M) por ultrafiltração e submetidas à ensaios de cristalização nas mesmas condições descritas para a proteólise *in situ*. Após uma semana, foi possível verificar cristais pequenos em forma de agulha em três condições de cristalização diferentes do kit *Index HT*, mostradas na FIG. 50.



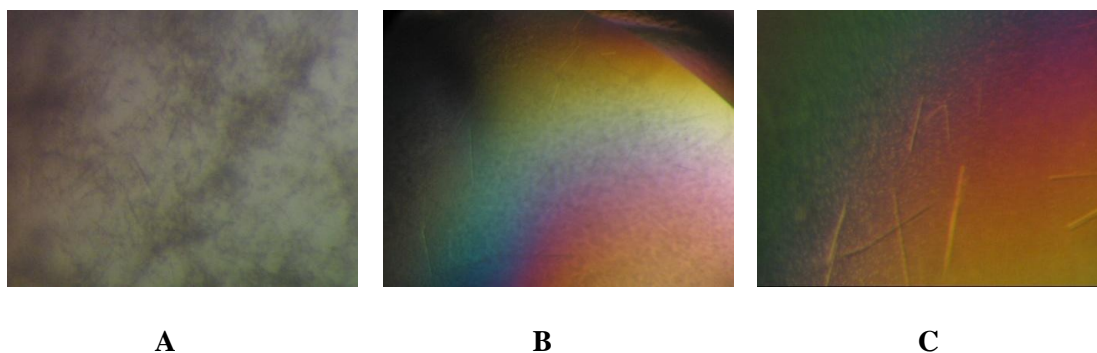
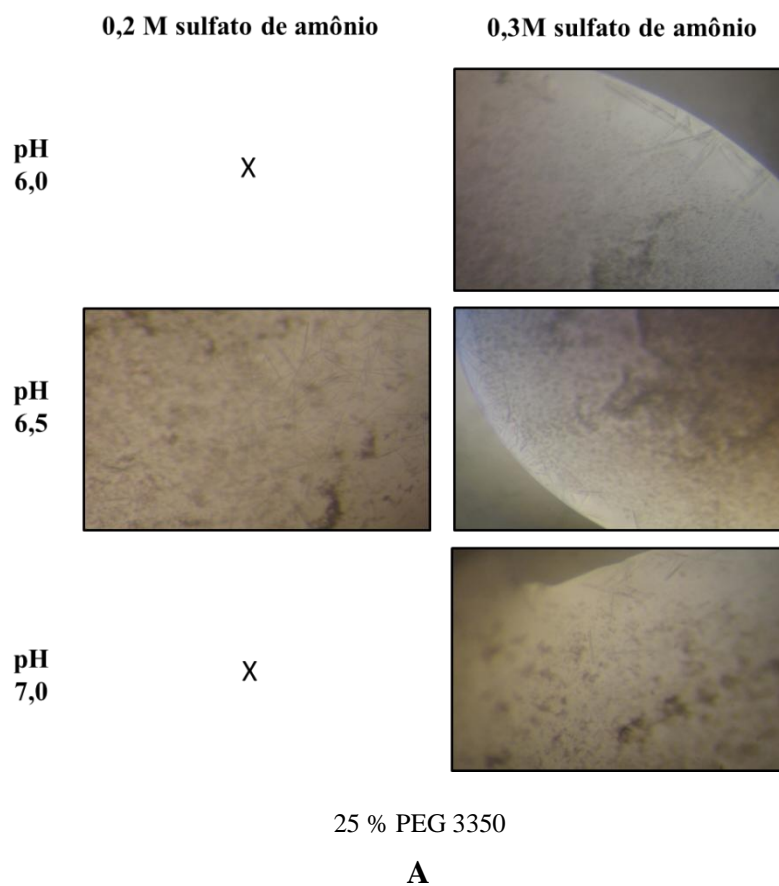


FIGURA 50 - Cristais na forma de agulhas obtidos para a SELD de *E. coli* na ausência de ligantes após proteólise com Quimiotripsina. Condições de cristalização: A- 1,5 M sulfato de amônio / 0,1 M BIS-TRIS pH 6,5 / 0,1 M cloreto de sódio. B- 0,2 M sulfato de amônio / 0,1 M Tris pH 8,5 / 25 % PEG 3350. C – 0,2 M acetato de amônio / 0,1 M HEPES pH 7,5 / 25 % PEG 3350.

A otimização dessas condições se baseou na variação do tipo e concentração de sal entre sulfato e acetato de amônio ou de sódio, além de variação do pH entre 6,0 e 8,5 e de concentração de PEG 3350 entre 20% e 30%. As condições em que se observaram cristais são mostradas a seguir (FIG. 51).



continua

continuação

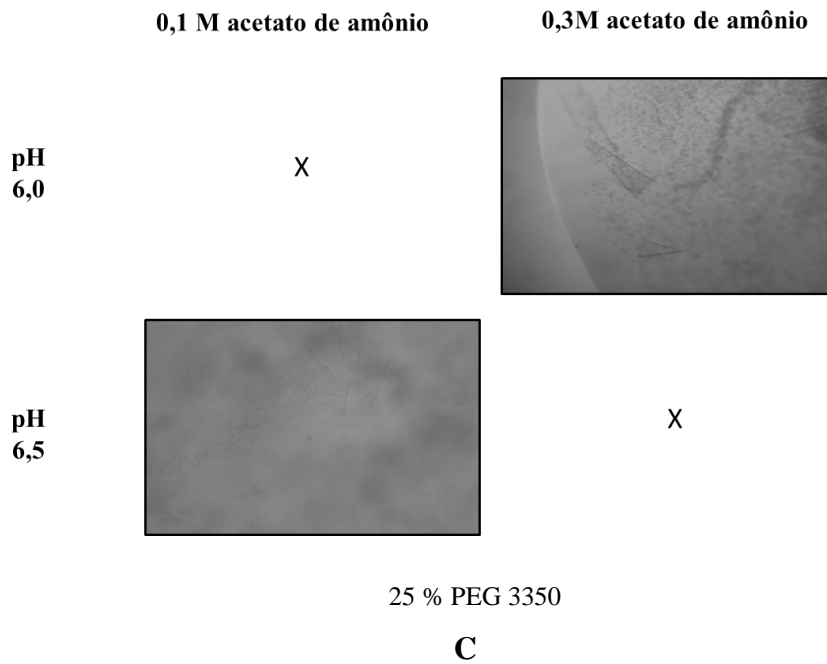
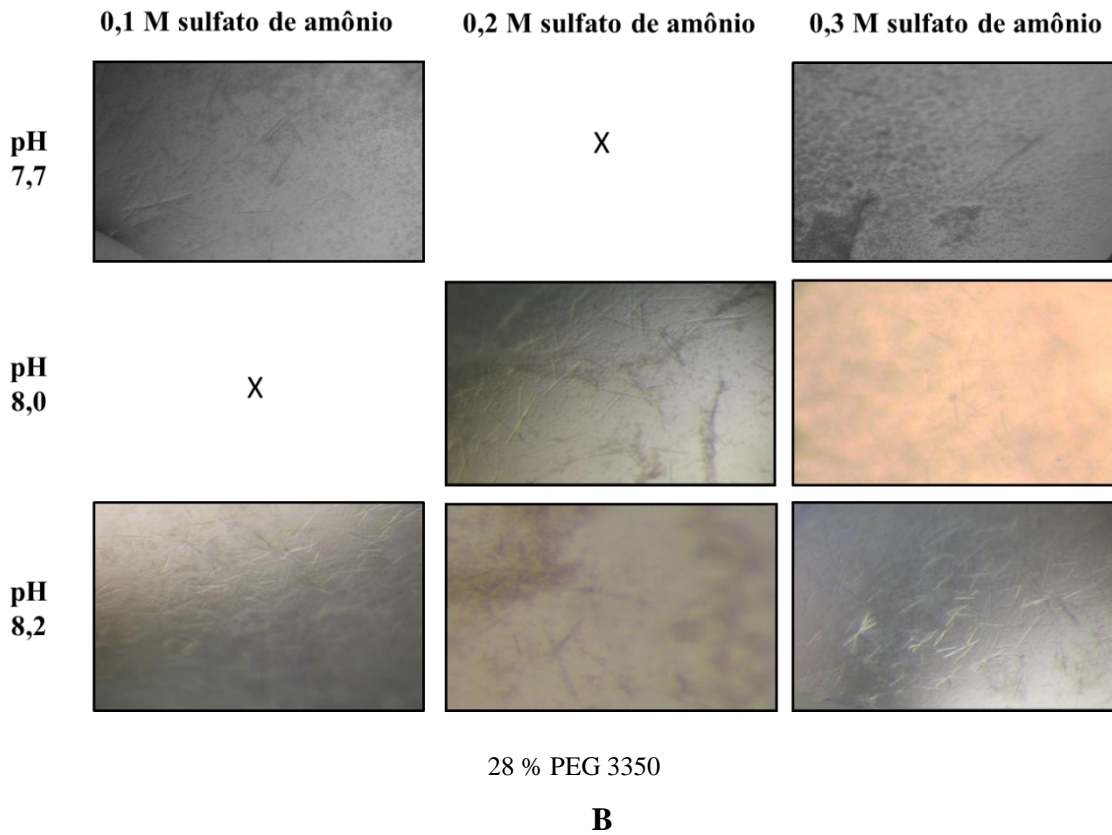


FIGURA 51 - Cristais na forma de agulhas obtidos para a amostra purificada de SELD de *E. coli* após proteólise com Quimiotripsina, na ausência de ligantes. São mostradas as condições de cristalização em que cresceram cristais, todos na forma de agulha, sob variação de pH, concentração de acetato (A e B) e sulfato de amônio (C).

Os resultados indicam que a proteína tende a se cristalizar na forma de pequenas agulhas em concentração de até 0,3 M de acetato ou sulfato de amônio em certos pHs. Os testes com acetato ou sulfato de sódio não levaram à formação de cristais da proteína. Porém, não se observou otimização significativa no tamanho nem no formato dos cristais após variação destes parâmetros. A formação de vários cristais na forma de pequenas agulhas pode ser devida à alta taxa de nucleação e o rápido crescimento dos cristais. Já nas condições em que não se verifica formação de cristais (marcadas com um X na FIG. 47), há formação de precipitado (não mostrado).

Deste modo, uma alternativa para lidar com esta situação seria realizar testes de cristalização em temperaturas menores, nas quais a energia térmica da proteína é menor ou mesmo por meio da técnica de *microbatch*, na qual a taxa de nucleação e o crescimento de cristais são controlados pela adição de uma camada de óleo sobre a gota de cristalização.

Experimentos preliminares de difração de raios-X dos cristais obtidos foram realizados no Laboratório Nacional de Luz Síncrotron (LNLS, Campinas -SP), utilizando-se a linha MX2 como fonte de raios-X no comprimento de onda 1,438 Å, à temperatura de 100K e 25% glicerol como protetor criogênico. O resultado (não mostrado) para todos os testes revelou padrão de difração não característico, não apresentando nenhum ponto de difração no detector. É provável que o resultado seja devido à formação quase-cristais, cuja característica morfológica é de um cristal de proteína, mas o arranjo interno das moléculas não é ordenado como o característico de um cristal.

A aluna de doutorado Lívia Maria Faim, em seu estágio no Laboratório de Cristalografia do professor Dr. Tom Blundell da Universidade de Cambridge (Reino Unido), realizou experimentos de *microseeding* automatizado, baseado na varredura de condições de cristalização pelo método da matriz esparsa, sendo que, neste caso, a solução do reservatório é adicionada a uma gota contendo solução de proteína e solução de sementes (pequenos cristais previamente obtidos), em razão 2:3:1, além de experimentos de *microbatch* com a SELD de *E. coli* em diversas condições de cristalização, porém sem obtenção de monocristal adequado para difração de raios-X com o objetivo de determinação da estrutura.

Novas construções da proteína devem ser feitas, por exemplo, na ausência do N-terminal flexível ou com a cauda de histidinas no C-terminal, com o objetivo de se obter uma proteína recombinante possível de ser cristalizada.

## 4.4 Estudo da interação da SELD de *E. coli* com o complexo SELA-SELC

### 4.4.1 Anisotropia de Fluorescência

A técnica de anisotropia de fluorescência foi utilizada para monitorar a titulação de SELD à solução contendo 5  $\mu\text{M}$  do complexo de SELA-SELC (1:1), sendo 10 nM de SELC marcado com fluoresceína. O resultado é mostrado na FIG. 52.

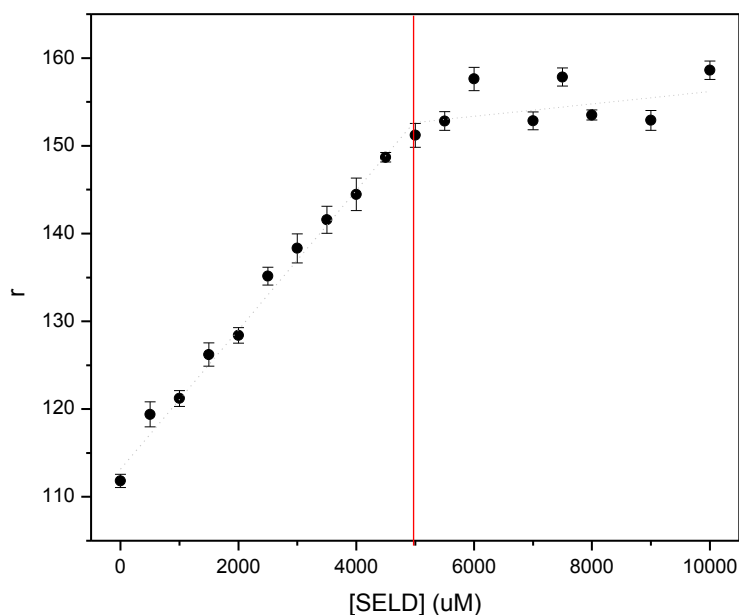


FIGURA 52 – Variação da anisotropia de fluorescência com a titulação de SELD de *E. coli* em amostra contendo 5  $\mu\text{M}$  do complexo SELA-SELC, sendo 10 nM de SELC marcado com fluoresceína. A linha em vermelho indica a razão estequiométrica de 1:1 (mol de monômeros de SELD : mol do complexo SELA-SELC). As barras de erro representam erros intrínsecos a cada medida no equipamento utilizado.

Observa-se um aumento significativo na anisotropia de fluorescência com a adição de SELD à solução, partindo de 110 até 160. O resultado, portanto, indica a interação da SELD com o complexo SELA-SELC presente em solução, já que um aumento na anisotropia de fluorescência é devido à menor velocidade de rotação do complexo quando comparada à situação em ausência de SELD. Próximo à concentração de 5  $\mu\text{M}$  de SELD parece ocorrer uma saturação do sistema e a partir desse ponto não há mudança significativa nos valores de na anisotropia de fluorescência. Como a concentração de SELA-SELC em solução é de 5  $\mu\text{M}$ , a estequiometria de ligação à SELD é estimada, portanto, como sendo de 1:1:1, em mols de monômeros.

#### 4.4.2 Espectrometria de massas com troca H/D

A partir do resultado de interação da proteína SELD de *E. coli* com o complexo SELA-SELC observado por meio da técnica de anisotropia de fluorescência, utilizou-se a técnica de espectrometria de massas com troca H/D para análise de possíveis regiões da SELD de *E. coli* que apresentam alterações na taxa de incorporação de deutério após interação com o complexo, indicando a possível região de interação.

Iniciou-se o experimento de espectrometria de massas com troca H/D com a amostra de SELD de *E. coli* pura. A deconvolução do espectro de massas da proteína intacta resultou em dois picos (FIG. 53), sendo que pico 2 corresponde a uma massa molecular de  $(72465 \pm 29)$  Da e o pico 1 corresponde a  $(36232 \pm 14)$  Da. Portanto, verificou-se uma mistura entre dímeros e monômeros, com erro relativo para a massa teórica do dímero e do monômero igual a 1,2 % para ambos os resultados. Apesar da detecção de monômero em solução, deve-se levar em conta que as condições experimentais, incluindo pH ácido igual a 2,5, podem alterar o equilíbrio dímero-tetrâmero-octâmero observado em pH 8,0 por ultracentrifugação analítica, resultando na formação de monômeros em solução, detectados por espectrometria de massas.

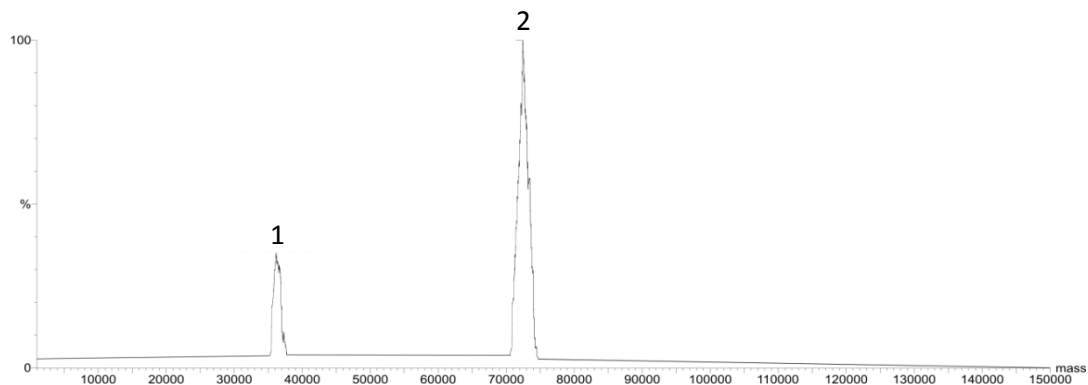


FIGURA 53 – Resultado da deconvolução do espectro de massas da SELD de *E. coli* apresentando dois picos, 1 e 2, correspondentes a massas moleculares de  $36232 \pm 14$  e  $72465 \pm 29$  Da, respectivamente.

A proteólise com pepsina da SELD de *E. coli* resultou em 41 peptídeos, cobrindo 68,7 % da sequência de aminoácidos da proteína. Foram realizados experimentos em diferentes tempos de incorporação de deutério para a amostra de proteína pura e com 30 minutos de incorporação de deutério em amostra após interação com o complexo SELA-SELC. O resultado é mostrado na FIG. 54. Os anexos 2 e 3 mostram os peptídeos detectados e sua porcentagem de incorporação de deutério para cada intervalo de tempo.

1 47  
 SELD - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE

t<sub>5</sub> - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE  
 t<sub>10</sub> - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE  
 t<sub>15</sub> - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE  
 t<sub>30</sub> - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE  
 int - MSENSIRLTQYSHGAGCGCKISPKVLETILHSEQAKFVDPNLLVGNE

48 94  
 SELD - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA  
 \* \* \* \* \* \* \* \*

# #  
 t<sub>5</sub> - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA  
 t<sub>10</sub> - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA  
 t<sub>15</sub> - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA  
 t<sub>30</sub> - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA  
 int - TRDDAAVYDLGNGT~~SV~~ISTTDF~~F~~MP~~I~~VDN~~P~~FDFGRIAATNAISDIFA

95 141  
 SELD - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS  
 \* \* \* \*

###  
 t<sub>5</sub> - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS  
 t<sub>10</sub> - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS  
 t<sub>15</sub> - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS  
 t<sub>30</sub> - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS  
 int - MGGKPI~~MA~~IA~~IL~~GWP~~I~~NKLS~~P~~EIA~~R~~EVTEGGRYACRQAGIALAGGHS

142 188  
 SELD - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A  
 \* \* \*

t<sub>5</sub> - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A  
 t<sub>10</sub> - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A  
 t<sub>15</sub> - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A  
 t<sub>30</sub> - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A  
 int - IDAPEPI~~F~~GLAVT~~G~~I~~V~~PT~~E~~RVKKNST~~A~~QAGCK~~L~~F~~L~~T~~K~~PLG~~I~~~~G~~V~~L~~T~~T~~A

continua

continuação

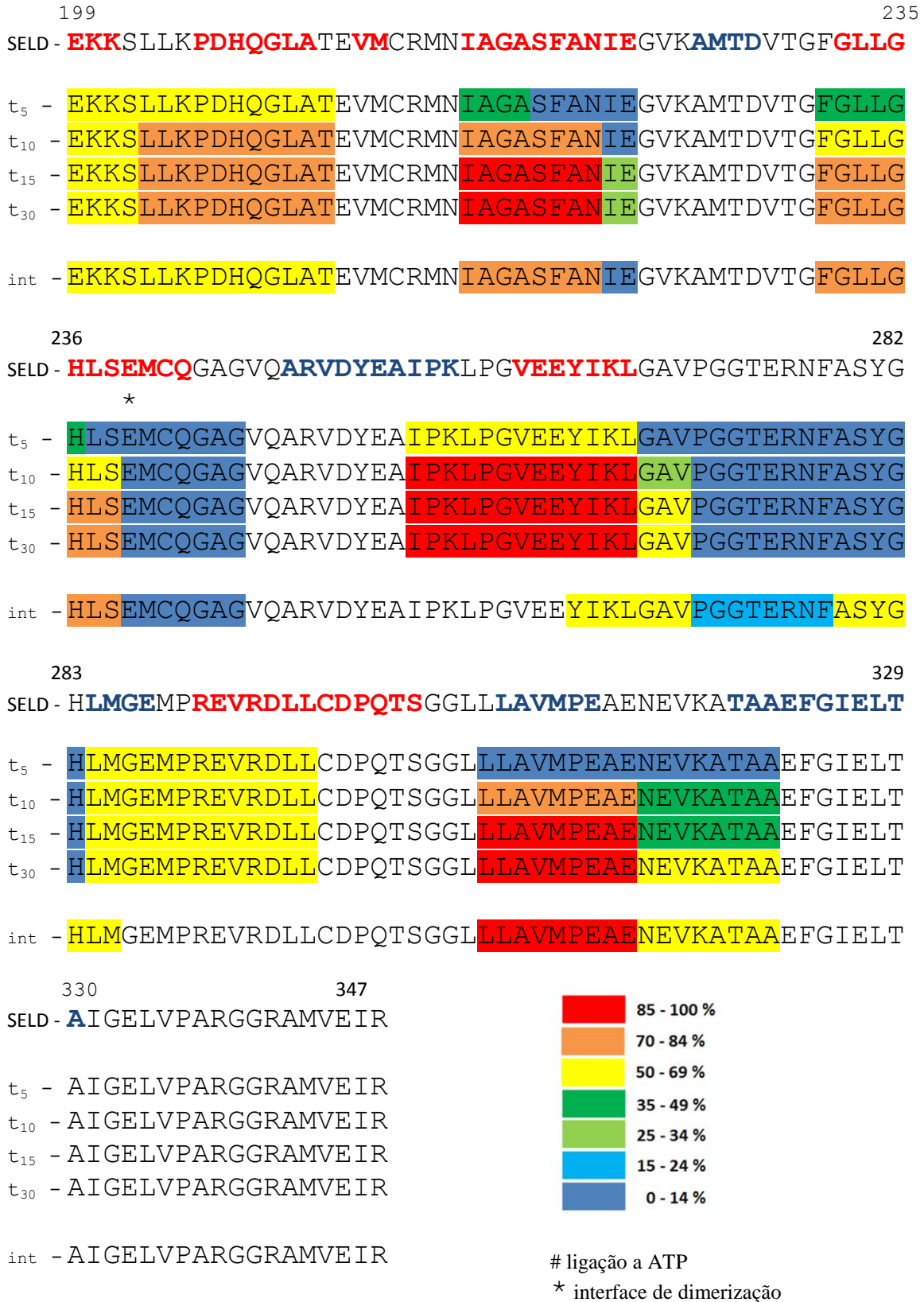


FIGURA 54 – Perfil de incorporação de deutério na SELD de *E. coli* em diferentes tempos de troca H/D:  $t_5 = 5$  min,  $t_{10} = 10$  min,  $t_{15} = 15$  min,  $t_{30} = 30$  min,  $int = 30$  min (no complexo SELD-SELA-SELC). A incorporação de deutério nas amostras  $t_5$ ,  $t_{10}$ ,  $t_{15}$  e  $t_{30}$  foi de 68,7 % enquanto que para a amostra  $int$  foi de 62,5 %. A legenda indica faixas de porcentagem de incorporação de deutério na estrutura da proteína. Em cinza estão marcados os resíduos de aminoácidos catalíticos.

O resultado obtido mostra, de maneira geral, que grande parte da proteína é acessível ao solvente, dada a alta taxa de incorporação de deutério observada para a amostra de proteína pura, o que concorda com o resultado do servidor *PISA*<sup>77</sup> (seção 4.1.2) para o modelo tridimensional da SELD de *E. coli*. A análise do resultado de incorporação de deutério em diferentes intervalos de tempo também mostra que a região N-terminal da proteína é acessível ao solvente e bastante flexível, sendo observada desde o intervalo de 5 minutos de incorporação de deutério uma alta taxa de troca H/D.

Regiões da sequência de aminoácidos da SELD entre os resíduos 136 e 146, 239 e 245, 271 e 283 não apresentaram incorporação significativa de deutério, mesmo após 30 minutos de incorporação. A análise destas regiões revela que estes resíduos se encontram interiorizados na proteína, com pouco acesso ao solvente e, por isso, apresentam baixa taxa de troca H/D. A região entre os resíduos 136 e 146, além do resíduo Glu(239), é predita como parte da interface de dimerização, como analisado na seção 4.1.2. Observa-se, ainda, que três dos resíduos preditos como responsáveis pela interação com ATP encontram-se em região com baixa taxa de incorporação de deutério (~25 % após 15 minutos), correspondendo a uma região de fita- $\beta$  como estrutura secundária.

Comparado ao resultado de incorporação de deutério da amostra de SELD pura, a proteína em interação com o complexo SELA-SELC mostrou uma taxa global de incorporação de deutério menor (62,5 %), o que indica que regiões da proteína ficaram inacessíveis ao solvente após a interação, como era de se esperar. Um resultado bastante interessante é que a região N-terminal da proteína (resíduos de 1 a 9) (FIG. 54 e 55) apresenta uma taxa de incorporação de deutério bastante baixa quando a SELD está em complexo com SELA-SELC, sendo que, quando pura, a taxa de incorporação é bastante alta mesmo para 5 minutos de incorporação de deutério. Este resultado indica que esta região fica inacessível ao solvente após a interação.

Além disso, a região entre os resíduos 43 e 54, correspondente a uma alça na superfície proteica próxima ao sítio catalítico (FIG. 54 e 55), também passa a apresentar uma baixa incorporação de deutério após a interação. Este é um possível sítio de interação com o complexo SELA-SELC. Outra região que também passa a apresentar baixa incorporação de deutério após interação com o complexo SELA-SELC é a compreendida entre os resíduos 120 e 129. Esta região no modelo da estrutura tridimensional da SELD está próxima da região entre os resíduos 43 e 54, o que indica novamente que esta região é um possível sítio de interação com o complexo SELA-SELC (FIG. 54 e 55).



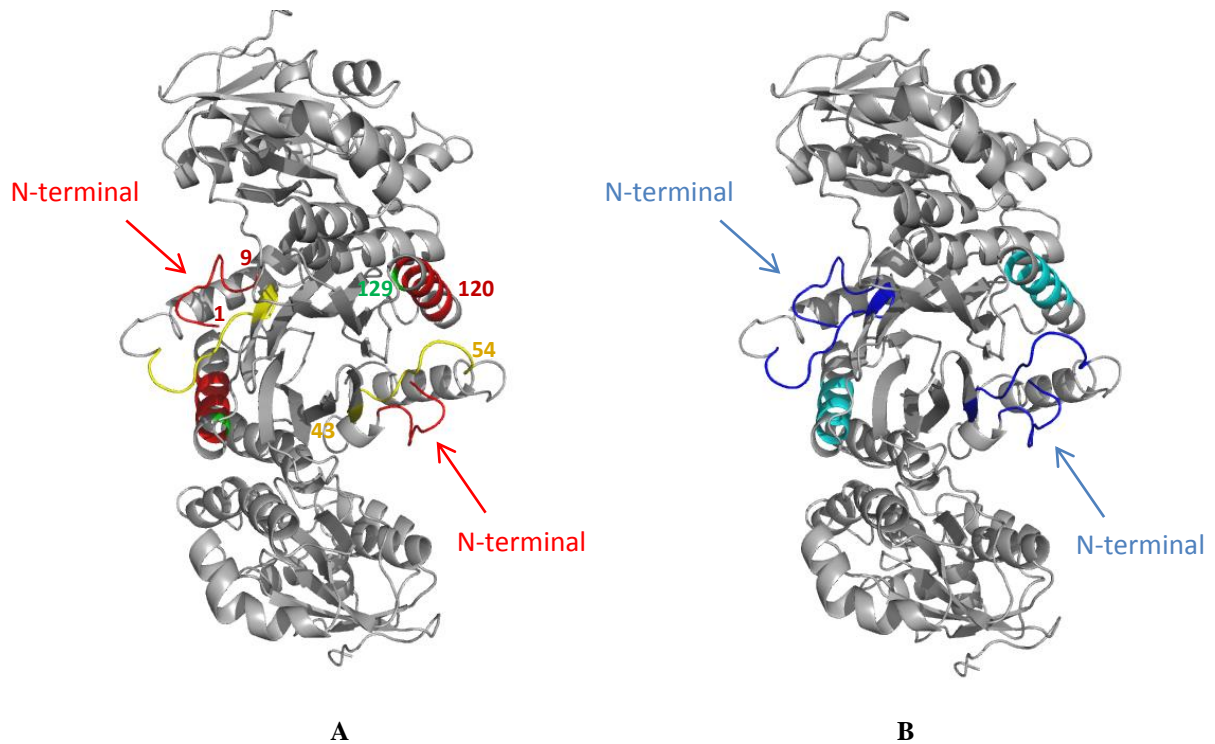


FIGURA 55 - Perfil de incorporação de deutério na SELD de *E. coli* com 30 minutos de troca H/D, para A – amostra pura de proteína e B – após interação com o complexo SELA-SEL, indicando apenas as regiões em que houve grande variação na taxa de incorporação de deutério. As setas indicam os N-terminais de cada monômero. Em azul, são representadas regiões com porcentagem de incorporação de deutério de 0 a 14 %, enquanto que em ciano a porcentagem de incorporação é de 15 a 24 %, em amarelo é de 50 a 69 % e em vermelho é de 85 a 100 %. Os números apresentados em A indicam a posição dos resíduos de aminoácidos nas regiões assinaladas, com suas respectivas cores (esses números são indicados em apenas 3 regiões, dado que as outras são simétricas às primeiras).

O anexo 4 mostra diferentes disposições do modelo de SELD de *E. coli*, onde se pode observar as regiões com maior porcentagem de incorporação de deutério, que indicam regiões bastante expostas ao solvente, bem como regiões com alta flexibilidade, que já apresentam alta taxa de incorporação de deutério para pequeno intervalo de tempo de incorporação.



## 5 Conclusões

Neste trabalho são apresentados os resultados de clonagem, expressão e purificação da proteína recombinante Selenofosfato Sintetase de *E. coli* e a caracterização do dímero como sua forma oligomérica predominante em solução. Adicionalmente, a interação da SELD com o complexo SELA-SELC é investigada por meio de Anisotropia de Fluorescência e Espectrometria de Massas com troca H/D.

Um estudo através de ferramentas de Bioinformática revela que a proteína pertence à família AIRS, formada por proteínas que tem ATP como substrato, com conservação de aminoácidos com esta função. Além disso, observou-se a conservação de aminoácidos responsáveis pela formação de interface de dimerização, sendo que algumas regiões contendo esses aminoácidos foram observadas por Espectrometria de Massas apresentando baixa incorporação de deutério, o que significa que possivelmente estão inacessíveis ao solvente devido à formação dessa interface.

A Cromatografia de Exclusão Molecular indicou a presença de uma mistura de dímeros e tetrâmeros da proteína em solução, com predominância de dímeros, o que foi confirmado através de Eletroforese em Gel Nativo. A observação de dímeros em solução já era esperada, já que aminoácidos responsáveis pela dimerização são conservados na família de proteínas AIRS. As propriedades hidrodinâmicas da amostra de SELD de *E. coli* foram estudadas por DLS, SAXS e AUC, sendo estimados os coeficientes de difusão translacional e de sedimentação das espécies em solução.

O fator de Perrin calculado utilizando tanto os coeficientes de difusão quanto dos coeficientes de sedimentação experimentais indica que as espécies são alongadas, como também observado para o modelo tridimensional do homodímero da proteína e para o modelo de SAXS. O experimento de AUC por velocidade de sedimentação revelou a existência de uma pequena concentração de octâmeros em solução, além das formas dimérica e tetramérica observadas por meio das outras técnicas biofísicas. O experimento de AUC por sedimentação em equilíbrio confirmou a existência de um equilíbrio entre oligômeros em solução, estimando as constantes de associação e dissociação entre dímeros e tetrâmeros e entre tetrâmeros e octâmeros em de  $(70 \pm 2)$  e  $(560 \pm 40) \mu\text{M}$ , respectivamente, sendo evidente um equilíbrio fraco entre essas formas oligoméricas. Adicionalmente, realizaram-se ensaios de cristalização da SELD de *E. coli* e foram obtidos pequenos cristais em forma de agulha em condição a ser otimizada.

A interação entre SELD e o complexo SELA-SELC foi verificada monitorando-se a variação da anisotropia de fluorescência ao se adicionar SELD em amostra contendo o complexo SELA-SELC, resultando em uma estequiometria estimada de 1:1. Uma análise mais detalhada da interação por meio de Espectrometria de Massas com troca H/D revela possíveis regiões de interação entre SELD e SELA-SELC, próximas aos dois sítios catalíticos do homodímero de SELD. Além disso, verifica-se que, após a interação, a região N-terminal da SELD fica inacessível ao solvente, o que condiz com a possível entrega à SELA do composto biologicamente ativo do selênio, que é tóxico para a célula.

## 6 Perspectivas

- Realizar ensaio de atividade com a SELD de *E. coli* em suas formas dimérica e tetramérica, caso seja possível isolá-las em solução.
- Proceder com os testes de cristalização da proteína com construções truncadas no N-terminal flexível e também com mutantes dos aminoácidos catalíticos Cys(17) e Lys(20).
- Investigar diferentes condições de interação da SELD com o complexo SELA-SELC, como alta força iônica e a presença de ligantes de SELD, como o ATP e o ADP.
- Verificar se a interação de SELD-SELA é dependente do SELC por meio de Calorimetria de Titulação Isotérmica (ITC).
- Analisar a estrutura do complexo SELD-SELA-SELC por meio de SAXS, Microscopia de Força Atômica, Microscopia Eletrônica de Transmissão (*Negative Staining*) e Criomicroscopia eletrônica.



## Referências

- 1 JOHANSSON, L.; GAFVELIN, G.; ARNER, E. S. J. Selenocysteine in proteins: properties and biotechnological use. **Biochimica Et Biophysica Acta-General Subjects**, Amsterdam, v. 1726, n. 1, p. 1-13, 2005.
- 2 PAPP, L. V. et al. From selenium to selenoproteins: synthesis, identity, and their role in human health. **Antioxidants & Redox Signaling**, New Rochelle, v. 9, n. 7, p. 775-806, 2007.
- 3 BERZELIUS, J. J. Lettre de M. Berzelius à M. Berthollet sur deux métaux nouveaux. **Annales de Chimie et de Physique**. Série 2, v. 7, p. 199-202, 1818.
- 4 BOYD, R. Selenium stories. **Nature Chemistry**. Halifax, v. 3, n. 7, p. 570-570, 2011.
- 5 RAYMAN, M. P. The importance of selenium to human health. **Lancet**, London, v. 356, n. 9225, p. 233-241, 2000.
- 6 VALDIGLESIAS, V. et al. In vitro evaluation of selenium genotoxic, cytotoxic, and protective effects: a review. **Archives of Toxicology**, La Coruna, v. 84, n. 5, p. 337-351, 2010.
- 7 SCHWARZ, K.; FOLTZ, C. M. Selenium as an integral part of factor-3 against dietary necrotic liver degeneration. **Journal of the American Chemical Society**, Washington, v. 79, n. 12, p. 3292-3293, 1957.
- 8 PINSENT, J. The need for selenite and molybdate in the formation of formic dehydrogenase by members of the coli-aerogenes group of bacteria. **Biochemical Journal**, London, v. 57, n. 1, p. 10-16, 1954.
- 9 LEDUC, D. L. et al. Overexpression of selenocysteine methyltransferase in Arabidopsis and Indian mustard increases selenium tolerance and accumulation. **Plant Physiology**, Rockville, v. 135, n. 1, p. 377-383, 2004.
- 10 KOHRLE, J. et al. Selenium in biology: facts and medical perspectives. **Biological Chemistry**, Berlin, v. 381, n. 9-10, p. 849-864, 2000.
- 11 BRASIL. Ministério da Saúde. Agência Nacional de Vigilância Sanitária (ANVISA). Resolução RDC n. 269, de 22 de setembro de 2005: Regulamento técnico sobre a ingestão

diária recomendada (idr) de proteína, vitaminas e minerais. **Diário Oficial da União**, Brasília,DF, 23 set. 2005.

12 MORENO-REYES, R. et al. Kashin-Beck osteoarthropathy in rural Tibet in relation to selenium and iodine status. **New England Journal of Medicine**, Boston, v. 339, n. 16, p. 1112-1120, 1998.

13 VANDERPAS, J. B. et al. Iodine and selenium deficiency associated with cretinism in northern zaire. **American Journal of Clinical Nutrition**, Bethesda, v. 52, n. 6, p. 1087-1093, 1990.

14 HATFIELD, D. L.; BERRY, M. J.; GLADYSHEV, V. N. **Selenium: its molecular biology and role in human health**. New York: Springer- Verlag, 2006.

15 COMBS, G. F.; COMBS, S. B. The nutritional biochemistry of selenium. **Annual Review of Nutrition**, Palo Alto, v. 4, p. 257-280, 1984.

16 KRYUKOV, G. V. et al. Characterization of mammalian selenoproteomes. **Science**, Washington, v. 300, n. 5624, p. 1439-1443, 2003.

17 ROTRUCK, J. T. et al. Selenium: biochemical role as a component of glutathione peroxidase. **Science**, Washington, v. 179, n. 4073, p. 588-590, 1973.

18 FLOHE, L.; GUNZLER, W. A.; SCHOCK, H. H. Glutathione peroxidase: selenoenzyme. **Febs Letters**, Amsterdam, v. 32, n. 1, p. 132-134, 1973.

19 CONE, J. E. et al. Chemical characterization of selenoprotein component of clostridial glycine reductase: identification of selenocysteine as organoselenium moiety. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 73, n. 8, p. 2659-2663, 1976.

20 LEINFELDER, W. et al. In vitro synthesis of selenocysteinyl-transfer RNA<sub>Auca</sub> from seryl-transfer RNA<sub>Auca</sub>: involvement and characterization of the seld gene-product. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 87, n. 2, p. 543-547, 1990.

21 KRYUKOV, G. V.; GLADYSHEV, V. N. The prokaryotic selenoproteome. **Embo Reports**, London, v. 5, n. 5, p. 538-543, 2004.



22 LOBANOV, A. V. et al. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. **Genome Biology**, London, v. 8, n. 9, p. 1-16, 2007.

23 COSTA, F. C. et al. Oxidative stress protection of Trypanosomes requires selenophosphate synthase. **Molecular and Biochemical Parasitology**, Amsterdam, v. 180, n. 1, p. 47-50, 2011.

24 HERAS, I. L.; PALOMO, M.; MADRID, Y. Selenoproteins: the key factor in selenium essentiality. State of the art analytical techniques for selenoprotein studies. **Analytical and Bioanalytical Chemistry**, Heidelberg, v. 400, n. 6, p. 1717-1727, 2011.

25 CHAMBERS, I. et al. The structure of the mouse glutathione-peroxidase gene: the selenocysteine in the active-site is encoded by the termination codon, TGA. **Embo Journal**, Oxford, v. 5, n. 6, p. 1221-1227, 1986.

26 ZINONI, F. et al. Nucleotide-sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 83, n. 13, p. 4650-4654, 1986.

27 LEINFELDER, W. et al. Gene for a novel transfer-RNA species that accepts L-serine and cotranslationally inserts selenocysteine. **Nature**, London, v. 331, n. 6158, p. 723-725, 1988.

28 BOCK, A. et al. Selenoprotein synthesis: an expansion of the genetic-code. **Trends in Biochemical Sciences**, Oxford, v. 16, n. 12, p. 463-467, 1991.

29 MARVINSKETCH. Advanced chemical drawing software. Disponível em: < <http://www.chemaxon.com/products/marvin/marvinsketch/> >. Acesso em: 15 out. 2011.

30 PETERSEN, E. F. et al. UCSF chimera - A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, Hoboken, v. 25, n. 13, p. 1605-1612, 2004.

31 ARNER, E. S. J. Selenoproteins-What unique properties can arise with selenocysteine in place of cysteine? **Experimental Cell Research**, San Diego, v. 316, n. 8, p. 1296-1303, 2010.

32 FOMENKO, D. E.; MARINO, S. M.; GLADYSHEV, V. N. Functional diversity of cysteine residues in proteins and unique features of catalytic redox-active cysteines in thiol oxidoreductases. **Molecules and Cells**, Seoul, v. 26, n. 3, p. 228-235, 2008.

33 CASTELLANO, S. On the unique function of selenocysteine - Insights from the evolution of selenoproteins. **Biochimica Et Biophysica Acta-General Subjects**, Amsterdam, v. 1790, n. 11, p. 1463-1470, 2009.

34 ALLMANG, C.; WURTH, L.; KROL, A. The selenium to selenoprotein pathway in eukaryotes: More molecular partners than anticipated. **Biochimica Et Biophysica Acta-General Subjects**, Amsterdam, v. 1790, n. 11, p. 1415-1423, 2009.

35 SCHON, A. et al. The selenocysteine-inserting opal suppressor serine transfer-RNA from *Escherichia coli* is highly unusual in structure and modification. **Nucleic Acids Research**, Oxford, v. 17, n. 18, p. 7159-7165, 1989.

36 FORCHHAMMER, K.; BOCK, A. Selenocysteine synthase from *Escherichia coli*: analysis of the reaction sequence. **Journal of Biological Chemistry**, Bethesda, v. 266, n. 10, p. 6324-6328, 1991.

37 GLASS, R. S. et al. Monoselenophosphate: synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound sepx. **Biochemistry**, Washington, v. 32, n. 47, p. 12555-12559, 1993.

38 BARON, C. et al. Solution structure of selenocysteine-inserting transfer RNA(Sec) from *Escherichia coli*: comparison with canonical transfer RNA(Ser). **Journal of Molecular Biology**, London, v. 231, n. 2, p. 274-292, 1993.

39 STOCK, T.; SELZER, M.; ROTHER, M. In vivo requirement of selenophosphate for selenoprotein synthesis in archaea. **Molecular Microbiology**, Malden, v. 75, n. 1, p. 149-160, 2010.

40 THANBICHLER, M.; BOCK, A. The function of SECIS RNA in translational control of gene expression in *Escherichia coli*. **Embo Journal**, Oxford, v. 21, n. 24, p. 6925-6934, 2002.

41 PALESKAVA, A.; KONEVEGA, A. L.; RODNINA, M. V. Thermodynamic and kinetic framework of selenocysteyl-tRNA(Sec) recognition by elongation factor SelB. **Journal of Biological Chemistry**, Bethesda, v. 285, n. 5, p. 1-23, 2010.

42 ZHANG, Y.; GLADYSHEV, V. N. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. **Bioinformatics**, Oxford, v. 21, n. 11, p. 2580-2589, 2005.

43 LACOURCIERE, G. M. Selenium is mobilized in vivo from free selenocysteine and is incorporated specifically into formate dehydrogenase H and tRNA Nucleosides. **Journal of Bacteriology**, Washington, v. 184, n. 7, p. 1940-1946, 2002.

44 MULLER, S.; HEIDER, J.; BOCK, A. The path of unspecific incorporation of selenium in *Escherichia coli*. **Archives of Microbiology**, Munich, v. 168, n. 5, p. 421-427, 1997.

45 ROTHER, M. et al. Selenoprotein synthesis in archaea. **Biofactors**, Amsterdam, v. 14, n. 1-4, p. 75-83, 2001.

46 CHAPPLE, C. E.; GUIGO, R. Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. **Plos One**, San Francisco, v. 3, n. 8, p. 1-14, 2008.

47 COPELAND, P. R. et al. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. **Embo Journal**, Oxford, v. 19, n. 23, p. 6634-6634, 2000.

48 SMALL-HOWARD, A. et al. Supramolecular complexes mediate selenocysteine incorporation in vivo. **Molecular and Cellular Biology**, Washington, v. 26, n. 6, p. 2337-2346, 2006.

49 STOCK, T.; ROTHER, M. Selenoproteins in Archaea and Gram-positive bacteria. **Biochimica Et Biophysica Acta-General Subjects**, Amsterdam, v. 1790, n. 11, p. 1520-1532, 2009.

50 HEIDER, J.; BOCK, A. Selenium metabolism in microorganisms. **Advances in Microbial Physiology**, London, v. 35, p. 71-109, 1993.

51 ZANNONI, D. et al. The bacterial response to the chalcogen metalloids Se and Te. In: POOLE, R. K. (Ed.). **Advances in microbial physiology**, San Diego: Elsevier Academic Press Inc. , 2008. v. 53. ISBN 0065-2911978-0-12-373713-7.

52 MATSUMOTO, E. et al. Structure of an N-terminally truncated selenophosphate synthetase from *Aquifex aeolicus*. **Acta Crystallographica Section F-Structural Biology and Crystallization Communications**, Oxford, v. 64, p. 453-458, 2008. DOI: 10.1107/S1744309108012074.

53 LI, C. L. et al. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the Escherichia coli purine biosynthetic pathway at 2.5 angstrom resolution. **Structure with Folding & Design**, London, v. 7, n. 9, p. 1155-1166, 1999.

54 MORAR, M. et al. Complexed structures of formylglycinamide ribonucleotide amidotransferase from Thermotoga maritima describe a novel ATP binding protein superfamily. **Biochemistry**, Washington, v. 45, n. 50, p. 14880-14895, 2006.

55 RANGARAJAN, E. S. et al. Structure of NiFe hydrogenase maturation protein HypE from Escherichia coli and its interaction with HypF. **Journal of Bacteriology**, Washington, v. 190, n. 4, p. 1447-1458, 2008.

56 MCCULLOCH, K. M. et al. Structural studies of thiamin monophosphate kinase in complex with substrates and products. **Biochemistry**, Washington, v. 47, n. 12, p. 3810-3821, 2008.

57 EHRENREICH, A. et al. Selenoprotein synthesis in Escherichia-coli - purification and characterization of the enzyme catalyzing selenium activation. **European Journal of Biochemistry**, Munich, v. 206, n. 3, p. 767-773, 1992.

58 VERES, Z. et al. Selenophosphate synthetase - enzyme properties and catalytic reaction. **Journal of Biological Chemistry**, Bethesda, v. 269, n. 14, p. 10597-10603, 1994.

59 WALKER, H.; FERRETTI, J. A.; STADTMAN, T. C. Isotope exchange studies on the Escherichia coli selenophosphate synthetase mechanism. **Proceedings of the National Academy of Sciences of the United States of America**, Bethesda, v. 95, n. 5, p. 2180-2185, 1998.

60 ITOH, Y. et al. Structure of selenophosphate synthetase essential for selenium incorporation into proteins and RNAs. **Journal of Molecular Biology**, London, v. 385, n. 5, p. 1456-1469, 2009.

61 KIM, I. Y.; VERES, Z.; STADTMAN, T. C. Escherichia-coli mutant seld enzymes: the cysteine-17 residue is essential for selenophosphate formation from ATP and selenide. **Journal of Biological Chemistry**, Rockville, v. 267, n. 27, p. 19650-19654, 1992.

62 TAMURA, T. et al. Selenophosphate synthetase genes from lung adenocarcinoma cells: SPS1 for recycling L-selenocysteine and Sps2 for selenite assimilation. **Proceedings of the National Academy of Sciences of the United States of America**, Bethesda, v. 101, n. 46, p. 16162-16167, 2004.

63 GASTEIGER, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, Oxford, v. 31, n. 13, p. 3784-3788, 2003.

64 LIN, H.-H.; TSENG, L.-Y. DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines. **Nucleic Acids Research**, Oxford, v. 38, p. W503-W507, 2010. DOI: 10.1093/NAR/GKQ514.

65 SCHULTZ, J. et al. SMART, a simple modular architecture research tool. **Proceedings of the National Academy of Sciences of the United States of America**, Bethesda, v. 95, n. 11, p. 5857-5864, 1998.

66 MARCHLER, A. B. et al. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. **Nucleic Acids Research**, Oxford, v. 30, n. 1, p. 281-283, 2002.

67 LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, Oxford, v. 23, n. 21, p. 2947-2948, 2007.

68 JONES, D. T. Protein secondary structure prediction based on position-specific scoring matrices. **Journal of Molecular Biology**, London, v. 292, n. 2, p. 195-202, 1999.

69 BAZZOLI, A.; TETTAMANZI, A. G. B.; ZHANG, Y. Computational protein design and large-scale assessment by I-TASSER structure assembly simulations. **Journal of Molecular Biology**, London, v. 407, n. 5, p. 764-776, 2011.

70 KACZANOWSKI, S.; ZIELENKIEWICZ, P. Why similar protein sequences encode similar three-dimensional structures? **Theoretical Chemistry Accounts**, New York, v. 125, n. 3-6, p. 643-650, 2010.

71 SAVES. Structural analysis and verification server. Disponível em: < <http://nihserver.mbi.ucla.edu/SAVES/> >. Acesso em: 12 maio 2011.

72 WIEDERSTEIN, X. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids research**, Oxford, v. 35, n. S, p. W407-W410, 2007.

73 SANCHEZ, R.; SALI, A. Comparative protein structure modeling: introduction and practical examples with modeller. **Protein Structure Prediction: Methods and Protocols**, Totowa, v. 143, p. 97-129, 2000. DOI: 10.1385/1-59259-368-2:97.

74 BLAST. Basic local alignment search tool. Disponível em: < <http://blast.ncbi.nlm.nih.gov/Blast.cgi> >. Acesso em: 12 maio 2011.

75 CLUSTALW2. Multiple sequence alignment. Disponível em: < <http://www.ebi.ac.uk/Tools/msa/clustalw2/> >. Acesso em: 12 maio 2011.

76 LYSKOV, S. GRAY, J. J. The RosettaDock server for local protein-protein docking. **Nucleic Acids Research**, Oxford, v. 36, n. 1, p. W233-W238, 2008.

77 KRISSINEL, E.; HENRICK, K. Detection of protein assemblies in crystals. **Computational Life Sciences, Proceedings**, Berlin, v. 3695, p. 163-174, 2005. DOI: 10.1007/11560500\_15.

78 ORTEGA, A.; AMOROS, D.; GARCIA DE LA TORRE, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. **Biophysical Journal**, Cambridge, v. 101, n. 4, p. 892-898, 2011.

79 NCBI. National center for biotechnology information. Disponível em: < <http://www.ncbi.nlm.nih.gov/> >. Acesso em: 12 nov. 2011.

80 BIOEDIT. Disponível em: < <http://www.mbio.ncsu.edu/bioedit/bioedit.html> >. Acesso em: 22 maio 2011.

81 GENE RUNNER. Molecular biology analysis program. Disponível em: < [http://www.uoguelph.ca/~djosephy/lab/handbook\\_generunner.html](http://www.uoguelph.ca/~djosephy/lab/handbook_generunner.html) >. Acesso em: 20 maio 2010.

82 SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 74, n. 12, p. 5463-5467, 1977.

83 CANTOR, C. R.; SCHIMMEL, P. R. **Biophysical chemistry: Part II - techniques for the study of biological structure and function**. New York: Freeman and Co., 1980.

84 K., S. R. **Protein purification: principles and practice**. New York: Springer-Verlag, 1982.

85 DICHROWEB. On-line analysis for protein Circular Dichroism spectra. Disponível em: < <http://dichroweb.cryst.bbk.ac.uk/html/home.shtml> >. Acesso em: 22 maio 2011.

86 BERNE, B. J.; PECORA, R. **Dynamic light scattering**. New York, Wiley interscience, 1976.

87 SVERGUN, D. I. Small-angle X-ray and neutron scattering as a tool for structural systems biology. **Biological Chemistry**, Berlin, v. 391, n. 7, p. 737-743, 2010.

88 HAMMERSLEY, A. P. **FIT2D**: an introduction and overview. Grenoble: 1997. ERSF internal report.

89 FISCHER, H. et al. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. **Journal of Applied Crystallography**, Malden, v. 43, p. 101-109, 2010. DOI:10.1107/S0021889809043076.

90 SVERGUN, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal of Applied Crystallography**, Malden, v. 25, p. 495-503, 1992. DOI:10.1107/S0021889892001663.

91 SVERGUN, D.; BARBERATO, C.; KOCH, M. H. J. CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. **Journal of Applied Crystallography**, Malden, v. 28, p. 768-773, 1995. DOI: 10.1107/S0021889895007047.

92 SVERGUN, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophysical Journal**, Cambridge, v. 76, n. 6, p. 2879-2886, 1999.

93 VOLKOV, V. V.; SVERGUN, D. I. Uniqueness of ab initio shape determination in small-angle scattering. **Journal of Applied Crystallography**, Malden, v. 36, p. 860-864, 2003. DOI:10.1107/S0021889803000268.

94 LEBOWITZ, J.; LEWIS, M. S.; SCHUCK, P. Modern analytical ultracentrifugation in protein science: A tutorial review. **Protein Science**, Plainview, v. 11, n. 9, p. 2067-2079, 2002.

95 SDENTERP. Disponível em: < <http://www.jphilo.mailway.com/download.htm> >. Acesso em: 22 maio 2011.

96 BORGES, J. C.; RAMOS, C. H. I. Analysis of molecular targets of Mycobacterium tuberculosis by Analytical Ultracentrifugation. **Current Medicinal Chemistry**, Sharjah, v. 18, n. 9, p. 1276-1285, 2011.

97 SCHUCK, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. **Biophysical Journal**, Cambridge, v. 78, n. 3, p. 1606-1619, 2000.

98 DAM, J. et al. Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions  $c(s)$ . **Biophysical Journal**, Cambridge, v. 89, n. 1, p. 619-634, 2005.

99 WERNIMONT, A.; EDWARDS, A. In situ proteolysis to generate crystals for structure determination: an update. **Plos One**, San Francisco, v. 4, n. 4, p. 1-5, 2009.

100 OCHI, T. et al. Perspectives on protein crystallisation. **Progress in Biophysics & Molecular Biology**, Oxford, v. 101, n. 1-3, p. 56-63, 2009.

101 LACOWICKZ, J. R. **Principles of fluorescence spectroscopy**. 3th ed. New York: Springer-Verlag, 2006.

102 SAIDEMBERG, D. M. et al. Shikimate Kinase (EC 2.7.1.71) from *Mycobacterium tuberculosis*: kinetics and structural dynamics of a potential molecular target for drug development. **Current Medicinal Chemistry**, Sharjah, v. 18, n. 9, p. 1299-1310, 2011.

103 MSDIGEST. MS-spectra prediction. Disponível em: < <http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest> >. Acesso em: 12 out. 2011.



ANEXO 1 – Quadro comparativo dos parâmetros massa molecular, raio de giro, raio hidrodinâmico, coeficiente de difusão translacional e coeficiente de sedimentação calculados para os estados oligoméricos de SELD de *E. coli* observados em solução, por meio das técnicas Cromatografia de Exclusão Molecular, Eletroforese em Gel Nativo, Espalhamento Dinâmico de Luz, Espalhamento de Raios-X a Baixo Ângulo e Ultracentrifugação Analítica (Velocidade de Sedimentação).

	Dímero	Tetrâmero	Octâmero
<b>Massa molecular (kDa)</b>			
MM <sub>teórica</sub>	73,94	147,87	295,74
MM <sub>Crom. Excl. Mol.</sub>	62 ± 15	123 ± 29	-
MM <sub>Gel Nativo</sub>	57 ± 8	112 ± 13	-
MM <sub>DLS</sub>	59,8	-	-
MM <sub>SAXS</sub>	85,4	-	-
MM <sub>SV</sub>	77 ± 7	148 ± 5	328 ± 40
MM <sub>SV+DLS</sub>	74 ± 1	-	-
<b>Raio (Å)</b>			
R <sub>esfera</sub> **	27,8	35,1	44,2
R <sub>g</sub> Modelo	28,3	-	-
R <sub>H</sub> Crom. Excl. Mol.	30 ± 1	40 ± 1	-
R <sub>H</sub> Gel Nativo	34 ± 1	40 ± 2	-
R <sub>H</sub> DLS	34,1	-	-
R <sub>g</sub> SAXS	33,1 ± 0,3 (Guinier)	-	-
	34,16 ± 0,06 (p(r))		
<b>Coefficiente de difusão translacional</b>			
<b>(x 10<sup>-7</sup> cm<sup>2</sup>/s)</b>			
D <sub>esfera</sub> **	7,69	6,11	4,85
D <sub>Modelo</sub>	5,70	-	-
D <sub>DLS</sub>	5,28 ± 0,14	-	-
<b>Coefficiente de sedimentação</b>			
<b>(x 10<sup>-13</sup> S)</b>			
S <sub>esfera</sub>	6,17	9,80	15,55
S <sub>Modelo</sub>	4,49	-	-
S <sub>SV</sub>	4,53 ± 0,07	7,1 ± 0,4	13,6 ± 0,5

\* R<sub>g</sub> corresponde ao raio de giro da molécula e R<sub>H</sub> corresponde ao raio hidrodinâmico da molécula.

\*\* R<sub>esfera</sub>, D<sub>esfera</sub> e S<sub>esfera</sub> correspondem aos parâmetros hidrodinâmicos calculados para uma esfera de massa molecular teórica correspondente a cada estado oligomérico da proteína, considerando a densidade típica de proteínas como sendo  $\rho = 1,35 \text{ g/cm}^3$ .



ANEXO 2 – Peptídeos identificados por espectrometria de massas e taxa de incorporação de deutério para cada intervalo de tempo para amostra pura da proteína SELD de *E. coli*.

m/z (av)	Sequência do peptídeo	Porcentagem de incorporação de deutério (%)			
		5 min	10 min	15 min	30 min
452,4901	(I)AGASF(A)	50,0	75,0	100,0	100,0
452,4901	(A)GTSFA(N)	50,0	75,0	100,0	100,0
487,5827	(F)GRIAA(T)	0	75,0	100,0	100,0
501,6504	(Y)IKLGA(V)	0	25,0	50,0	50,0
516,5749	(L)SPEIA(R)	0	100,0	100,0	100,0
541,5877	(L)AGGHSI(D)	0	0	20,0	20,0
551,6235	(A)SFANI(E)	0	0	25,0	25,0
562,6663	(A)VMPEA(E)	0	50,0	100,0	100,0
570,2406	(I)STTDF(F)	0	25,0	50,0	50,0
576,6334	(A)SYGHL(M)	0	0	0	0
585,7282	(I)LGWPI(N)	0	0	0	0
591,6451	(I)VDNPF(D)	33,3	100,0	100,0	100,0
602,7122	(A)PEPIF(G)	0	0	0	0
609,7507	(F)GLLGHL(S)	40,0	60,0	80,0	80,0
619,7652	(A)QAGCKL(F)	0	0	0	0
630,7666	(F)GLAVTGI(V)	0	100,0	100,0	100,0
673,8574	(F)AMGGKPI(M)	0	75,0	100,0	100,0
684,7310	(I)LHSEQA(K)	0	80,0	80,0	100,0
717,7577	(I)STTDF(M)	40,0	40,0	40,0	40,0
717,8889	(A)EKKSLI(K)	0	0	0	0
725,8222	(L)SEMCQGA(G)	0	0	0	0
730,9067	(L)LAVMPEA(E)	0	80,0	100,0	100,0
749,3651	(A)QAGCKLF(L)	33,3	83,3	100,0	100,0
772,9249	(A)PEPIFGL(A)	25,0	25,0	25,0	25,0
794,8908	(L)KPDHQGL(A)	0	80,0	80,0	80,0
805,0553	(A)MGGKPI(M)	16,7	100,0	100,0	100,0
861,9326	(A)ENEVKATA(A)	0	42,9	42,9	57,1
878,0192	(L)TTAEKSL(L)	57,1	57,1	57,1	57,1
887,0704	(I)SPKVLETI(L)	100,0	100	100,0	100,0
898,1404	(L)TKPLGIGVL(T)	57,1	71,4	85,7	85,7
966,1075	(-)MSENSIRL(T)	71,4	71,4	85,7	100,0
979,1306	(L)LKPDHQGLA(T)	57,1	71,4	71,4	71,4
1011,214	(W)PINKLSPEI(A)	83,3	100	100,0	100,0
1032,189	(I)PKLPGVEEY(L)	66,7	100	100,0	100,0
1048,060	(L)VGNETRDAA(V)	37,5	37,5	75,0	75,0
1059,234	(I)STTDFM(V)	57,1	57,1	57,1	57,1
1067,153	(A)REVELEGGRY(A)	55,6	77,8	100,0	100,0
1080,299	(F)FMPIVDNPF(D)	16,7	16,7	16,7	16,7
1195,388	(F)MPIVDPFDF(G)	83,3	83,3	83,3	83,3
1209,312	(I)AREVTEGGRYA(C)	0	20,0	30,0	30,0
1298,406	(A)VPGGTERNFASY(G)	0	0	10,0	10,0
1365,580	(L)MGEMPREVRDL(L)	55,6	55,6	55,6	55,6

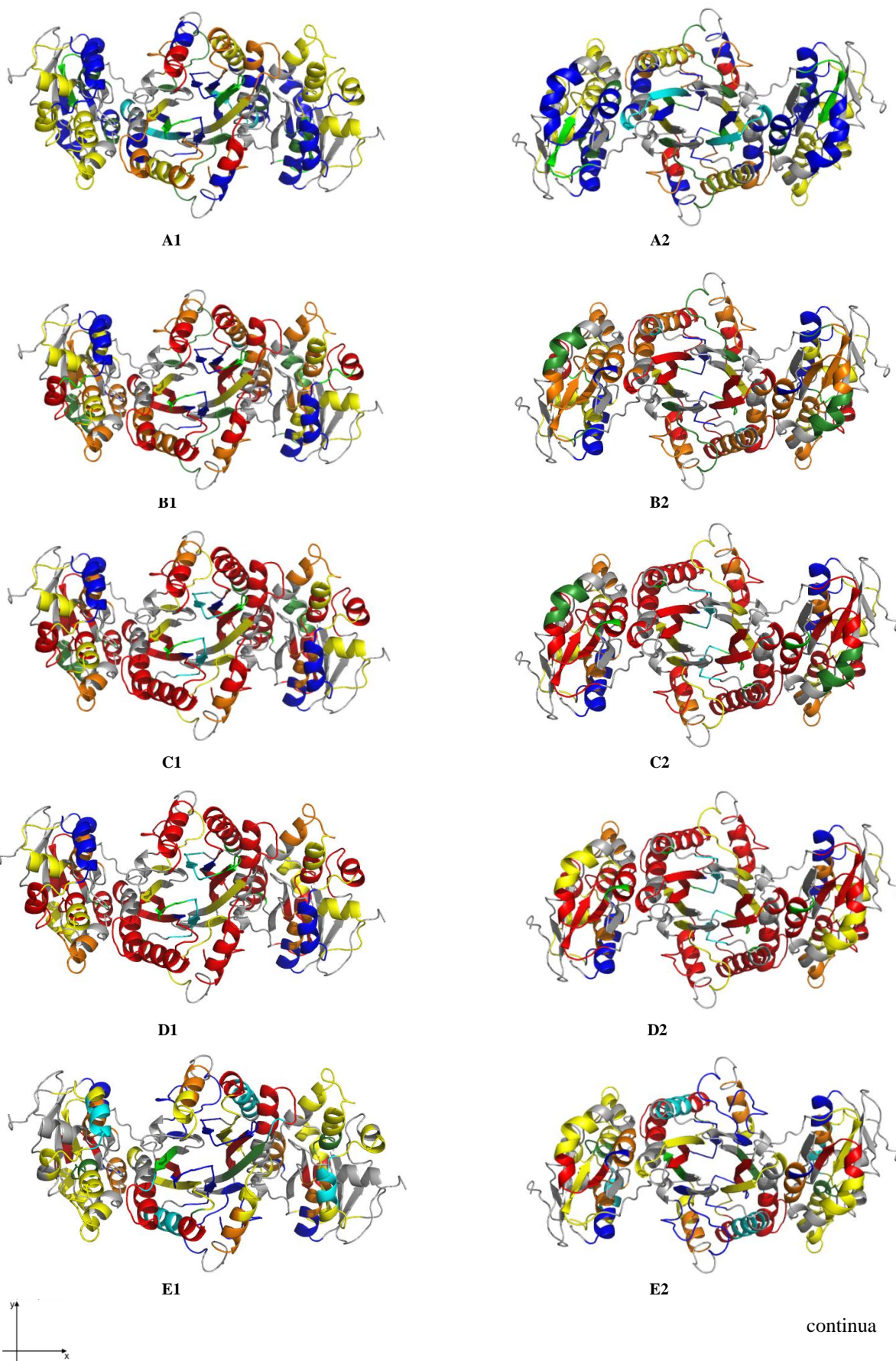


ANEXO 3 – Peptídeos identificados por espectrometria de massas e taxa de incorporação de deutério para cada intervalo de tempo para amostra da proteína SELD de *E. coli* em interação com o complexo SELA-SELC.

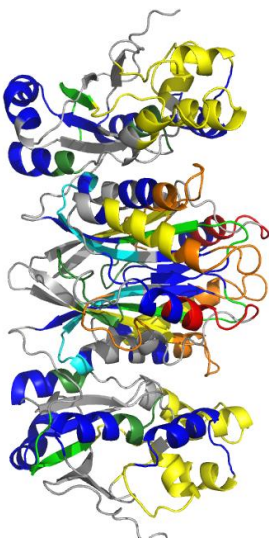
m/z (av)	Sequência do peptídeo	Porcentagem de incorporação de deutério
		(%) 30 min
452,4901	(I)AGASF(A)	75,0
452,4901	(A)GTSFA(N)	75,0
487,5827	(F)GRIAA(T)	100,0
501,6504	(Y)IKLGA(V)	50,0
516,5749	(L)SPEIA(R)	100,0
541,5877	(L)AGGHSI(D)	0
551,6235	(A)SFANI(E)	0
562,6663	(A)VMPEA(E)	100,0
570,2406	(I)STTDF(F)	25,0
576,6334	(A)SYGHL(M)	50,0
585,7282	(I)LGWPI(N)	66,7
591,6451	(I)VDNPF(D)	100,0
602,7122	(A)PEPIF(G)	0
609,7507	(F)GLLGHL(S)	80,0
619,7652	(A)QAGCKL(F)	20,0
630,7666	(F)GLAVTGI(V)	100,0
673,8574	(F)AMGGKPI(M)	100,0
684,7310	(I)LHSEQA(K)	80,0
717,7577	(I)STTDF(F)	20,0
717,8889	(A)EKKSLI(K)	0
725,8222	(L)SEMCQGA(G)	0
730,9067	(L)LAVMPEA(E)	100,0
749,3651	(A)QAGCKLF(L)	50,0
772,9249	(A)PEPIFGL(A)	0
794,8908	(L)KPDHQL(A)	60,0
805,0553	(A)MGGKPI(M)	66,7
861,9326	(A)ENEVKATA(A)	57,1
878,0192	(L)TTAEKKSL(L)	57,1
887,0704	(I)SPKVLETI(L)	66,7
898,1404	(L)TKPLGIGVL(T)	42,9
966,1075	(-)MSENSIRL(T)	0
979,1306	(L)LKPDHQLA(T)	57,1
1011,214	(W)PINKLSPEI(A)	33,3
1048,060	(L)VGNETRDAA(V)	0
1059,234	(I)STTDFMPI(V)	14,3
1067,153	(A)REVELEGGRY(A)	22,2
1080,299	(F)FMPIVDNPF(D)	16,7
1195,388	(F)MPIVDPFDF(G)	16,7
1209,312	(I)AREVTEGGRYA(C)	10,0
1298,406	(A)VPGGTERNFASY(G)	20,0



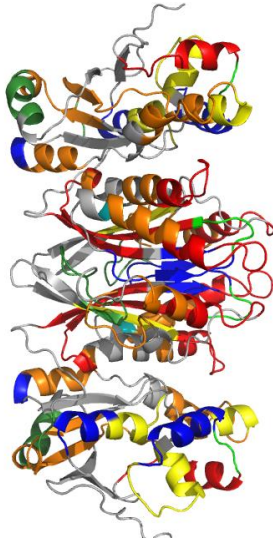
ANEXO 4 - Perfil de incorporação de deutério na SELD de *E. coli* em diferentes tempos de troca H/D:  $t_5 = 5$  min (A),  $t_{10} = 10$  min (B),  $t_{15} = 15$  min (C),  $t_{30} = 30$  min (D),  $int = 30$  min (no complexo SELD-SELA-SELCA) (E). Os índices 1 e 2 representam imagens giradas de  $180^\circ$  em torno do eixo x e os índices 3 e 4 representam  $180^\circ$  em torno do eixo y.



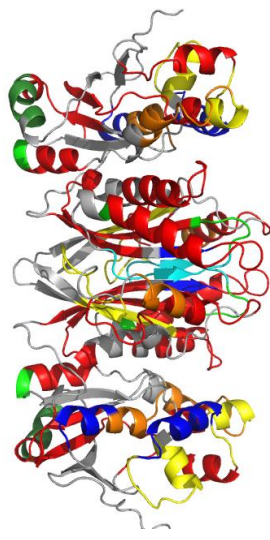
continuação



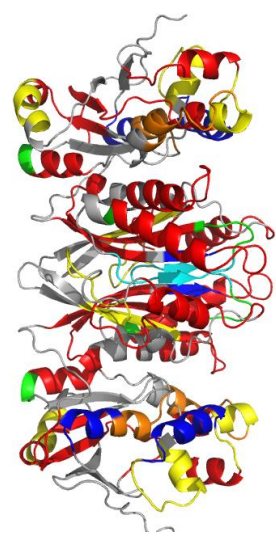
A3



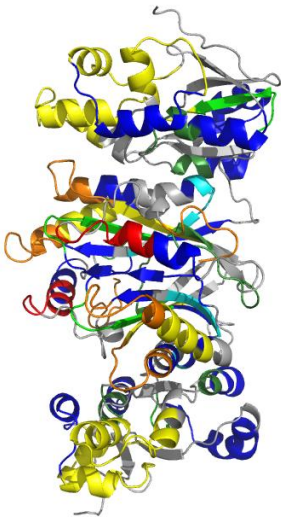
B3



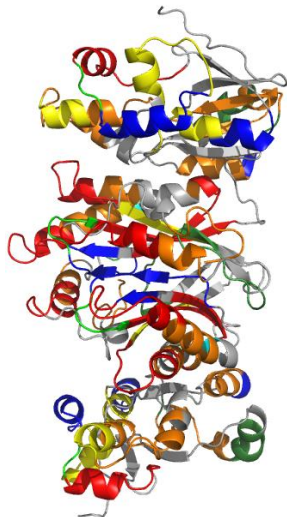
C3



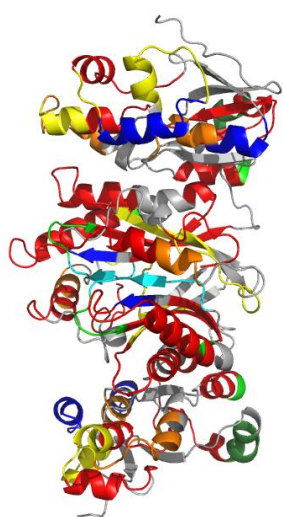
D3



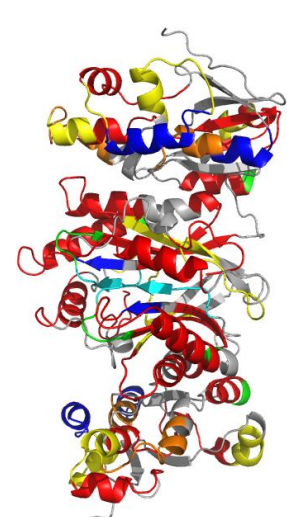
A4



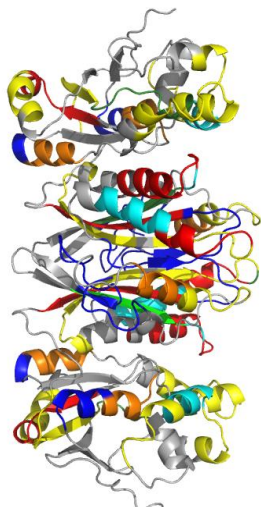
B4



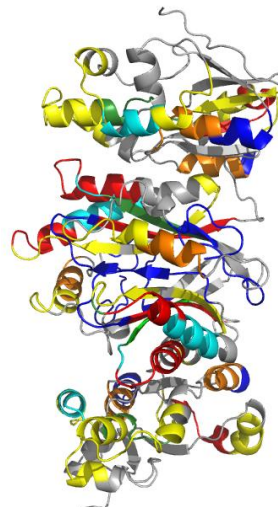
C4



D4



E3



E4

