

Universidade de São Paulo  
Instituto de Física de São Carlos

Alexandre Cassago

Determinação estrutural da proteína Selenocisteína Sintase de *Escherichia coli*

São Carlos

2010



Alexandre Cassago

Determinação estrutural da proteína Selenocisteína Sintase de *Escherichia coli*

Tese 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 Doutor em Ciências.

Área de concentração: Física Aplicada

Opção Física Biomolecular.

Orientador: Prof. Dr. Otavio Henrique Thiemann

São Carlos

2010

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 IFSC/USP

Cassago, Alexandre

Determinação estrutural da proteína Selenocisteína Sintase de  
*Escherichia coli*. / Alexandre Cassago; orientador Otavio Henrique  
Thiemann.—São Carlos, 2010.

110 p.

Tese (Doutorado – Programa de Pós-Graduação em Física –  
Área de concentração: Física Aplicada – opção Física Biomolecular)  
– Instituto de Física de São Carlos da Universidade de São Paulo

1. Selenocisteína Sintase. 2. Selenocisteína. 3. tRNA<sup>sec</sup>. I.  
Título

## FOLHA DE APROVAÇÃO

Alexandre Cassago

Tese apresentada ao Instituto de Física de  
São Carlos da Universidade de São Paulo  
para obtenção do título de Doutor em  
Ciências.  
Área de Concentração: Física Aplicada -  
Opção: Física Biomolecular

Aprovado(a) em: 27.08.2010

Comissão Julgadora

Prof(a). Dr(a). Otavio Henrique Thiemann

Instituição: IFSC/USP

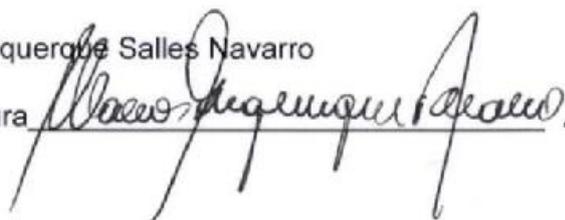
Assinatura



Prof(a). Dr(a). Marcos Vicente de Albuquerque Salles Navarro

Instituição: IFSC/USP

Assinatura



Prof(a). Dr(a). Cristiano Luis Pinto de Oliveira

Instituição: IF/USP

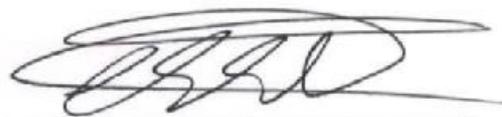
Assinatura



Prof(a). Dr(a). Christian Macagnan Probst

Instituição: FIOCRUZ

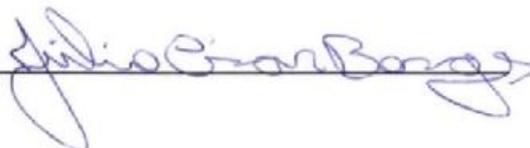
Assinatura



Prof(a). Dr(a). Júlio Cesar Borges

Instituição: IQSC/USP

Assinatura





## **DEDICATÓRIA**

Aos meus pais, Herminio Cassago Junior e Maria Rita de Vasconcellos Cassago pelo amor, dedicação e exemplos de vidas que procuro seguir. E a minha esposa Carolina Aparecida de Guzzi Cassago, por completar minha vida a cada instante.

## **AGRADECIMENTOS**

Aos meus pais Herminio e Maria Rita e minha irmã Ana Paula pela união e amor que se desenvolve ao longo de nossas vidas tornando-nos uma família verdadeiramente feliz.

À minha esposa Carolina por seu amor e por nossa futura filha Cecília.

Ao professor Dr. Otavio Henrique Thiemann pela orientação desde meu mestrado e principalmente pela amizade desenvolvida durante todos esses anos, que renderão ainda muitos frutos.

Aos professores Dr. Eduardo Horjales Reboredo, Dr. Humberto d'Muniz Pereira e Dr. Richard Chales Garret pela ajuda, instrução e amizade.

Aos professores Dra. Iris L. Torriani e Dr. Cristiano L. P. Oliveira pelo trabalho em conjunto.

Ao Dr. Alexander Golubev, Dra. Leandra Watanabe, Dra. Patrícia R. de Moura e Dra. Sandra Krauchenco pelas aulas e discussões sobre cristalização de proteínas.

A José Fernando Ruggiero Bacheга e Lucas Bleicher pela imensurável ajuda e futuras rodas de queijo e vinho.

À Angela Maria Fala, Juliana R. T. de Souza e Larissa Romanello pela ajuda e amizade.

Aos companheiros de grupo de pesquisa Alécio A. P. Junior, Daiana E. Martil, Fernanda C. Costa, Gustavo Mercaldi, Ivan R. e Silva, Jaqueline P. Evangelista, Lívia M. Fain, Lívia R. Manzine, Marcos M. de Souza, Marco T. A. da Silva, Maria Amélia O. Dotta, Victor E. A. Caldas, Simone M. Duarte, Susana A. Scuaccio e aos colegas do Instituto de Física de São Carlos Alessandro S. Nascimento, Amanda Bernardes, Ana Carolina M. Figueira, Ana C. Puhl, Cesar M. Camilo, Daniela B. B. Trivella, Eliene Bogo, Fernanda Batista, Flávio R. Rosseto, Francieli Colussi, Frederico M. Ferreira, Kelven Ulisses, Leonardo H. F. Lima, Lis S. Miotto, Marcelo

Liberato, Marcos Navarro, Maria A. M. Santos, Maria Luiza Voltatodio, Mario O. Neto, Mario Sanches, Marisa A. de Lima, Matheus Postigo, Maycou A. Deriggi, Napoleão F. Valadares, Nathalia Rodrigues, Nayara Cavalcante, Ney R. Leite, Paulo Sergio, Rafael Guido, Renata A. A. Mesquita, Rodrigo V. Portugal, Thiago Moda, Vitor H. B. Serrão, Viviane I. Serpa, Wanius J. G. de Silva, alguns não mais convivendo diretamente conosco, mas sempre por nós lembrados.

À constante ajuda de Andressa P. A. Pinto e Derminda Isabel Morais nos experimentos realizados no Laboratório de Biofísica do Instituto de Física de São Carlos.

À biblioteca SBI-IFSC e seus funcionários altamente qualificados e sempre prestativos

Ao Instituto de Física de São Carlos (IFSC) pela infra-estrutura e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro para o desenvolvimento desse estudo de doutorado e minha formação acadêmica.

## **EPIGRAFE**

"Dans les champs de l'observation le hasard ne favorise que les esprits préparés."

Louis Pasteur

## RESUMO

CASSAGO, A. **Determinação estrutural da proteína Selenocisteína Sintase de *Escherichia coli***. 2010. 110 p. Tese (Doutorado) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

A biossíntese do 21<sup>o</sup> aminoácido, Selenocisteína (Sec – U), envolve uma complexa maquinaria enzimática composta, em eubactérias, pela Selenocisteína Sintase (SELA), Fator de Elongação de Selenocisteína (SELB), Selenofosfato Sintetase (SELD) e tRNA de Inserção Selenocisteína (tRNA<sup>sec</sup>). Em arqueobactérias e eucariotos existem ainda O – fosforil tRNA<sup>sec</sup> Kinase (PSTK), SepSecS como SELA, EFSec como SELB, SPS1 e 2 como SELD e Proteína Ligante ao SECIS 2 (SBP2). O resíduo Selenocisteína é incorporado à proteína nascente no códon semelhante ao UGA de terminação identificado como local para incorporação de Sec, pela presença da Sequência de Inserção de Selenocisteína (SECIS), juntamente ao códon UGA na região codificante em bactérias e na região 3' não codificante em arqueobactérias e eucariotos. SELA desempenha um papel central nessa via de biossíntese pela modificação do resíduo de Serina carregado ao tRNA<sup>sec</sup> pela enzima Seril-tRNA Sintetase (SerRS) convertendo-o em Selenocisteína. Essa enzima forma um complexo homodecamérico que reconhece e liga-se especificamente a Seril-tRNA<sup>sec</sup>. A interação específica entre SELA e o tRNA permanece ainda não determinada. Nosso objetivo é a investigação estrutural por Espalhamento de Raios-X a Baixos Ângulos (SAXS) e cristalização da SELA e SELA-tRNA<sup>sec</sup> de *Escherichia coli*. Dados de SAXS determinaram parâmetros dimensionais como dimensão máxima, massa molecular e raio de giro. O modelo *ab-initio* foi calculado assumindo a simetria P5<sub>2</sub> de projeções de Microscopia Eletrônica de Transmissão (TEM). Os cristais obtidos do complexo SELA-tRNA mostraram o grupo espacial e dimensões da cela, apesar da baixa resolução dos dados. Para melhorar os estudos estruturais um modelo para proteína SELA de *Escherichia coli* foi construído usando o alinhamento da sequência de aminoácidos e o PDB, da proteína SELA putativa, de *Methanococcus jannaschii*, que apesar da baixa identidade resultou em um modelo muito bom. Adicionalmente, uma Análise de Acoplamento Estatístico (SCA) foi

realizada baseada em alinhamentos múltiplos da proteína SELA, ordenando os aminoácidos mais conservados e a relação existente entre eles.

Palavras – chave: Selenocisteína Sintase. Selenocisteína. tRNA<sup>sec</sup> de Inserção para Selenocisteína.

## ABSTRACT

CASSAGO, A. **Structural determination of Selenocysteine Synthase from *Escherichia coli***. 2010. 110 p. Tese (Doutorado) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

The biosynthesis of the 21<sup>th</sup> amino acid, Selenocysteine (Sec – U), requires complex enzymatic machinery composed in eubacteria of: Selenocysteine Synthase (SELA), Selenocysteine Specific Elongation Factor (SELB), Selenophosphate Synthetase (SELD) and a specific Selenocysteine Inserting tRNA (tRNA<sup>sec</sup>). In archaeobacteria and eukaryotes there are O – phosphoryl tRNA<sup>sec</sup> Kinase (PSTK), SepSecS as SELA, EFSec as SELB, SPS1 and 2 as SELD and SECIS Binding Protein 2 (SBP2). The Selenocysteine residue is incorporated into a nascent protein at a UGA like stop codon signaling as a Sec incorporation site by the presence of a Selenocysteine Insertion Sequence (SECIS), embedding the UGA codon in the coding region in bacteria and in a 3' UTR in archaea and eukarya. SELA plays a central role in this pathway by modifying the Serine residue charged into the tRNA<sup>sec</sup> by Seryl-tRNA Synthetase (SerRS) and converting it into Selenocysteine. This enzyme forms a homodecameric complex that specifically recognizes and binds to Seryl-tRNA<sup>sec</sup>. The specific interaction of SELA and its tRNA remains unclear. Our aim is the structural investigation by Small Angle X ray Scattering (SAXS) and crystallization of *Escherichia coli* SELA and SELA-tRNA<sup>sec</sup>. SAXS data determined dimensional parameters as maximum dimension, molecular mass and radius of gyration. *Ab-initio* model calculation was made assuming a P5<sub>2</sub> symmetry from Transmission Electron Microscope (TEM) projections. Crystals of SELA-tRNA complex shown the space-group and cell dimensions, although its low resolution. To improve the structural studies a SELA model of *E. coli* was built using the amino acid sequences alignment and the PDB from *Methanococcus jannaschii*, SELA putative protein, which although the lower identities result in a very good model. In addition, a Statistical Coupling Analysis (SCA) was performed based on a multiple sequence

alignment of SELA, ordering the most preserved amino acid and the relation between them.

Keywords: Selenocysteine Synthase. Selenocysteine. Selenocysteine inserting tRNA<sup>sec</sup>.

## LISTA DE FIGURAS

- Figura 1. 1** - Aminoácido Selenocisteína. visualização 2D a esquerda e 3D a direita. (Fonte Wikipedia <http://en.wikipedia.org/wiki/Selenocysteine>)..... 30
- Figura 1. 2** - Estrutura secundária e terciária de tRNAs. Em A- estrutura geralmente encontrada em tRNAs; Em B- tRNA de Inserção de Selenocisteína (tRNA<sup>sec</sup>) de eubactérias; Em C- tRNA<sup>sec</sup> de arqueobactérias e eucariotos; Em D e E- tRNA<sup>sec</sup> de *Escherichia coli* representado respectivamente pela estrutura secundária e terciária<sup>17, 36, 37</sup>..... 32
- Figura 1. 3** - Sequências de Inserção Sec (SECIS). Em A- SECIS para eubactérias (*E. coli*) localizado logo após o códon UGA para Sec. Delimitado pelo retângulo encontra-se a região reconhecida pela proteína SELB a exatamente 11 nucleotídeos do códon UGA; Em B- SECIS para arqueobactérias (*M. maripaludis*) localizado na região 3'-UTR a aproximadamente 25 nucleotídeos do stop códon; Em C e D- SECIS para eucariotos (mamíferos) em que respectivamente é ilustrada a forma I e II, contendo os espaçamentos indicados pela legenda a esquerda, o quarteto AUGA e a presença de AA ou CC na região do loop apical<sup>35, 38, 39</sup>..... 32
- Figura 1. 4** - Etapas envolvidas na biossíntese e incorporação de Selenocisteína de *Escherichia coli* conforme descrito no texto. A etapa 4 pode ser visualizada em detalhes na figura 1.5 na página seguinte. .... 36
- Figura 1. 5** - Esquema detalhado da conversão do aminoácido serina em selenocisteína pela enzima Selenocisteína Sintase (SEL A) parcialmente representada em azul. Também indicado ligado à enzima SELA o cofator enzimático Piridoxal 5'- Fosfato: 1- Ligação do seril-tRNA<sup>sec<sub>uca</sub></sup> ao cofator piridoxal 5'-fosfato. 2- liberação de uma molécula de água com a formação de um composto intermediário (aminoacrilil-tRNA<sup>sec<sub>uca</sub></sup>). Nessa etapa o composto intermediário pode tautomerisar liberando Piruvato (espontaneamente), ou devido a um agente redutor (borohidratado de potássio) liberar alanil-tRNA<sup>ec<sub>uca</sub></sup>. 3- transferência do selênio reduzido e ativado à molécula intermediária, resultando no selenocisteil-tRNA<sup>sec<sub>uca</sub></sup>. 4- liberação do aminoácido selenocisteína, disponibilizando a enzima SEL A e o cofator Piridoxal 5'- Fosfato para reinício do ciclo<sup>40</sup> ..... 37

- Figura 1. 6** - Imagens da proteína Selenocisteína Sintase (SELA) de *E.coli* obtidas pela técnica de Microscopia Eletrônica de Transmissão (TEM) por “*negative stain*”, utilizando como marcadores acetato de urânio (A, C, D) e oxalato de urânio (B). Em A- vista frontal da proteína SELA, ilustrando a simetria encontrada; em B- vista lateral, ilustrando uma subunidade dimérica. Em C e D- vista frontal e lateral respectivamente da proteína SELA ligada ao composto intermediário aminoacrilil-tRNA<sup>sec</sup>. (Reproduzido de Harald Engelhardt<sup>46</sup>)..... 39
- Figura 3. 1** - Esquema para amplificação e transcrição *in vitro* do gene *selC* de *E. coli*. Em azul a região promotora T7 RNA Polimerase, seguida em alaranjado pela sequência do gene *selC*. Abaixo o transcrito de 95rnt com sua terminação CCA comum a todos os tRNAs..... 46
- Figura 3. 2** - Amplificação e transcrição *in vitro* do gene de inserção selenocisteína tRNA<sup>sec</sup> (*selC*) e tRNA<sup>sec</sup> visualizado em gel de agarose 2% - TAE [1X] corado com brometo de etídeo. Em A- coluna 1- padrão molecular “1Kb Plus DNA Ladder”; coluna 2- amplificação do gene *selC* portando a região promotora T7; coluna 3- transcrição do tRNA<sup>sec</sup>. Em B- coluna 1- padrão molecular “1Kb Plus DNA Ladder”; coluna 2- transcrição do tRNA<sup>sec</sup> pelo protocolo adaptado da literatura; coluna 3- transcrição do tRNA<sup>sec</sup> utilizando o “Kit Megashortscript – Ambion”..... 47
- Figura 3. 3** - Gráfico de energia, resultante das análises de restrições, para os modelos da proteína SELA gerados pelo programa MODELLER. O modelo de número 96 apresenta a menor energia (2228,3464)..... 49
- Figura 3. 4** - Distribuição dos resíduos de aminoácidos no diagrama de Ramachandran. A- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* - PDB 2AEU; B- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* - PDB 2AEV; C- estrutura modelada da proteína SELA de *E. coli*..... 49
- Figura 3. 5** - Alinhamento entre as sequências da proteína Selenocisteína Sintase de *E. coli* (SELAeco) e *M. jannaschii* (SELA<sub>mj</sub>). As setas indicam os resíduos localizados nas regiões não permitidas ( ↓ ), generosamente permitidas( ↓ ↓ ), Lisina do sítio ativo ( ↓ ) e região conservada estruturalmente próxima ao sítio ativo ( ↓ ) descritas no texto. .... 51
- Figura 3. 6** - Modelo tridimensional da proteína SELA gerado pelo programa MODELLER. A- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* (PDB 2AEU); B- modelo gerado da proteína SELA de *E. coli*; C- sobreposição das duas estruturas..... 52

- Figura 3. 7** - A- PDB da proteína SELA modelado pelo programa MODELLER submetido as análises de  $\Delta G^{\text{stat}}$  para conservação dos resíduos para proteína SELA de *E. coli*. Os valores de conservação aumentam do azul para o vermelho. B- Em destaque as setas indicam os resíduos Lisina-295 do sítio ativo ( ↓ ), o resíduo Asparagina-85 estruturalmente próxima ao sítio ativo ( ↓ ) e dois resíduos também conservados nessa região, Histidina-92 e Glutamina-301 ( ↓ ). ..... 55
- Figura 3. 8** - Análise de correlação entre os resíduos da proteína SELA de *E. coli*. A- modelo gerado pelo programa MODELLER com gradiente de conservação de resíduos; B, C, D, E, F, G e H- sobreposição à estrutura SELA modelada em cinza dos resíduos que possuem correlação resultante respectivamente aos grupos 1, 2, 3, 4, 5, 6, 7 agrupados na tabela 3.9. .... 61
- Figura 3. 9** - Curvas de espalhamento obtidas experimentalmente (pontos) e respectivo ajuste pelo GNOM extrapolando a  $q=0$  (linha cheia). Em preto, proteína SELA e em azul, proteína SELA-tRNA. .... 64
- Figura 3. 10** - Função de distribuição de distâncias a partir da curvas de espalhamento obtidas experimentalmente. Em preto proteína SELA e em azul proteína SELA-tRNA. .... 65
- Figura 3. 11** - Modelo *ab initio*, gerados pelo programa GASBOR. A- proteína Selenocisteína Sintase (SELA); B- SELA – tRNA; C- sobreposição das duas estruturas. .... 66
- Figura 3. 12** - Modelo *ab initio*, gerados pelo programa GASBOR e imagens tratadas pelo programa PYMOL. A- proteína Selenocisteína Sintase (SELA); B- SELA – tRNA. As linhas negras e respectivos números indicam as medidas da subunidade dimérica: 1- comprimento lateral; 2- comprimento em direção ao centro; 3- espessura e o círculo- diâmetro interno do orifício central. Valores encontrados na tabela 3.11. A seta negra indica um dos cinco orifícios encontrado apenas na estrutura SELA-tRNA. .... 67
- Figura 3. 13** - Foto dos cristais obtidos ao longo das triagens e refinamentos das condições de cristalização para proteína Selenocisteína Sintase de *E. coli*. Em A- Cristal Screen 2 – 21: 2.0 M Sodium Chloride, 0.1 M MES pH 6.5, 0.2 M Na/K Phosphate; Em B- Sodium Chloride – B3: 0.1 M

MES pH 6.0, 2.0 M Sodium Chloride; Em C- refinamento dos cristais A e B - 26% MPD, 0.1M Mes pH 6.2; Em D- refinamento dos cristais A e B - 29% MPD, 0.1M Mes pH 6.4; Em E a L condições descritas na tabela 3.12 (página 67); Em M a P condições descritas na tabela 3.13 (página 71); Em Q- cristal condição N formado em solução crio de 22% de glicerol; Em R- cristal condição N formado após “microseeding”; Em S- cristal condição N formado em solução crio 20% de etilenoglicol; Em T- refinamento cristal N crescido após digestão com Tripsina em 1.6 M do agente precipitante. Cristais S e T apresentaram melhores padrões de difração..... 70

**Figura 3. 14** - Cromatografia da proteína SELA após gel filtração (Superdex 200) e análise em SDS-PAGE: coluna 1- padrão de massa molecular; coluna 2- proteína SELA antes da coluna de gel filtração; coluna 3- proteína SELA após coluna de gel filtração. .... 72

**Figura 3. 15** - Padrão de difração para a proteína SELA e SELA-tRNA<sup>sec</sup>. Em A- proteína SELA, em B- proteína SELA-tRNA<sup>sec</sup>. Ambas cristalizadas conforme descrito na tabela 3.14 – M, N e visualizados na figura 3.13 – M, N. .... 75

**Figura 3. 16** - Mapas dos vetores pET (Novagen) destacando-se em vermelho os locais de clivagem para as enzimas *Dra* III e *Sap* I. Em A- vetor pETDuet -1, resistente a ampicilina. Em B- vetor pET-29a, resistente a kanamicina. Figura cedida pela doutoranda Lívia Regina Manzine. .... 80

**Figura 3. 17** - Análise em SDS-PAGE dos ensaios de protease para proteína SELA-2. Em A- incubação com Tripsina – 60 min/25° C; Em B- incubação com Papaína – 20 min/25° C. Coluna 1- padrão de massa molecular; coluna 2- proteína SELA-2 controle; colunas 3 a 7- proteína SELA-2 incubada em concentrações crescentes de protease 20, 40, 60, 80,100 ug/mL respectivamente. A seta negra indica respectivamente a protease Tripsina 23 kDa e Papaína 23,8 kDa. .... 83

**Figura 3. 18** - Padrão de difração para a proteína SELA-tRNA<sup>sec</sup>. Em A- melhor padrão de difração para proteína SELA1-tRNA<sup>sec</sup> congelada em solução crio-protetora 20% etilenoglicol, em B- proteína SELA2-tRNA<sup>sec</sup> digerida com protease Tripsina pela metodologia A, descrita no texto. Ambas cristalizadas conforme descrito na tabela 3.11 N (com variação do agente precipitante para 1.6M para SELA2-tRNA) e visualizados na figura 3.13 – S e T. .... 84

- Figura 3. 19** - Análises das regiões flexíveis do modelo para proteína SELA de *E. coli* (colorido) sobreposto ao modelo cristalográfico de SELA de *M. jannaschii* (prateado). Em A- sobreposição dos dois modelos; B- evidenciando a região flexível para o N-Terminal (nTerm); C- evidenciando a região flexível para a segunda alça (alc2); D- evidenciando a região flexível para a terceira alça (alc3). ..... 87
- Figura 3. 20** - Figura ilustrativa dos plasmídeos pET29a contendo o gene SelA de *E. coli* e o mesmo plasmídeo mostrando as três regiões deletadas. Em A- plasmídeo pET29a-SelA inteiro; B- plasmídeo pET29a-SelA com a região N-Terminal, alc2 e alc3 deletadas respectivamente em marrom, vermelho e verde. .... 89
- Figura 3. 21** - Amplificação do gene SelA-mut\_nTerm e busca por plasmídeos recombinantes (PTZ-SelA-mut\_nTerm) pela técnica de PCR de colônia visualizado em gel de agarose 1% - TAE [1X] corado com brometo de etídeo. Em A- coluna 1- padrão molecular “1Kb plus DNA ladder”; coluna 2- amplificação do gene SelA-mut\_nTerm (1209 pb); coluna 3- amplificação controle do gene SelA (1392 pb). Em B- coluna 1- padrão molecular “1Kb plus DNA ladder”; coluna 2, 4, 7, 8, 9- colônias 12, 14, 17, 18, 19 PTZ-SelA-mut\_nTerm positivas respectivamente (amplificação do gene de interesse e parte da região de múltipla-clonagem do vetor PTZ totalizando aproximadamente 1500 pb); coluna 10- controle negativo (amplificação da região de múltipla-clonagem do vetor PTZ de aproximadamente 300 pb); coluna 3, 5, 6- nenhuma amplificação ocorreu. .... 90
- Figura 3. 22** - Busca por plasmídeos recombinantes (pET29a-SelA-mut\_nTerm) pela técnica de PCR de colônia visualizado em gel de agarose 1% - TAE [1X] corado com brometo de etídeo e testes de expressão das colônias positivas em SDS-PAGE. Em A- coluna 1- padrão molecular “1Kb plus DNA ladder”; coluna 2 a 11- colônias 1 a 10 pET29a-SelA-mut\_nTerm positiva (amplificação do gene de interesse e parte da região de múltipla-clonagem do vetor pET29a, totalizando 1407 pb); coluna 12- controle positivo (amplificação do gene SelA inteiro e parte da região de múltipla-clonagem, totalizando 1585 pb). .... 91

## LISTA DE TABELAS

- Tabela 1. 1** - Selenoproteínas encontradas em procariotos e mamíferos organizadas em ordem alfabética. Dados obtidos a partir de revisões e trabalhos de busca e identificações de Selenoproteínas<sup>4, 48, 49</sup>. ..... 42
- Tabela 3. 1** - Distribuição em porcentagem dos resíduos de aminoácidos no diagrama de Ramachandran para os dois PDBs cristalográficos de *M. jannaschii* e o PDB modelado para *E. coli*. ..... 50
- Tabela 3. 2** - Grupo 1 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 57
- Tabela 3. 3** - Grupo 2 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 57
- Tabela 3. 4** - Grupo 3 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 58
- Tabela 3. 5** - Grupo 4 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 58
- Tabela 3. 6** - Grupo 5 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 59
- Tabela 3. 7** - Grupo 6 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências. .... 59

<b>Tabela 3. 8</b> - Grupo 7 de correlação entre os resíduos da proteína SELA de <i>E. coli</i> . Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.....	59
<b>Tabela 3. 9</b> - Reunião dos sete grupos de correlação entre os resíduos da proteína SELA de <i>E. coli</i> . Na primeira coluna o número de resíduos para cada grupo, seguido pelos respectivos resíduos, posições e característica de sua cadeia lateral: em azul – resíduos polares não carregados; em marrom – resíduos apolares; em verde – resíduos ácidos e em vermelho – resíduos básicos. ....	60
<b>Tabela 3. 10</b> - Análise de correlação/anti-correlação para acoplamento entre procariotos e arqueobactérias. Os números correspondem à média dos logaritmos dos p-valor observado para correlações entre membros de uma comunidade em relação aos membros da outra. Sinais positivos são atribuídos quando há correlação, enquanto sinais negativos são usados para anti-correlação, isto é, a presença de um resíduo diminui a frequência de outro. ....	62
<b>Tabela 3. 11</b> - Parâmetros estruturais para a proteína SELA e SELA-tRNA, obtidos por Microscopia Eletrônica de Transmissão (TEM) e Espalhamento de Raios-X a Baixos Ângulos (SAXS). ....	67
<b>Tabela 3. 12</b> - Fatoriais comerciais de cristalização utilizados para triagens de cristalização da proteína SELA e SELA-tRNA <sup>sec</sup> . ....	69
<b>Tabela 3. 13</b> - Micro-cristais e cristais da proteína SELA obtidos em triagens de cristalização após a adição da coluna de gel filtração Superdex 200 no protocolo de purificação.....	73
<b>Tabela 3. 14</b> - Cristais da proteína SELA e SELA-tRNA <sup>sec</sup> obtidos em triagens de cristalização após a adição da coluna de gel filtração Superdex 200 no protocolo de purificação.....	74
<b>Tabela 3. 15</b> - Informações sobre as condições de cristalização e análises dos padrões de difrações obtidos para a proteína Selenocisteína Sintase ligada ao seu tRNA.....	85

**Tabela 3. 16** - Oligonucleotídeos sintetizado para as deleções das regiões flexíveis encontrados na sequência da proteína SELA. Em verde é mostrado uma pequena sequência utilizada para ancoragem pelas enzimas de restrição e que será eliminada após digestão, em vermelho os sítios para enzimas *Nde* I e *Hind* III (SelA-mut\_nTerm), *Bgl* II (SelA-mut\_alc2) e *Hind* III (SelA-mut\_alc3) e o grifo negro abaixo da trinca de nucleotídeos indica a fase de leitura para as respectivas deleções..... 88

**Tabela 3. 17** - Grupos candidatos a mutação sítio-dirigida analisados pelo servidor “Surface Entropy Reduction Prediction Server” (SERp)<sup>88</sup>. Resíduos propostos a mutação são destacados em verde; resíduos possíveis de mutação e com alta entropia são destacados em azul; resíduos com baixa entropia são destacados em amarelo; demais resíduos não são destacados (<http://nihserver.mbi.ucla.edu/SER/intro.php>)..... 93

**Tabela 3. 18** - Oligonucleotídeos sintetizado para as mutações sítio-dirigidas na sequência gênica da proteína SELA. Em vermelho localizam-se os nucleotídeos mutados, o grifo negro indica o códon responsável pela tradução em Alanina e os demais nucleotídeos seguem a sequência nativa do gene para hibridação e extensão da reação de amplificação. .... 94

## LISTA DE ABREVIATURAS E SIGLAS

3' UTR	Região 3' não traduzida de mRNA ( <i>3' Untranslated mRNA Regions</i> )
DNA	Ácido Desoxirribonucléico ( <i>Deoxyribonucleic Acid</i> )
EMBL	<i>European Molecular Biology Laboratory</i>
IUBMB	<i>International Union of Biochemistry and Molecular Biology</i>
LNLS	Laboratório Nacional de Luz Síncrotron
MAFFT	<i>Mult Alignment using Fast Fourier Transform</i>
MPD	2-Methyl-2,4-pentaneidiol
mRNA	Ácido Ribonucléico mensageiro ( <i>Messenger Ribonucleic Acid</i> )
NCBI	<i>National Center for Biotechnology Information</i>
PCR	Reação em Cadeia de Polimerase ( <i>Polimerase Chain Reaction</i> )
PDB	<i>Protein Data Bank</i>
PFAM	<i>Protein Multiple Sequence Alignments</i>
PLP	Piridoxal 5' – Fosfato ( <i>Pyridoxal 5' – Phosphate</i> )
PSTK	O – fosforil tRNA <sup>sec</sup> Kinase ( <i>O – phosphoryl tRNA<sup>sec</sup> Kinase</i> )
SAXS	Espalhamento de Raios-X a Baixos Ângulos ( <i>Small Angle X-Ray Scattering</i> )
SBP2	Proteína Ligante ao SECIS 2 ( <i>SECIS Binding Protein 2</i> )
SCA	Análises de Acoplamento Estatístico ( <i>Statistical coupling analysis</i> )
Se	Selênio
Sec	Selenocisteína

SECIS	Sequência de Inserção de Selenocisteína ( <i>Selenocysteine Insertion Sequence</i> )
SELA	Selenocisteína Sintase ( <i>Selenocysteine Synthase</i> )
SELB	Fator de Elongação de Selenocisteína ( <i>Selenocysteine Specific Elongation Factor</i> )
EFSec	
selC	Gene de Inserção Selenocisteína – tRNA <sup>sec</sup> ( <i>Selenocysteine Inserting tRNA gene</i> )
SELD	Selenofosfato Sintetase ( <i>Selenophosphate Synthetase</i> )
SepSecS	Selenocisteína Sintase ( <i>Selenocysteine Synthase</i> ) para
mSecS	arqueobactérias e eucariotos
SERp	<i>Surface Entropy Reduction Prediction Server</i>
SerRS	Seril-tRNA Sintetase ( <i>Seryl-tRNA Synthetase</i> )
SLA / LP	<i>Soluble Liver Antigen / Liver Pancreas</i>
SPS1	Selenofosfato Sintetase 1 e 2 ( <i>Selenophosphate Synthetase 1 and 2</i> )
SPS2	para arqueobactérias e eucariotos
tRNA <sup>sec</sup>	tRNA de Inserção Selenocisteína ( <i>Selenocysteine Inserting tRNA</i> )

## LISTA DE SÍMBOLOS

Å	Ângstrom
kDa	Quilodalton ( $10^3$ G/Mol)
L	Litro
mM	Milimolar
MM	Massa Molecular
nm	Nanômetro
° C	Graus Celsius
pH	Potencial Hidrogeniônico
pK <sub>a</sub>	Constante de Dissociação Ácida
ΔG	Varição de Energia (em unidades arbitrárias)



## SUMÁRIO

<b>CAPÍTULO 1</b> .....	27
1 SELÊNIO, SELENOCISTEÍNA E SUAS IMPORTÂNCIAS PARA OS SERES VIVOS.....	27
1.1 Selênio .....	27
1.2 Selenocisteína .....	29
1.3 Selenocisteína Sintase (SELA) de <i>Escherichia coli</i> .....	38
1.4 Selenoproteínas .....	40
<b>CAPÍTULO 2</b> .....	43
2 OBJETIVOS .....	43
<b>CAPÍTULO 3</b> .....	45
3 PROTEÍNA SELENOCISTEÍNA SINTASE E tRNA <sup>SEC</sup> .....	45
3.1 Preparação da proteína Selenocisteína Sintase e tRNA <sup>SEC</sup> .....	45
3.2 Interação proteína SELA e tRNA <sup>SEC</sup> .....	47
3.3 Modelagem da proteína SELA de <i>Escherichia coli</i> .....	48
3.4 Estudos de conservação e acoplamento estatístico da proteína SELA.....	53
3.5 Ensaios estruturais da proteína SELA e SELA-tRNA <sup>SEC</sup> por SAXS.....	63
3.6 Ensaios estruturais da proteína SELA e SELA-tRNA <sup>SEC</sup> por difração de raio-X .....	68
3.6.1 Proteína Selenocisteína Sintase .....	68
3.6.2 Proteína Selenocisteína Sintase-2 .....	79
3.6.3 Ensaios de proteases, deleções e mutações sítio-dirigida para proteína Selenocisteína Sintase-2 .....	81
<b>CAPÍTULO 4</b> .....	95
4 DISCUSSÃO .....	95
4.1 PERSPECTIVAS FUTURAS .....	100
<b>REFERÊNCIAS</b> .....	103



## **CAPÍTULO 1**

### **1 SELÊNIO, SELENOCISTEÍNA E SUAS IMPORTÂNCIAS PARA OS SERES VIVOS**

#### **1.1 Selênio**

O elemento Selênio (Se) foi descoberto em 1817 por Berzelius e Gahn enquanto examinavam sedimentos de plantas em Gripsholm na Suécia<sup>1-5</sup>. Teve sua atividade atribuída a efeitos tóxicos, decorrentes observações e estudos ao longo dos anos, como o primeiro manuscrito que se tem notícia em 1856 por Madison, cirurgião da guarda do Forte Randall em Nebraska nos Estados Unidos, que observou enfermidades fatais em cavalos que pastavam em determinadas áreas próximas ao forte<sup>1</sup>. Em 1929, na Dakota do Sul, estudos em grãos considerados tóxicos e posteriormente em 1933, busca pelos princípios tóxicos desses grãos, que sugeriam ser o elemento Selênio, confirmado no ano seguinte<sup>1</sup>.

Com uma massa atômica de 78,96, o elemento Selênio, compartilha muitas propriedades químicas com o elemento enxofre (32,07), encontrado na forma de inorgânica de  $H_2Se$ ,  $H_2Se_2O_3$ ,  $H_2SeO_3$  e  $H_2SeO_4$  homólogos respectivamente a Sulfito de Hidrogênio, Tiosulfato, Sulfito, Sulfato. Quando em compostos orgânicos, geralmente é mais instável e reativo que o correspondente análogo sulfúrico, propriedade responsável pela alta toxicidade quando incorporado indiscriminado, no lugar do enxofre<sup>6,7</sup>.

Até meados dos anos de 1950, devido à sua natureza extremamente tóxica, características urticantes e forte odor de alho detectável e persistente mesmo a baixíssimas concentrações, pouco interesse de pesquisa foi destinado ao estudo do elemento Selênio<sup>8, 9</sup>. Alguns nutricionistas de animais e químicos interessaram-se pela estrutura química do composto *Se* presente em plantas, principalmente do gênero *Astragalus* que apresentam altíssimas taxas de *Se*. Entretanto, tais estudos não obtiveram sucesso provavelmente devido à instabilidade dos compostos que continham Selênio<sup>3</sup>.

Entretanto a partir da descoberta da presença do elemento Selênio em proteínas e sua atuação em inúmeros processos fisiológicos em meados de 1950, fez crescer enormemente o interesse para pesquisas relacionadas ao *Se* e seu papel biológico<sup>2, 3, 6</sup>.

Atualmente o Selênio é considerado um elemento essencial na dieta humana e de outros mamíferos<sup>3, 10</sup>, sendo atribuídos benefícios significativos a saúde. Dentre eles são destacados: agente quimiopreventivo ao câncer, devido à forte ação antioxidante; redutor da expressão viral, retardando a progressão da AIDS em pacientes HIV<sup>+</sup>; redutor de doenças cardíacas e outras desordens musculares; em processos de desenvolvimento e sistema imune; fertilidade masculina dentre outros<sup>5, 8, 10, 11</sup>. Porém para que os benefícios citados acima sejam obtidos é importante ater a quantidade de Selênio ingerida, sendo recomendável respectivamente 55 e 70 µg/dia para mulheres e homens adultos de acordo com o Conselho Nacional de Pesquisa Norte Americano<sup>7, 12</sup>. Apesar de síndromes típicas como: Doença Muscular Bovina (*White Muscle Disease in cattle*), Miopatias Suína (*Mulberry Heart Disease in pig*), Diátese Exsudativa em aves (*Exudative Diathesis in Chicks and Poults*), desordens neurológicas e danos por tecidos queratinizados (*Keratinaceous Tissues*) encontradas em animais domésticos em vários países, a falta de Selênio não é considerada um problemas para humanos em países desenvolvidos que possuam uma dieta variada<sup>9, 13</sup>. Doença inevitavelmente associada à deficiência de Selênio são reportadas em remotas áreas rurais da China com extremo déficit de Selênio em seus solos e com produção local de alimentos, como na região de Keshan. A população dessas regiões apresenta cardiomiopatias graves, infecções virais, disfunção da tireóide, problemas de fecundidade masculina, problemas de disposição/humor e concentração, além de problemas vasculares<sup>7</sup>. Entretanto,

grandes doses de Selênio são tóxicas causando a morte, desse modo, quantidades inferiores a 20 µg/dia são consideradas insuficientes e acima de 800 µg/dia tóxicas para seres humanos<sup>13</sup>.

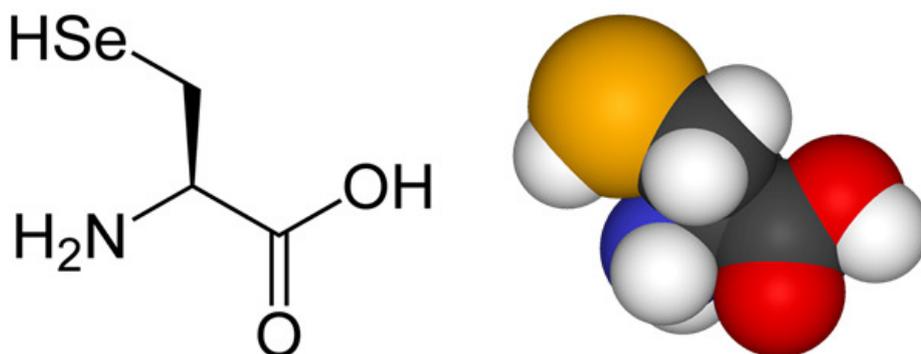
A obtenção de Selênio pelos organismos vivos ocorre através das plantas, que absorvem o Se em sua forma inorgânica, proveniente dos solos ricos em Selênio ou suplementados via fertilizantes, convertendo-o em formas orgânicas de baixa massa molecular como o aminoácido Selenometionina (SeMet), presentes em grande quantidades nos legumes, soja, cereais e grãos em geral. A SeMet é maior fonte de Selênio encontrada nos tecidos das plantas e principal precursor do aminoácido Selenocisteína (Sec)<sup>4, 7</sup>. O Selênio também pode ser obtido via suplementação alimentar encontrado nos complexos de vitaminas e sais minerais amplamente comercializados.

## 1.2 Selenocisteína

Apesar de numerosos aminoácidos resultantes de modificações pós-traducionais comumente encontrados em proteínas maduras, o código genético universal constitui-se de 22 aminoácidos. Aminoácidos esses, pré-existentes e com tRNAs próprios para inserção na cadeia polipeptídica durante o processo de tradução.

O aminoácido Selenocisteína (Sec) é considerado o vigésimo primeiro aminoácido ao logo dos 22 conhecidos<sup>10</sup> (figura 1.1). Descoberto em meados de 1960<sup>14</sup>, está presente nas três linhagens descendentes (eubactérias, arqueobactérias e eucariotos) sugerindo que sua origem precede a separação do ancestral primordial comum<sup>15, 16</sup>. Apesar de ampla presença nem todos os organismos são capazes de incorporar Selenocisteína, tais como, plantas superiores e fungo<sup>17</sup>. Sua estrutura química é similar a do aminoácido cisteína,

diferindo apenas pela substituição do elemento enxofre pelo selênio. Entretanto sua estrutura eletrônica proporciona maior estabilidade, menores valores de  $pK_a$ , uma maior atividade nucleofílica e um alto potencial redutor presente em proteínas com atividades anti-oxidantes<sup>3, 7, 17</sup>.



**Figura 1. 1** - Aminoácido Selenocisteína. visualização 2D a esquerda e 3D a direita. (Fonte Wikipedia <http://en.wikipedia.org/wiki/Selenocysteine>).

Inserido na cadeia polipeptídica em resposta ao códon UGA, por uma complexa via de incorporação capaz de diferenciar o códon para Sec ou para terminação da síntese protéica<sup>16, 18</sup>.

As proteínas e estruturas de RNA envolvidas na biossíntese, direcionamento e incorporação do aminoácido Selenocisteína na cadeia polipeptídica são:

**Seril-tRNA Sintetase (SerRS)** – proteína dimérica (95 kDa) pertencente a classe II das aminoacil-tRNA Sintetases, responsável pela aminoacilação do tRNA<sup>sec</sup> com o resíduo serina. Capaz de aminoacilar Serina tanto para o tRNA<sup>ser</sup> (99%) como para o tRNA<sup>sec</sup> (1%)<sup>19-21</sup>, pois é incapaz diferenciar a região anticódon do tRNA reconhecendo apenas suas grandes regiões do braço variável (maiores de 10 nucleotídeos em comparação aos 4-5 encontrados nos demais tRNAs)<sup>22, 23</sup>;

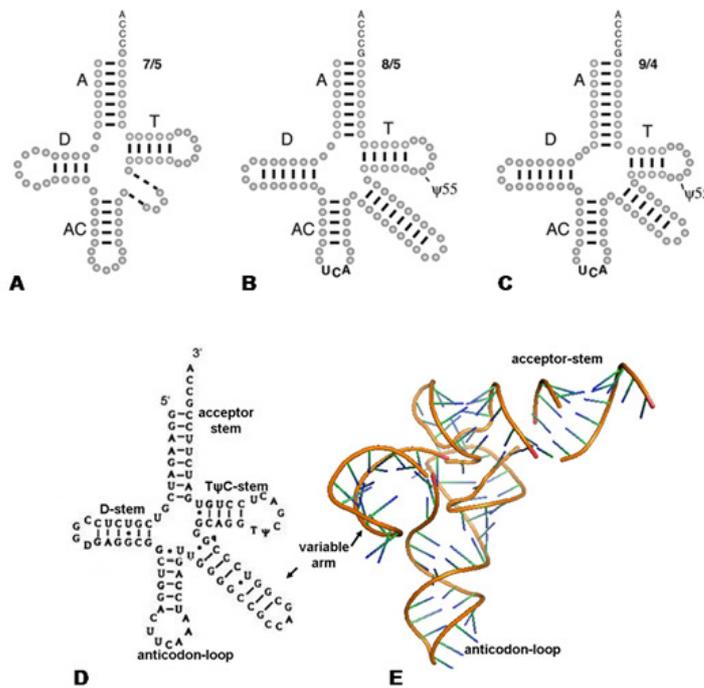
**Selenocisteína Sintase (SELA ou SepSecS)** – proteína homodecamérica (506 kDa) em eubactérias<sup>24, 25</sup>, homodimérica (84 kDa) em arqueobactéria (putativa a de eubactéria)<sup>25, 26</sup> e homotetramérica (220 kDa) em arqueobactérias e eucariotos<sup>25</sup>. É responsável pela conversão do resíduo Serina em Selenocisteína.

**Fator de Elongação de Selenocisteína (SELB ou EFSec)** – proteína monomérica (68,8 kDa) em eubactérias<sup>27</sup> e (63,5 kDa) em arqueobactérias e eucariotos<sup>28, 29</sup> cuja região N-terminal compartilha similaridades com o Fator de Iniciação-2 (IF-2) e Fator de Elongação Tu (EF-Tu). Possui alta afinidade a GTP e ao Selenociteil-tRNA<sup>sec</sup>, considerada como o Fator de Elongação para via de Selenocisteína<sup>27</sup>.

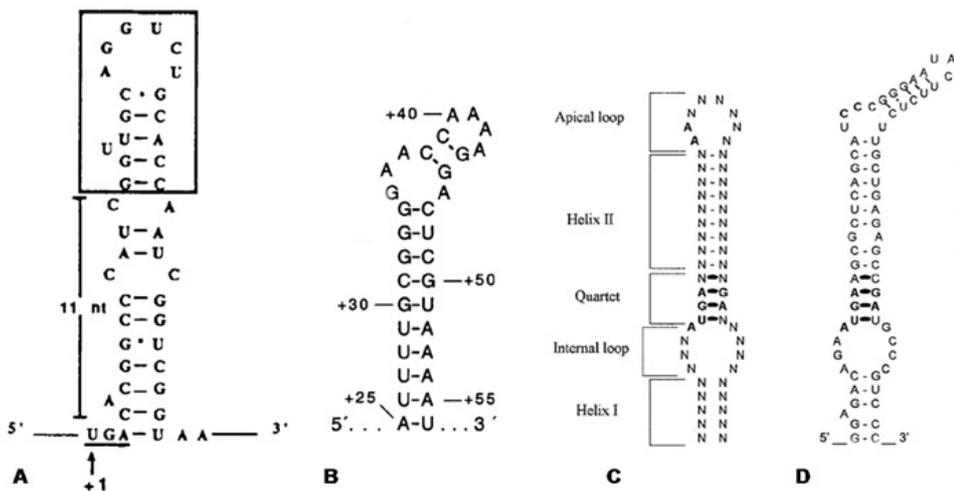
**tRNA de Inserção de Selenocisteína (SELC)** – tRNA próprio dessa via com características distintas dos demais tRNAs conhecidos. É o maior tRNA com tamanho de 95 e 90 ribonucleotídeos, distribuição 8/5 e 9/4 pares de bases nos braço acceptor / braço T respectivamente para eubactérias e eucariotos, 6 pares de base em seu braço D, maior região variável já encontrada, além da trinca UCA complementar ao códon de terminação em sua região anticódon<sup>17</sup> (figura 1.2).

**Selenofosfato Sintetase (SELD ou SPS1 / SPS2)** – proteína homodimérica (73,2 kDa) em eubactérias<sup>30</sup> e (90 / 95,4kDa) em eucariotos<sup>31, 32</sup>, responsável pelo fornecimento de Selênio na forma de monoselenofosfato para biossíntese de Selenocisteína<sup>16, 17, 33</sup>.

**Sequência de Inserção Sec (SECIS)** – uma alça (*steam loop*) formada pelo mRNA, localizada na região codificante, em eubactérias, e na região 3' – UTR, em arqueobactérias e eucariotos, o elemento SECIS possui uma estrutura conservada para cada um das três linhagens, podendo assumir duas formas nos eucariotos (figura 1.3)<sup>10, 34, 35</sup>. É responsável pela ligação dos elementos *trans* como SELB, Sec-tRNA<sup>sec</sup>, SBP2 e na identificação do códon UGA de inserção para Selenocisteína<sup>4, 16</sup>.



**Figura 1. 2** - Estrutura secundária e terciária de tRNAs. Em **A**- estrutura geralmente encontrada em tRNAs; Em **B**- tRNA de Inserção de Selenocisteína (tRNA<sup>Sec</sup>) de eubactérias; Em **C**- tRNA<sup>Sec</sup> de arqueobactérias e eucariotos; Em **D** e **E**- tRNA<sup>Sec</sup> de *Escherichia coli* representado respectivamente pela estrutura secundária e terciária<sup>17, 36, 37</sup>.



**Figura 1. 3** - Sequências de Inserção Sec (SECIS). Em **A**- SECIS para eubactérias (*E. coli*) localizado logo após o códon UGA para Sec. Delimitado pelo retângulo encontra-se a região reconhecida pela proteína SELB a exatamente 11 nucleotídeos do códon UGA; Em **B**- SECIS para arqueobactérias (*M. maripaludis*) localizado na região 3'-UTR a aproximadamente 25 nucleotídeos do stop códon; Em **C** e **D**- SECIS para eucariotos (mamíferos) em que respectivamente é ilustrada a forma I e II, contendo os espaçamentos indicados pela legenda a esquerda, o quarteto AUGA e a presença de AA ou CC na região do loop apical<sup>35, 38, 39</sup>.

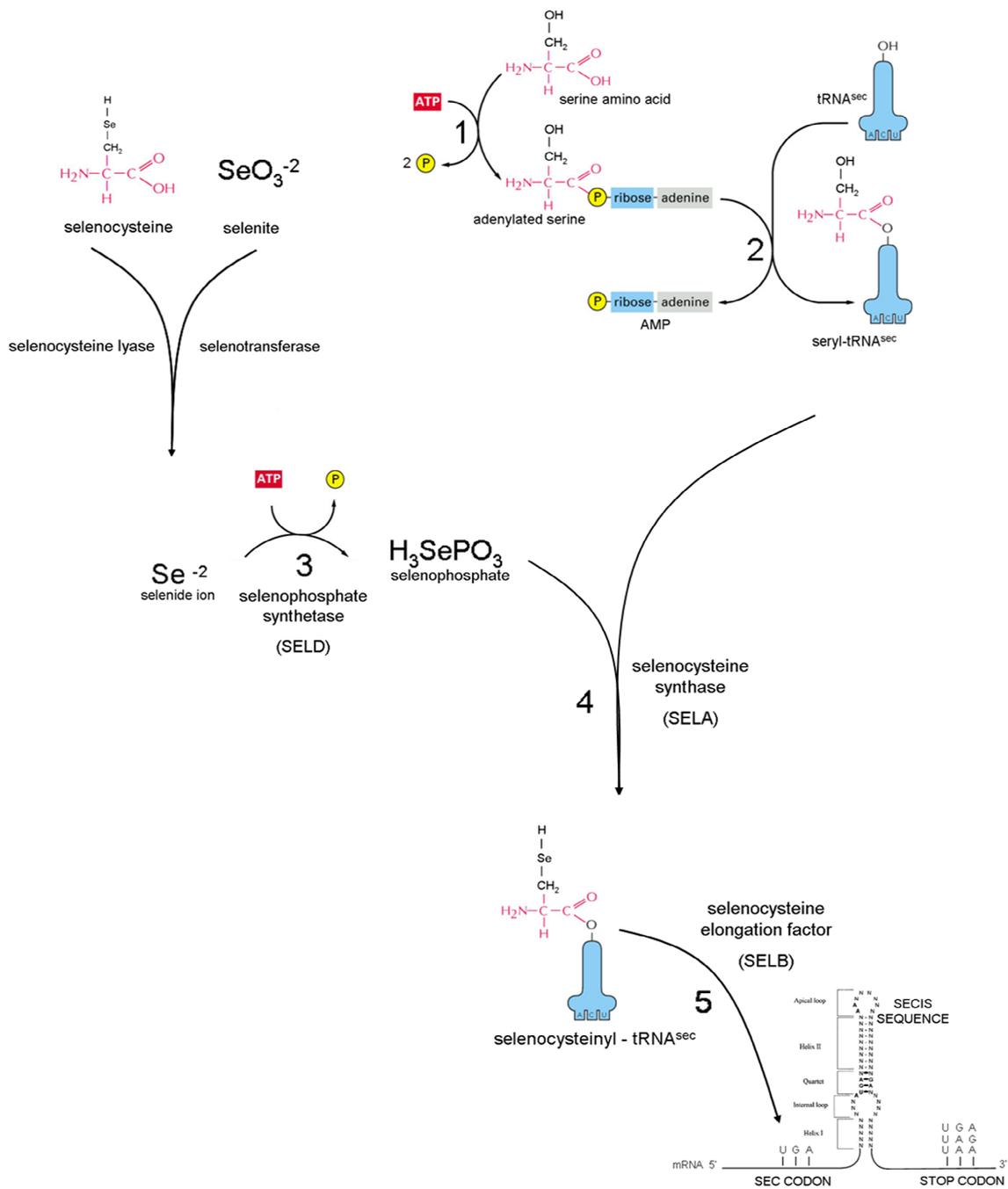
A biossíntese de Selenocisteína inicialmente ocorre com a aminoacilação do tRNA<sup>sec</sup> ao resíduo Serina pela proteína Seril-tRNA Sintetase. Após a liberação do tRNA aminoacilado (Seril-tRNA<sup>sec</sup>) (Figura 1.4 – etapa 1 a 2) a Serina é convertida à Selenocisteína pela ação da proteína Selenocisteína Sintase (Figura 1.4 – etapa 4). Em bactérias a reação ocorre de maneira direta em que cada monômero da proteína está ligado covalentemente a uma molécula de Piridoxal 5' – Fosfato (PLP), um cofator enzimático responsável pela interação com o Seril-tRNA<sup>sec</sup>, resultando na formação de uma Base de Schiff (ligação dupla entre um nitrogênio e um carbono) entre o grupamento amino  $\alpha$  do resíduo de serina e o grupamento azo-formil do piridoxal, resultando na eliminação de água e síntese de uma molécula intermediária denominada de aminoacrilil-tRNA<sup>sec</sup>. A partir desse composto intermediário o aminoácido insaturado pode tautomerizar para uma forma instável, imino, que espontaneamente hidrolisa a piruvato e amônia. Alternativamente a dupla ligação pode ser reduzida por um agente redutor (borohidrato de potássio – KBH<sub>4</sub>) resultando em alanil-tRNA<sup>sec</sup> 40.

O passo seguinte consiste na transferência do selênio reduzido e ativado sob a forma de monoselenofosfato para a molécula intermediária (aminoacrilil-tRNA<sup>sec</sup>) resultando no vigésimo primeiro aminoácido Selenocisteína liberado em seguida<sup>16, 41</sup> (Figura 1.5). Em arqueobactérias e eucariotos a conversão ocorre após a fosforilação do Seril-tRNA<sup>sec</sup> em Sep-tRNA<sup>sec</sup> através da enzima O – fosforil tRNA<sup>sec</sup> Kinase (PSTK). Desse modo, a proteína conhecida por SLA/LP (*Soluble Liver Antigen / Liver Pancreas*) de 48 kDa, co-imunoprecipitada com o tRNA<sup>sec</sup> por anticorpos em pacientes com severa hepatite crônica auto-imune, teve seu cDNA identificado por dois diferentes grupos como sendo a Selenocisteína Sintase para arqueobactérias e eucariotos<sup>16</sup>. Em ratos foi nomeada por mSecS pelos grupos dos respectivos pesquisadores Dieter Söll e Osamu Nureki<sup>42</sup> enquanto em humanos e arqueobactérias foi nomeada SepSecS pelo grupo do pesquisador Marcus C. Wahl<sup>25</sup>. Também é uma enzima pertencente à família de proteínas dependentes de PLP, composta de dois monômeros que interagem formando um dímero que se associa formando um homotetramero, que após eliminação do Sep fosfato o monoselenofosfato interage com o grupamento aminoacrilil-tRNA<sup>sec</sup> conforme passos descritos para SELA de bactéria<sup>16</sup>.

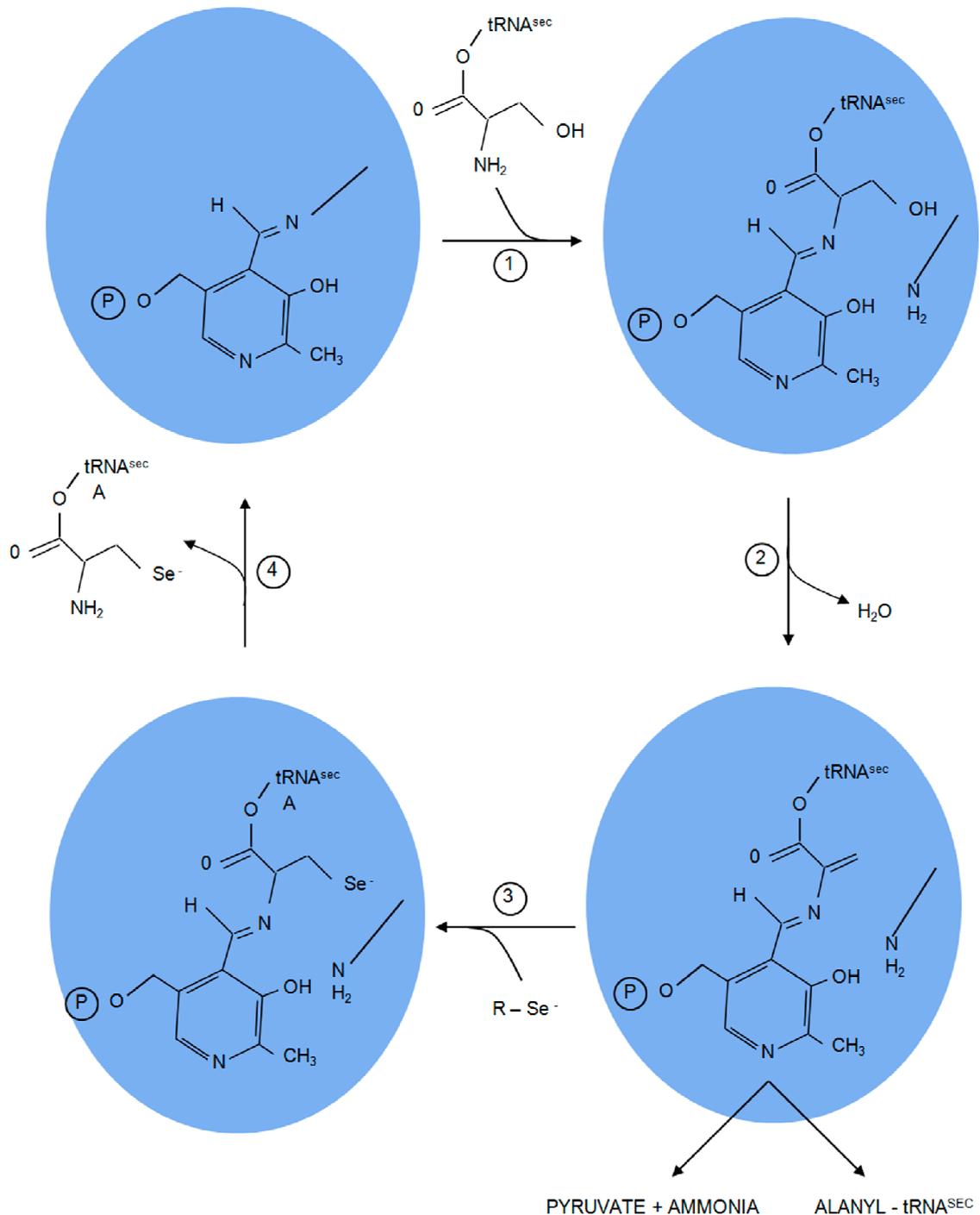
O Selênio é obtido a partir do Selenito, um sal com ânion divalente ( $\text{SeO}_3^{2-}$ ), ou de Selenocisteínas respectivamente pelas proteínas Selenotransferases e Selenocisteína  $\beta$  Liase que reduzem o Selênio<sup>43, 44</sup> (Figura 1.4). O Selênio reduzido é então ativado na forma de monoselenofosfato pela enzima Selenofosfato Sintetase, numa reação dependente de ATP (Figura 1.4 – etapa 3). Em bactérias é conhecida apenas uma forma de Selenofosfato Sintetase, denominada SELD, enquanto em arqueobactérias e eucariotos existem duas proteínas: SPS1 e SPS2<sup>45</sup>. SELD e SPS apresentam apenas 32% de identidade<sup>17, 31</sup>, entretanto SPS compartilham alta identidade sequencial e atividade para produção do doador de Selênio, monoselenofosfato, utilizado pela Selenocisteína Sintase. A principal diferença entre as duas SPS deve-se ao fato de SPS2 ser uma Selenoproteína, isto é, possuir em sua composição peptídica o aminoácido Selenocisteína. Como Selenocisteína é requerido para a tradução de uma selenoproteína, caso contrário a sua síntese são interrompidas resultando em proteínas incompletas, foi proposto que a proteína SPS1 contribuísse para a manutenção basal dos níveis de Selenocisteína, função posteriormente transferida para SPS2<sup>17</sup>. Entretanto, estudos mais recentes identificaram que apesar de SPS1 apresentar atividade *in vitro*, apenas SPS2 é essencial para síntese de selenoproteínas<sup>16</sup>.

O direcionamento do aminoácido Selenocisteína à cadeia polipeptídica das Selenoproteínas, no ribossomo durante a tradução, deve-se a ação do Fator de Elongação de Selenocisteína. Entretanto para que haja tal incorporação é necessária a existência de uma alça (*steam loop*) *cis*-mRNA, isto é, na mesma fita de RNA mensageiro denominada de Sequência de Inserção de Selenocisteína (Figura 1.4 – etapa 5). O elemento SECIS, como é conhecido, atua como sinalizador para que o códon UGA presente no mRNA seja identificado como um códon para Selenocisteína e não um códon de terminação. Sua localização é responsável pelo número de Selenocisteína incorporado nas Selenoproteínas. Em bactérias sua posição localizada logo após o códon de inserção para Selenocisteína e previamente ao códon de terminação, ainda na fase aberta de leitura, permite que apenas um resíduo seja incorporado e sempre no final da sequência polipeptídica. Em arqueobactérias e eucariotos sua localização na região 3'-UTR permite que juntamente com fatores *trans*, como a proteína SBP2, se associem ao EFSec

carregado com o Sec-tRNA<sup>sec</sup> para inserção ao longo de posições variadas nas Selenoproteínas, num processo ainda não elucidado em eucariotos<sup>4, 10</sup>.



**Figura 1. 4** - Etapas envolvidas na biossíntese e incorporação de Selenocisteína de *Escherichia coli* conforme descrito no texto. A etapa 4 pode ser visualizada em detalhes na figura 1.5 na página seguinte.

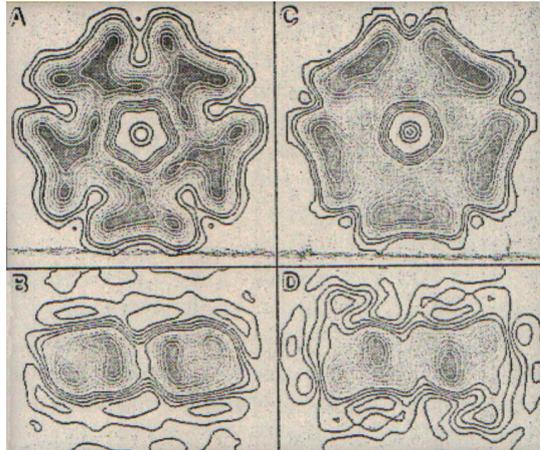


**Figura 1.5** - Esquema detalhado da conversão do aminoácido serina em selenocisteína pela enzima Selenocisteína Sintase (SELA) parcialmente representada em azul. Também indicado ligado à enzima SELA o cofator enzimático Piridoxal 5'- Fosfato: 1- Ligação do seril-tRNA<sup>sec</sup> ao cofator piridoxal 5'-fosfato. 2- liberação de uma molécula de água com a formação de um composto intermediário (aminoacrilil-tRNA<sup>sec</sup>). Nessa etapa o composto intermediário pode tautomerizar liberando Piruvato e Amônia (espontaneamente), ou devido a um agente redutor (borohidratado de potássio) liberar alanil-tRNA<sup>sec</sup>. 3- transferência do selênio reduzido e ativado à molécula intermediária, resultando no selenocisteil-tRNA<sup>sec</sup>. 4- liberação do aminoácido selenocisteína, disponibilizando a enzima SELA e o cofator Piridoxal 5'- Fosfato para reinício do ciclo<sup>40</sup>.

### 1.3 Selenocisteína Sintase (SELA) de *Escherichia coli*

A proteína Selenocisteína Sintase (SELA) de *Escherichia coli* foi primeiramente caracterizada em 1991 pelo grupo de August Böck<sup>24, 40</sup>. A partir da clonagem do seu gene em vetor de expressão foi possível a determinação da sequência de nucleotídeos com 1392 pb (463 aminoácidos), calcular sua massa molecular para 50,667 kDa e pI de 6,01. A superexpressão e o desenvolvimento de uma metodologia de purificação, possibilitaram experimento que determinaram que a proteína SELA apresentava 54% de  $\alpha$  – hélices em sua estrutura secundária, massa molecular de aproximadamente 51 kDa em gel de poliacrilamida desnaturante e de aproximadamente 600 kDa em gel de poliacrilamida não desnaturante e por cromatografia de exclusão de tamanho. Experimentos também confirmaram sua atividade enzimática na conversão do resíduo Serina em Selenocisteína, desde que ligado ao tRNA<sup>sec</sup> (Seril-tRNA<sup>sec</sup>) e na presença da Selenofosfato (provido pela enzima Selenofosfato Sintetase) e de Piridoxal 5' – Fosfato (PLP) cofator enzimático ligado covalentemente à proteína SELA na proporção 1 : 1 (monômero de SELA : PLP). A descrição passo a passo da conversão Serina – Selenocisteína é apresentada no item 1.2 Selenocisteína e figura 1.4 e 1.5.

O primeiro estudo estrutural realizado com a proteína SELA datado de 1992 por Microscopia Eletrônica de Transmissão (TEM), realizados por Herald Engelhardt<sup>46</sup>, mostrou que a proteína possuía uma simetria de ordem cinco, cinco subunidades diméricas, interconectadas por uma base semelhante a um anel (figura 1.6 – **A** – **B**). As dimensões calculadas foram: um diâmetro de 19 nm; um orifício central de aproximadamente 4 nm; cada subunidade dimérica com 8,5 nm de comprimento por 5 a 6 nm de espessura.



**Figura 1. 6** - Imagens da proteína Selenocisteína Sintase (SELA) de *E.coli* obtidas pela técnica de Microscopia Eletrônica de Transmissão (TEM) por “*negative stain*”, utilizando como marcadores acetato de urânio (A, C, D) e oxalato de urânio (B). Em **A**- vista frontal da proteína SELA, ilustrando a simetria encontrada; em **B**- vista lateral, ilustrando uma subunidade dimérica. Em **C** e **D**- vista frontal e lateral respectivamente da proteína SELA ligada ao composto intermediário aminoacrilil-tRNA<sup>sec</sup>. (Reproduzido de Harald Engelhardt<sup>46</sup>).

Devido a grande massa (31 kDa), bem como um grande braço variável presente no tRNA<sup>sec</sup> é possível sua detecção na imagens de TEM. Desse modo, uma razão molar de uma molécula de Seril-tRNA<sup>sec</sup> por monômero de SELA foi preparada, promovendo intermediários até o carregamento total dos sítios de ligação a proteína (1 Seril-tRNA<sup>sec</sup> : 1 monômero de SELA). Os dados obtidos revelaram que a ligação ocorreu nas extremidades das subunidades dimérica, mas não internamente ao orifício central, de modo que apenas uma molécula de tRNA ligava-se por dímero. A ligação ao dímero pode ser tanto no topo como na base do dímero. Nenhuma imagem com dois tRNAs por dímero foi observada (figura 1.6 – **C – D**).

Sabendo-se que uma molécula de PLP (265,2 Da) liga-se covalentemente a um monômero de SELA (50,667 kDa); que os dados experimentais por cromatografia de exclusão de tamanho e gel não desnaturante apontaram valores de massa próximos a 600 kDa; que os cálculos de massas obtidos por TEM para 493 +/- 20 kDa (com 95% de confiabilidade) e que a imagens obtidas por TEM indicaram uma simetria de cinco subunidades diméricas. Ficou determinado que a proteína SELA apresenta-se homodecamérica constituída por cinco subunidades diméricas e com massa molecular de 509,322 kDa.

Buscas por proteínas homólogas a SELA de *E. coli* em arqueobactérias e eucariotos identificaram apenas em arqueobactérias uma proteína homodimérica putativa, MJ0158, cristalizada em *Methanococcus jannaschii*<sup>26</sup>. Em arqueobactérias e eucariotos foi identificado como Selenocisteína Sintase a proteína até então conhecida como SLA/LP (*Soluble Liver Antigen / Liver Pancreas*), na forma homotetramérica e renomeada para SepSecS<sup>25</sup>.

A proteína SELA de *E. coli* teve sua nomenclatura sistematizada para Selenofosfato:L-seril-tRNA<sup>sec</sup> Selênio Transferase (*Selenophosphate: L-seryl-tRNA<sup>sec</sup> Selenium Transferase*) e EC 2.9.1.1, segundo a “*International Union of Biochemistry and Molecular Biology*” (<http://www.chem.qmul.ac.uk/iubmb/>). Análises de alinhamentos múltiplos seguido por mutações sítio-dirigidas identificaram o resíduo Lisina (K-295) como responsável pela ligação ao cofator PLP e diretamente relacionado a sua atividade enzimática<sup>47</sup>.

## 1.4 Selenoproteínas

A principal forma biológica do elemento Selênio encontra-se quando incorporado em proteínas, Selenoproteínas, através do aminoácido Selenocisteína<sup>16</sup>. A maior parte das Selenoproteínas já caracterizadas são enzimas, envolvidas em processos de oxi-redução, atuando como antioxidantes. Devido as propriedades químicas do elemento Selênio (menores valores de pK<sub>a</sub> e maior atividade nucleofílica comparado ao Enxofre) o resíduo Selenocisteína localizado no sítio ativo, proporciona atividade aumentada comparada as demoninadas Cys-Proteínas (proteínas homólogas as Selenoproteínas que possuem o resíduo Cisteína substituindo o resíduo Selenocisteína)<sup>7, 16, 17</sup>.

Apesar da maior atividade presente nas Selenoproteínas, inúmeras Cys-Proteínas são encontradas, gerando interesse sobre as origens e direções seguidas

pelas Selenoproteínas: O resíduo Selenocisteína seria um resíduo ancestral que vem sofrendo grande perda ao longo da evolução devido a sua alta reatividade, ou estaria sofrendo pressões evolutivas para o desenvolvimento das Selenoproteínas?

A resposta ainda não é conclusiva, sabendo-se da existência de Selenoproteínas nas três linhagens de organismos existente (eubactérias, arqueobactérias e eucariotos) o que sugere uma origem primitiva, precedente a separação do ancestral comum<sup>15, 16</sup>. Entretanto, também é sabido que nem todos os organismos existentes apresentam Selenoproteínas, como fungos e plantas superiores e que dessa forma as selenoproteínas não são essenciais para a vida por si só<sup>7, 17</sup>.

Com relação à grande quantidade de Cys-Proteínas encontradas em bactérias, arqueobactérias e eucariotos, amplamente distribuída em procariotos e aproximadamente 80% em humanos, tais proteínas compensam suas baixas atividades com uma maior quantidade de proteínas expressas por células<sup>7</sup>.

A busca por Selenoproteínas nos diferentes genomas já seqüenciados tem identificado através de metodologias computacionais e posterior comprovação por técnicas de biologia molecular um número cada vez maior de proteínas<sup>48, 49</sup>. Até o momento as Selenoproteínas são agrupadas em 30 famílias, distribuídas entre as espécies levando-se em conta a distribuição geográfica do elemento Selênio, organismos aquáticos apresentam maiores quantidades de Selenoproteínas comparados aos organismos terrestres<sup>49, 50</sup>. Revisões e trabalhos de busca ou identificações de Selenoproteínas distribuíram 31 Selenoproteínas entre os procariotos, 30 entre os peixes, 25 entre os mamíferos<sup>4, 48, 49</sup>. A tabela 1.1 ilustra as Selenoproteínas identificadas para procariotos e mamíferos.

**Tabela 1. 1** - Selenoproteínas encontradas em procariotos e mamíferos organizadas em ordem alfabética. Dados obtidos a partir de revisões e trabalhos de busca e identificações de Selenoproteínas<sup>4, 48, 49</sup>.

<b>Selenoproteínas de Procariotos</b>	<b>Selenoproteínas de Mamíferos</b>
AhpD-like protein	Glutathione Peroxidase 1
Arsenate Reductase	Glutathione Peroxidase 2
CMD domain-containing protein	Glutathione Peroxidase 3
Deiodinase-like protein	Glutathione Peroxidase 4
DsbA-like protein	Glutathione Peroxidase 6
DsbG-like protein	Iodothyronine Deiodinase 1
DsrE-like protein	Iodothyronine Deiodinase 2
F420-Reducing Hydrogenase Alpha Subunit	Iodothyronine Deiodinase 3
F420-Reducing Hydrogenase, Delta Subunit	Selenophosphate Synthetase 2
Formate Dehydrogenase Alpha Chain	Selenoprotein 15 kDa
Formylmethanofuran Dehydrogenase	Selenoprotein H
Glutaredoxin	Selenoprotein I
Glutathione Peroxidase	Selenoprotein K
Glutathione S-transferase	Selenoprotein M
Glycine Reductase Selenoprotein A	Selenoprotein N
Glycine Reductase Selenoprotein B	Selenoprotein O
HesB-like protein	Selenoprotein P
Heterodisulfide Reductase	Selenoprotein R
Hypothetical Protein 1	Selenoprotein S
Methylviologen-Reducing Hydrogenase	Selenoprotein T
Molybdopterin Biosynthesis MoeB protein	Selenoprotein V
NADH:Ubiquinone Oxidoreductase	Selenoprotein W
OsmC-like protein	Thioredoxin Reductase 1
Peroxiredoxin	Thioredoxin Reductase 2
Proline Reductase PrdB	Thioredoxin Reductase 3
Prx-like protein	
Rhodanase-Related Sulfurtransferase	
Selenophosphate Synthetase	
SelW-like protein	
Thiol-Disulfide Isomerase-like protein	
Thioredoxin	

## CAPÍTULO 2

### 2 OBJETIVOS

O presente estudo tem como objetivo a determinação estrutural da proteína Selenocisteína Sintase (SELA) de *Escherichia coli*, em sua forma apo e ligada o tRNA de Inserção Selenocisteína (tRNA<sup>sec</sup>), utilizando se de:

**Modelagem Tridimensional** – modelo em formato PDB produzido a partir de alinhamentos sequenciais e estruturas cristalográficas homólogas resolvidas à alta resolução;

**Estudos de Conservação e Acoplamento Estatístico** – análises da conservação e relação entre as posições dos aminoácidos, buscando identificar resíduos relacionados à oligomerização da estrutura homodecamérica, a ligação ao tRNA<sup>sec</sup> e ao sítio ativo;

**Espalhamento de Raios-X a Baixos Ângulos (SAXS)** – estudo estrutural a baixa resolução, realizados em solução, proporcionando a determinação estrutural mais próxima da encontrada fisiologicamente;

**Cristalização e Difração de Raios-X de Monocristais** – triagens a partir de fatoriais comerciais de cristalização e condições de refinamento visando a cristalização e melhoria dos cristais obtidos para posterior difração e determinação da estrutura cristalográfica.

Dando continuidade aos estudos iniciados durante o mestrado do aluno Alexandre Cassago, os objetivos acima fazem parte da linha de pesquisa estabelecida pelo grupo do Prof. Dr. Otavio Henrique Thiemman, que visa estudar a via de síntese do aminoácido Selenocisteína em diferentes organismos.

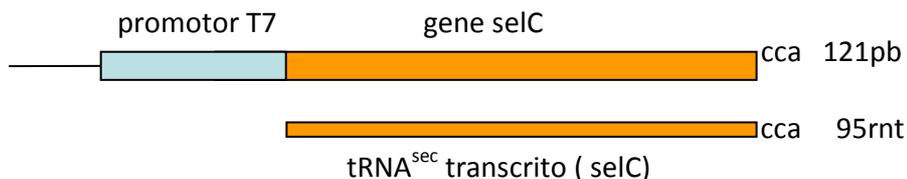
## CAPÍTULO 3

### 3 PROTEÍNA SELENOCISTEÍNA SINTASE E tRNA<sup>SEC</sup>

#### 3.1 Preparação da proteína Selenocisteína Sintase e tRNA<sup>sec</sup>

A proteína Selenocisteína Sintase (SELA) inicialmente descrita e purificada em 1991<sup>24, 40</sup>, teve seu gene amplificado, pela técnica de PCR (*Polimerase Chain Reaction*), a partir de DNA genômico de bactéria *Escherichia coli* e posterior subclonagem em vetor de expressão pET29a (Novagen). Após análise por sequenciamento, a construção pET29a-SELA foi transformada em bactérias *E. coli* cepas BL21(DE3) competentes para expressão. A purificação seguiu nova metodologia desenvolvida, proporcionando maiores rendimentos de proteínas, necessários aos estudos estruturais<sup>51</sup>.

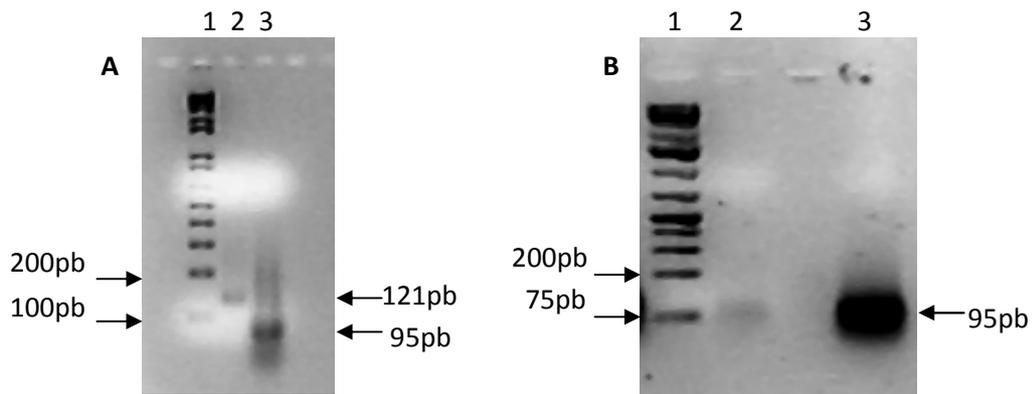
O gene de Inserção Selenocisteína – tRNA<sup>sec</sup> (selC), cuja sequência depositada no banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov>) sob os códigos de acesso Y00299<sup>52</sup> e X15993<sup>53</sup>, foi amplificado utilizando os oligonucleotídeos SELC-1eco2 (5'-ACGAATTCTAATACGACTCACTATAGGGAAGATCGTCTCTCCGGTGAGGCGGCTGGACTTCAAATCCAGTTG-3') e SELC-2eco2 (5'-TGGGGAAGATCACAGGAGTCGAACCTGCCCGGGACCGCTGGCGGCCCAA CTGGATTTGAAGTCCAG-3'). Tais oligonucleotídeos possuem uma região interna complementar de 20 nucleotídeos (grifos contínuos) e de uma sequência promotora T7 RNA polimerase (grifo descontínuo) em sua região 5' (SELC-1eco2) (figura 3.1).



**Figura 3. 1** - Esquema para amplificação e transcrição *in vitro* do gene *selC* de *E. coli*. Em azul a região promotora T7 RNA Polimerase, seguida em alaranjado pela sequência do gene *selC*. Abaixo o transcrito de 95rnt com sua terminação CCA comum a todos os tRNAs.

O fragmento de 121pb amplificado (região promotora T7 + gene de Inserção Selenocisteína tRNA<sup>sec</sup>) foi purificado de gel de agarose 2% (Wizard SV Gel and PCR Clean Up System - Promega) e utilizado como DNA molde para a reação de transcrição *in vitro* do tRNA<sup>sec</sup>.

Inicialmente utilizou-se para transcrição *in vitro* variações dos protocolos encontrados na literatura<sup>54, 55</sup>, em que grande quantidade de DNA molde (2µg), tempo de incubação (24h) e quantidade de reagentes além da enzima T7 RNA polimerase (Invitrogen) eram empregados, resultando infelizmente em quantidades pequenas de tRNA transcrito. Visando o ganho de rendimento e praticidade na reação de transcrição, foi empregado o “kit Megashortscript – Ambion” desenvolvido para transcrição de pequenas sequências de RNA. Essa nova abordagem resultou em menores quantidades de DNA molde (0,2µg), tempo de incubação (4 - 8h) e praticidade na preparação da reação. O tRNA transcrito foi purificado pelo protocolo de precipitação por fenol/cloroformio<sup>56</sup>. A comparação entre os transcritos provenientes dos dois protocolos são visualizados na figura 3.2, em que a transcrição pela Ambion apresentou superioridade de 13,5 vezes mais tRNA em quatro horas de incubação.



**Figura 3. 2** - Amplificação e transcrição *in vitro* do gene de inserção selenocisteína tRNA<sup>sec</sup> (selC) e tRNA<sup>sec</sup> visualizado em gel de agarose 2% - TAE [1X] corado com brometo de etídeo. Em **A**- coluna 1- padrão molecular "1Kb Plus DNA Ladder"; coluna 2- amplificação do gene selC portando a região promotora T7; coluna 3- transcrição do tRNA<sup>sec</sup>. Em **B**- coluna 1- padrão molecular "1Kb Plus DNA Ladder"; coluna 2- transcrição do tRNA<sup>sec</sup> pelo protocolo adaptado da literatura; coluna 3- transcrição do tRNA<sup>sec</sup> utilizando o "Kit Megashortscript – Ambion".

### 3.2 Interação proteína SELA e tRNA<sup>sec</sup>

A proteína Selenocisteína Sintase (SELA) purificada em tampão I (20 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub> pH 7,5) contendo 10 μM do cofator enzimático Piridoxal 5'- Fosfato (PLP) foi concentrada após a adição de glicerol (10% concentração final) até atingir as concentração utilizadas nos experimentos estruturais.

O enovelamento tridimensional do tRNA<sup>sec</sup> foi conseguido inicialmente pela ressuspensão do mesmo em tampão II (40 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub> pH 7,5; 40 mM KCl; 20 mM MgCl<sub>2</sub> e 10 mM DTT). Após sua ressuspensão, o tRNA<sup>sec</sup> foi aquecido em banho até 70° C por dez minutos para sua total desnaturação e através do lento resfriamento da temperatura até 25° C para seu enovelamento.

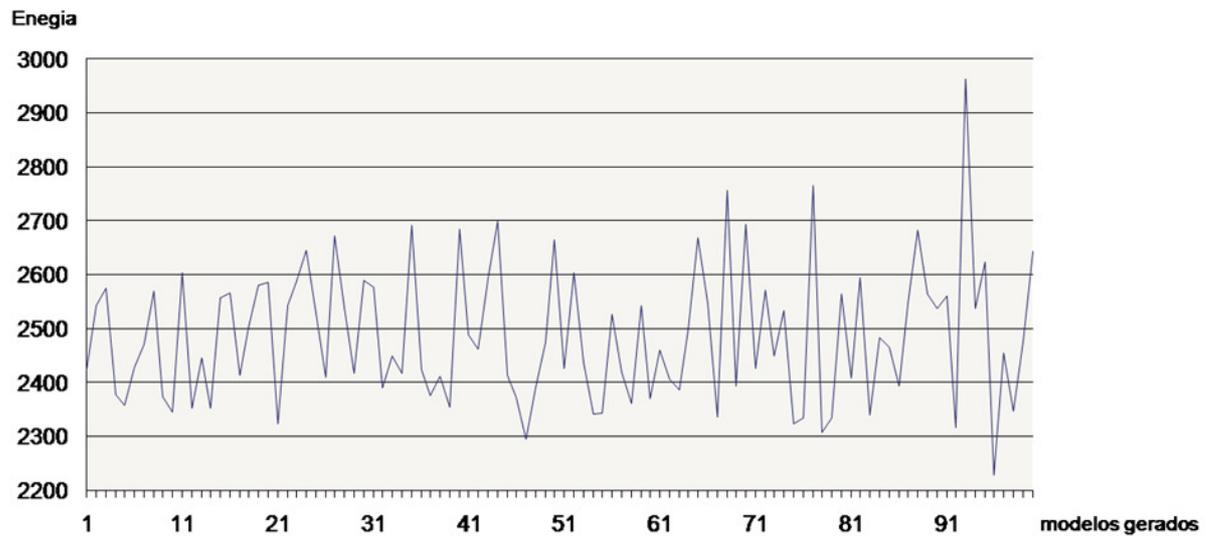
O complexo SELA-tRNA<sup>sec</sup> foi obtido incubando-se volumes iguais de proteína e tRNA<sup>sec</sup> a 25° C por 20 minutos. Desse modo, a concentração do tampão fosfato de potássio pH 7,5 manteve-se a 20mM, enquanto as concentrações de KCl, MgCl<sub>2</sub> e DTT foram reduzidas a metade.

Para a determinação do tampão de ressuspensão, bem como, temperatura de desnaturação/enovelamento do tRNA, tempo de incubação e cálculos das relações molares entre proteína-tRNA, diferentes fontes da literatura foram utilizadas<sup>22-24, 40, 46, 47</sup>.

### 3.3 Modelagem da proteína SELA de *Escherichia coli*

A proteína Selenocisteína Sintase (SELA) de *Escherichia coli* foi modelada utilizando o programa MODELLER<sup>57</sup>. Um programa capaz de gerar modelos tridimensionais no formato PDB, a partir de alinhamentos entre a sequência protéica de interesse, a ser modelada, com sequências protéicas de estruturas já resolvidas. Desse modo, o programa calcula automaticamente restrições para as distâncias  $C_{\alpha} - C_{\alpha}$  e ângulos diédricos na sequência alvo, a partir do alinhamento com o modelo tridimensional. As restrições são obtidas de análises estatísticas entre relações de pares de estruturas homólogas, proveniente de bancos de dados de proteínas com estruturas tridimensionais conhecidas. Para cada modelo gerado um valor de energia resultante das análises de restrição é obtido e utilizado como guia na busca de um modelo mais próximo da realidade.

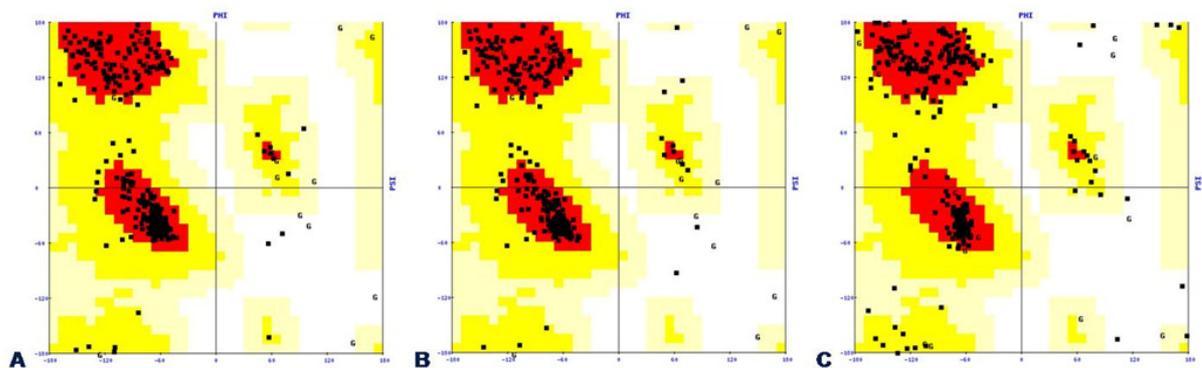
O modelo para proteína SELA foi preparado a partir do alinhamento, com 30% de identidade, entre as sequências protéicas SELA de *E. coli* (procarionte) e SELA putativa de *Methanococcus jannaschii* (arqueobactéria), cuja estrutura cristalográfica foi depositada no "Protein Data Bank" sob os códigos 2AEU e 2AEV<sup>26</sup>. Após o alinhamento entre as duas sequências pelo servidor MULTALING<sup>58</sup>, um total de cem modelos foram gerados, ordenados e analisados com relação às respectivas energias (figura 3.3).



**Figura 3. 3** - Gráfico de energia, resultante das análises de restrições, para os modelos da proteína SELA gerados pelo programa MODELLER. O modelo de número 96 apresenta a menor energia (2228,3464).

Os modelos que possuíam menores energias foram validados pelo servidor PROCHECK<sup>59</sup>. Desse modo, a qualidade estereoquímica da estrutura protéica é analisada tanto resíduo por resíduo, quanto sua geometria como um todo. Essa análise verifica quão normal ou não-usual é a geometria de um dado resíduo, a partir de comparações estereoquímicas com estruturas refinadas a altas resoluções.

O modelo de número 96 apresentou a menor energia e as melhores distribuições de resíduos nas regiões permitidas no diagrama de Ramachandran comparado aos PDBs cristalográficos de arqueobactéria (figura 3.4).



**Figura 3. 4** - Distribuição dos resíduos de aminoácidos no diagrama de Ramachandran. **A**- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* - PDB 2AEU; **B**- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* - PDB 2AEV; **C**- estrutura modelada da proteína SELA de *E. coli*.

As análises do servidor PROCHECK ordenam os resíduos em regiões mais favoráveis, permitidas, generosamente permitidas e não permitidas, sugerindo que um bom modelo deva ter mais de 90% dos seus resíduos nas regiões mais favoráveis. Entretanto, aqueles resíduos indicados em regiões não usuais não necessariamente resultam de erros, podendo ser resultados de características particulares da estrutura, por exemplo, de sítios ativos e seus ligantes. Em vista disso, tais resíduos devem ser analisados cuidadosamente. As análises de distribuição dos resíduos dos PDBs de *M. jannaschii* 2AEU / 2AEV e PDB modelado de *E. coli*, são visualizados na tabela 3.1.

**Tabela 3. 1** - Distribuição em porcentagem dos resíduos de aminoácidos no diagrama de Ramachandran para os dois PDBs cristalográficos de *M. jannaschii* e o PDB modelado para *E. coli*.

	SELA de <i>M. jannaschii</i>		SELA de <i>E. coli</i>
	2AEU	2AEV	modelo 96
<b>Regiões mais favoráveis</b>	91,7 %	91,1 %	84,8 %
<b>Regiões permitidas</b>	7,3 %	7,3 %	11,6 %
<b>Regiões generosamente permitidas</b>	0,3 %	0,9 %	2,5 %
<b>Regiões não permitidas</b>	0,6 %	0,6 %	1,1 %

Devido ao não alinhamento da região N-terminal da proteína SELA de *E. coli* e *M. jannaschii*, o alinhamento usado para gerar o modelo PDB de *E. coli* iniciou-se na posição 51 da mesma e nono da estrutura cristalográfica (figura 3.5). Evitando dessa forma, a criação de uma longa região flexível e irreal no modelo procarioto.

	1				50
SELAeco	MTTETRFLYS	QLPAIDRLLR	DSSFLSLRDT	YGHTRVVELL	RQMLDEAREV
SELAmj	-----	-----	-----	-----	--MLSDYEEF
	51				100
SELAeco	IRGSQTLPAW	CENWAQEVDA	RLTKEAQSAL	RPVINLTGTV	LHTNLGRALQ
SELAmj	LR-----	LEKARKIILE	ILNEKGRDAL	YDLSGLSGGF	LIDEKDKALL
	101				150
SELAeco	AEAAVEAVAQ	AMRSPVTLEY	DLDDAGRGRH	DRALAQLLRC	ITGAEDACIV
SELAmj	NTY----IGS	SYFAEKVNEY	GLKHLGGDEN	DK-----	-----CVG
	151				200
SELAeco	NNNAAVLLM	LAATASGKEV	VVSRGELVEI	GGAFRIPDVM	RQAGCTLHEV
SELAmj	FNRTSSAILA	TILALKPKKV	IHYLPELPGH	PSIERSCKIV	NAKYFESDKV
	201				250
SELAeco	GTTNRTHAND	YRQAVNENTA	LLMKVHTSNY	SIQGFTKAID	EAELVALGKE
SELAmj	GEILNKIDKD	TLVIITGSTM	DLKVIELN-	----FKKVIN	TA-----KN
	251				300
SELAeco	LDVPPVTDLG	SGSLVDLSQY	GLPKPEMPQE	LIAAGVSLVS	FSGDKLLGGP
SELAmj	KEAIVFVDDA	SGARVRL---	-LNFQPPALK	L---GADLVV	TSTDKLMEGP
	301				350
SELAeco	QAGIIVGKKE	MIARLQSHPL	KRALRADKMT	LAALATLRL	YLHPEALSEK
SELAmj	RGLLLAGKKE	LVDKIYIEGT	KFGLEAQPLL	LAGIYRALKN	FNLERIRKAF
	351				400
SELAeco	LPTLRLLTRS	AEVIQIQAGR	LQAPLAAHYG	AEFAVQVMPC	LSQIGSGSLP
SELAmj	ERAKNFDLSK	IEKLNKELKA	IDDNINIVYE	RTPTGFVIKR	VYK--DDTIN
	401				450
SELAeco	VDRLPSAALT	FTPHDGRGS-	HLESLAARWR	ELPVPVIGRI	YDGRLLWLHLR
SELAmj	IKKLIIEIGFN	LLKNYGIITI	TVAGMPGASK	SLRIDLTSR-	-DAERIDDNY
	451	465			
SELAeco	CLEDEQRFLE	MLLK-			
SELAmj	IIKAIVESIK	MAFKS			

**Figura 3. 5** - Alinhamento entre as sequências da proteína Selenocisteína Sintase de *E. coli* (SELAeco) e *M. jannaschii* (SELAmj). As setas indicam os resíduos localizados nas regiões não permitidas (↓), generosamente permitidas (↓), Lisina do sítio ativo (↓) e região conservada estruturalmente próxima ao sítio ativo (↓) descritas no texto.

O diagrama de Ramachandran para o modelo 96 indica quatro resíduos na região não permitida (A-104, Y-180, A-283 e F-383). Os três primeiros resíduos pertencem a regiões não alinhadas, visualizados como alças entre as

estruturas secundárias da proteína, o quarto resíduo apesar de alinhar a uma Prolina, mantém sua conformação muito parecida ao PDB molde. Seis dos nove resíduos colocados em regiões generosamente permitidas (A-59, E-106, A-107, S-268, A-284 e I-440) também são resíduos não alinhados, os três resíduos restantes (A-76, E-423 e S-424) alinham respectivamente a G-26, A-334 e G-335 da proteína de arqueobactéria (figura 3.5). Os demais aminoácidos localizam-se nas regiões mais favoráveis e permitidas, e quando somados (96,4 %) apresentam valores próximos aos encontrados nos modelos cristalográficos.

Também é visualizado, na figura 3.5, o resíduo Lisina ligante ao cofator PLP (K-295 e 208 respectivamente em *E. coli* e *M. jannashii*)<sup>26</sup>, também descrito em *D. baculatum*, *H. influenzae* e *M. thermoacetica*<sup>47</sup>. Tanto a Lisina do sítio ativo como os resíduos ao seu redor, encontram-se distribuídos na região mais favorável do diagrama de Ramachandran.

A proteína SELA de arqueobactéria é considerada putativa devido a falta de atividade, encontrada nos ensaios *in vitro*, em converter o tRNA<sup>sec</sup> carregado tanto pela serina (Ser-tRNA<sup>sec</sup>) como pela serina fosforilada (Sep-tRNA<sup>sec</sup>) em selenocisteína (Sec-tRNA<sup>sec</sup>)<sup>26</sup>. Apesar da baixa identidade sequencial (30 %) e diferença de 89 resíduos quando comparada a sequência procariótica, a SELA de arqueobactéria apresenta-se dimérica com grande semelhança as subestruturas (extremidades diméricas) do decâmero formado pela SELA de *E. coli*<sup>26, 46</sup>. Resultante dessas semelhanças estruturais foi possível gerar um bom modelo como visualizado na figura 3.6. As análises e tratamento das imagens foram realizados pelo programa PyMOL<sup>60</sup>.



**Figura 3. 6** - Modelo tridimensional da proteína SELA gerado pelo programa MODELLER. **A**- estrutura cristalográfica da proteína putativa SELA de *M. jannaschii* (PDB 2AEU); **B**- modelo gerado da proteína SELA de *E. coli*; **C**- sobreposição das duas estruturas.

### 3.4 Estudos de conservação e acoplamento estatístico da proteína SELA

Com o objetivo de investigar a conservação e relação entre posições e resíduos estatisticamente relacionados, extrapolando essa relação para identificar resíduos responsáveis pela oligomerização da estrutura, sítios de ligação ao tRNA e resíduos relacionados ao sítio ativo, estudos de conservação e acoplamento estatístico (SCA) foram utilizados, assim como realizado por Bachega e colaboradores em 2009 para a família de superóxido-dismutases<sup>61</sup>.

As análises foram realizadas em conjunto com o Dr. Lucas Bleicher, que desenvolveu os programas e realizou os procedimentos descrito a seguir com base nos trabalhos de Ranganathan<sup>62</sup>. Segundo a metodologia de Ranganathan, dois parâmetros são explorados:

- a conservação posicional ( $\Delta G^{\text{stat}}$ );
- o acoplamento estatístico entre duas posições ( $\Delta\Delta G^{\text{stat}}$ ).

Desse modo, resíduos altamente conservados em uma dada posição ao longo de um alinhamento são representados por altos valores de  $\Delta G^{\text{stat}}$ , enquanto  $\Delta\Delta G^{\text{stat}}$  quantifica o efeito da distribuição de aminoácidos em uma dada posição tendo-se um determinado resíduo em outra posição

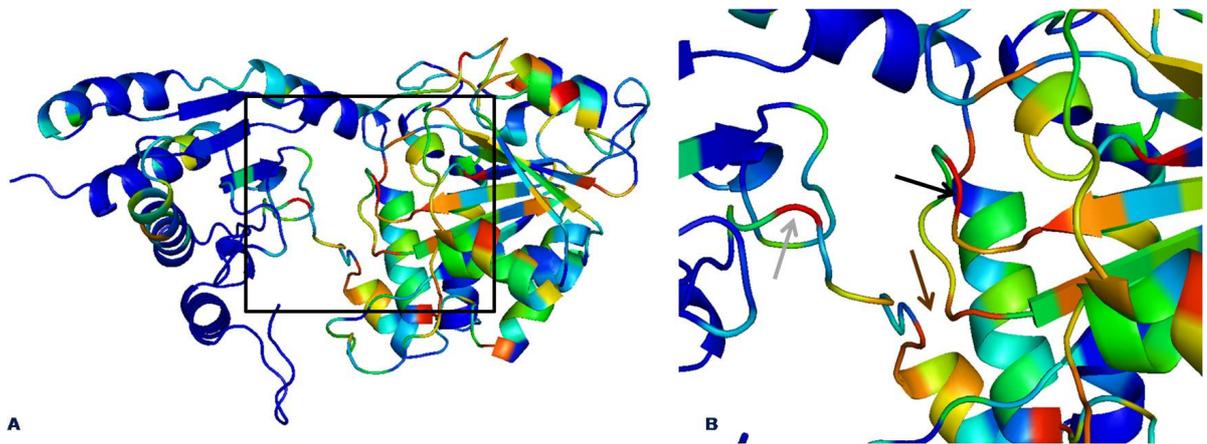
A principal diferença encontrada na abordagem de Bleicher<sup>63</sup> ocorre em  $\Delta\Delta G^{\text{stat}}$ , que além de correlacionar posições ao longo da sequência de interesse também especificam quais são os aminoácidos a ocuparem essas posições (correlação) e quais aminoácidos não podem ocupar tais posições (anti-correlação). Exemplificando,  $\Delta\Delta G_{ij=ALA}$  irá medir quanto à distribuição dos aminoácidos na posição *i* varia, quando uma Alanina ocupar a posição *j*, um procedimento conhecido por perturbação, e os aminoácidos cujas frequências aumentam ou diminuem devido a essa perturbação são registrados.

A análise da perturbação é muito mais informativa quando especificam quais resíduos estão envolvidos, e não apenas quais posições, visto que o conhecimento

dos aminoácidos possibilita a identificação das interações realizadas entre eles. Assim, pode-se gerar uma lista de correlação entre resíduos mensurando a variação de frequência de um resíduo numa posição na presença de outro resíduo em outra posição, e calculando a probabilidade dessa variação ter ocorrido ao acaso. Esse valor, no caso de uma correlação significativa, é bem baixo (ex.  $p < 10^{-15}$ ), assim pode se utilizar o valor de  $\log(P)$  para quantificar essa correlação (nesse exemplo, o valor seria 15). Pode-se então agrupar os resíduos em “comunidades”, isto é, conjuntos de resíduos que possuem forte correlação entre si e fraca correlação com os demais.

Para os estudos de conservação e acoplamento estatístico entre os resíduos encontrados na sequência da proteína SELA *E. coli*, um banco de dados (PF03841) contendo 654 sequências de proteínas distribuídas em 259 espécies entre procariotos (639), arqueobactérias (14) e um fragmento para *Schistosoma mansoni*, foi obtido pelo servidor PFAM (<http://pfam.sanger.ac.uk/>), que alinha as sequências, com base na similaridade de domínios depositados, com a sequência de referência. A partir dos códigos fornecidos pelo PFAM, as sequências completas foram obtidas no servidor Expasy (<http://www.expasy.ch/>) e posteriormente no programa Jalview<sup>64</sup> alinhadas pelo aplicativo MAFFT<sup>65</sup> e submetidas a análises de redundância de 90%, de modo a subtrair fragmentos redundantes, sequências com mutações pontuais ou repetidas, porém, com nomes ou códigos diferentes. Após a subtração das sequências redundantes, as 237 sequências resultantes foram utilizadas nos estudos de conservação e acoplamento estatístico (SCA).

A figura 3.7 ilustra o resultado de  $\Delta G^{\text{stat}}$ , isto é, a conservação dos resíduos de aminoácidos para a proteína SELA de *E. coli*, visualizados no modelo gerado pelo programa MODELLER<sup>57</sup>.



**Figura 3. 7 - A-** PDB da proteína SELA modelado pelo programa MODELLER submetido as análises de  $\Delta G^{\text{stat}}$  para conservação dos resíduos para proteína SELA de *E. coli*. Os valores de conservação aumentam do azul para o vermelho. **B-** Em destaque as setas indicam os resíduos Lisina-295 do sítio ativo ( ↓ ), o resíduo Asparagina-85 estruturalmente próxima ao sítio ativo ( ↓ ) e dois resíduos também conservados nessa região, Histidina-92 e Glutamina-301 ( ↓ ).

É possível observar resíduos mais conservados no interior da sequência protéica visualizada a direita da figura 3.7 **A**, em contrapartida ao N e C terminal quase completamente sem conservação, lado esquerdo em azul da mesma figura. Também é observado em destaque (figura 3.7 **B**) o resíduo Lisina (K-295) ligante ao cofator PLP (setas preta), o resíduo Asparagina (N-85) (seta cinza) ambos com alto grau de conservação e os resíduos Histidina (H-92) e Glutamina (Q-301), com menor grau de conservação, em regiões flexíveis próximas ao sítio ativo (seta marrom).

Quando famílias de proteínas assumem estados oligoméricos distintos ou mesmo ligam-se à moléculas distintas é esperado que essas interações não sejam devido a apenas um ou outro aminoácido, mas um grupo de aminoácidos possivelmente relacionados estruturalmente entre si<sup>61</sup>. A proteína SELA encontrada amplamente em procaríotos, possui organização estrutural homodecamérica e liga-se ao tRNA carregando o resíduo serina (Ser-tRNA<sup>sec</sup>). Os estudos de acoplamento estatísticos visam identificar resíduos relacionados com os contatos entre os monômeros bem como em contato com o tRNA, visto que o contato com o resíduo de serina é conhecidamente realizado com cofator PLP ligado a Lisina-295<sup>40</sup>.

Dessa forma, inicialmente as análises de correlação entre os resíduos da proteína SELA foram realizados considerando apenas correlações com valores menores de  $10^{-10}$  de probabilidade de ocorrerem ao acaso. Devido à grande

quantidade de dados gerados, restringiu-se para valores menores de  $10^{-20}$ . O resultado foram sete grupos com resíduos relacionados entre si como visualizados nas tabelas 3.2 a 3.8 . Para entender os dados expostos nas seguintes tabelas é preciso observar que:

- a primeira linha refere-se a resíduos e posições encontrados na sequência da proteína SELA de *E. coli* e que apresentaram correlação;

- a primeira coluna refere-se a resíduos e posições correlacionadas aos resíduos da primeira linha;

- a segunda coluna refere-se à porcentagem dos resíduos encontrados na primeira coluna ao longo do alinhamento realizados entre todas as sequências em estudo, isto é, que ocorrem naturalmente na população de sequências utilizadas nesse estudo;

- os valores ao longo da tabela referem-se às porcentagens de acoplamento entre os resíduos da primeira linha com os da primeira coluna. Os valores apresentados em verde referem-se às porcentagens superiores a 15% comparados aos valores encontrados no alinhamento das sequências (segunda coluna). Assim, na tabela 3.2, por exemplo, observa-se que a frequência de Glicinas na posição 262 (numeração referente à SELA de *E. coli*) é de 82,8%, e essa frequência aumenta para 98,5% na presença de uma Fenilalanina na posição 235.

De modo a facilitar a identificação e visualização dos resíduos que possuem correlações, a tabela 3.9 reúne os resíduos dos sete grupos destacando o respectivo resíduo, posição e classificação quanto à característica de sua cadeia lateral diferenciado por cores. A figura 3.8 combina os resíduos das tabelas com a estrutura da proteína SELA de *E. coli* modelada pelo programa MODELLER. Inicialmente, o modelo com o gradiente de conservação de resíduos seguido por cada grupo de correlação sobreposto separadamente ao modelo com a respectiva coloração de seus resíduos segundo a tabela 3.9.

**Tabela 3. 2** - Grupo 1 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	H92	N94	R127	R130	I180	R185	T203	S228	N229	F235	G262	G397	P400
H92	81,0	-	98,0	98,0	96,9	97,9	97,5	97,0	98,0	98,0	97,9	97,5	97,5	95,5
N94	82,8	99,0	-	98,5	97,9	99,5	98,0	97,5	99,0	99,5	99,5	98,0	98,5	97,5
R127	83,6	100,0	99,5	-	99,0	99,5	99,5	99,0	99,5	99,5	100,0	99,5	99,5	97,0
R130	81,5	96,4	96,4	96,5	-	96,4	96,0	96,5	95,9	96,4	96,9	95,9	95,9	95,0
I180	80,7	96,4	97,0	96,0	95,4	-	96,0	94,9	96,4	97,0	96,9	95,4	95,9	94,5
R185	84,0	100,0	99,5	100,0	99,0	100,0	-	99,0	99,5	99,5	100,0	99,5	99,5	97,0
T203	83,2	98,5	98,0	98,5	98,5	97,9	98,0	-	98,0	98,0	98,5	98,0	98,0	96,0
S228	82,8	99,0	99,0	98,5	97,4	99,0	98,0	97,5	-	99,5	99,0	98,5	99,0	97,0
N229	82,8	99,0	99,5	98,5	97,9	99,5	98,0	97,5	99,5	-	99,5	98,5	99,0	97,5
F235	81,9	97,9	98,5	98,0	97,4	98,4	97,5	97,0	98,0	98,5	-	97,5	98,0	96,0
G262	82,8	98,5	98,0	98,5	97,4	97,9	98,0	97,5	98,5	98,5	98,5	-	98,5	96,0
G397	82,8	98,5	98,5	98,5	97,4	98,4	98,0	97,5	99,0	99,0	99,0	98,5	-	96,5
P400	83,6	97,4	98,5	97,0	97,4	97,9	96,5	96,5	98,0	98,5	97,9	97,0	97,5	-

**Tabela 3. 3** - Grupo 2 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	R97	E119	N152	S173	F184	D210	H226	G260	D294	R325	G395
R97	85,3	-	98,5	99,5	100,0	100,0	99,5	99,5	99,5	100,0	100,0	99,0
E119	86,1	99,5	-	99,5	100,0	100,0	99,5	99,0	99,5	100,0	100,0	99,0
N152	85,3	99,5	98,5	-	100,0	100,0	99,5	99,0	99,5	100,0	100,0	99,0
S173	83,6	98,0	97,1	98,0	-	98,5	98,0	97,5	98,0	98,5	98,5	97,5
F184	84,5	99,0	98,0	99,0	99,5	-	99,0	98,5	99,0	99,5	99,5	98,5
D210	85,3	99,5	98,5	99,5	100,0	100,0	-	99,0	100,0	100,0	100,0	99,0
H226	85,7	100,0	98,5	99,5	100,0	100,0	99,5	-	99,5	100,0	100,0	99,0
G260	85,3	99,5	98,5	99,5	100,0	100,0	100,0	99,0	-	100,0	100,0	99,0
D294	84,9	99,5	98,5	99,5	100,0	100,0	99,5	99,0	99,5	-	100,0	99,0
R325	84,9	99,5	98,5	99,5	100,0	100,0	99,5	99,0	99,5	100,0	-	99,0
G395	84,0	97,5	96,6	97,5	98,0	98,0	97,5	97,1	97,5	98,0	98,0	-

**Tabela 3. 4** - Grupo 3 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	L86	D131	A183	K321	L347	L351	S396	L421
L86	24,4	-	94,9	90,0	100,0	93,2	90,0	98,3	84,6
D131	24,8	96,6	-	90,0	98,2	91,5	88,3	96,6	83,1
A183	25,2	93,1	91,5	-	94,6	88,1	86,7	93,1	83,1
K321	23,5	96,6	93,2	88,3	-	91,5	88,3	96,6	84,7
L347	24,8	94,8	91,5	86,7	96,4	-	90,0	94,8	83,1
L351	25,2	93,1	89,8	86,7	94,6	91,5	-	93,1	81,4
S396	24,4	98,3	94,9	90,0	100,0	93,2	90,0	-	84,7
L421	24,8	87,9	83,1	81,7	89,3	83,1	80,0	86,2	-

**Tabela 3. 5** - Grupo 4 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	L118	G128	N153	V157	L161	G182	M190	F291
L118	77,7	-	96,0	95,0	97,6	97,2	98,3	98,7	97,2
G128	72,7	89,7	-	89,0	92,7	92,8	92,2	92,3	89,3
N153	76,1	93,0	93,1	-	94,6	93,3	95,5	97,4	92,1
V157	69,3	87,0	88,4	87,8	-	89,4	91,1	92,9	87,6
L161	75,6	94,6	96,5	92,8	97,6	-	97,2	98,1	94,4
G182	75,2	95,1	95,4	94,5	98,8	96,7	-	98,1	95,5
M190	65,1	82,7	82,7	83,4	87,3	84,4	84,9	-	81,4
F291	74,4	93,0	94,4	90,1	93,9	92,8	94,4	92,9	-

**Tabela 3. 6** - Grupo 5 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	H207	I239
H207	60,5	-	85,9
I239	62,6	88,9	-

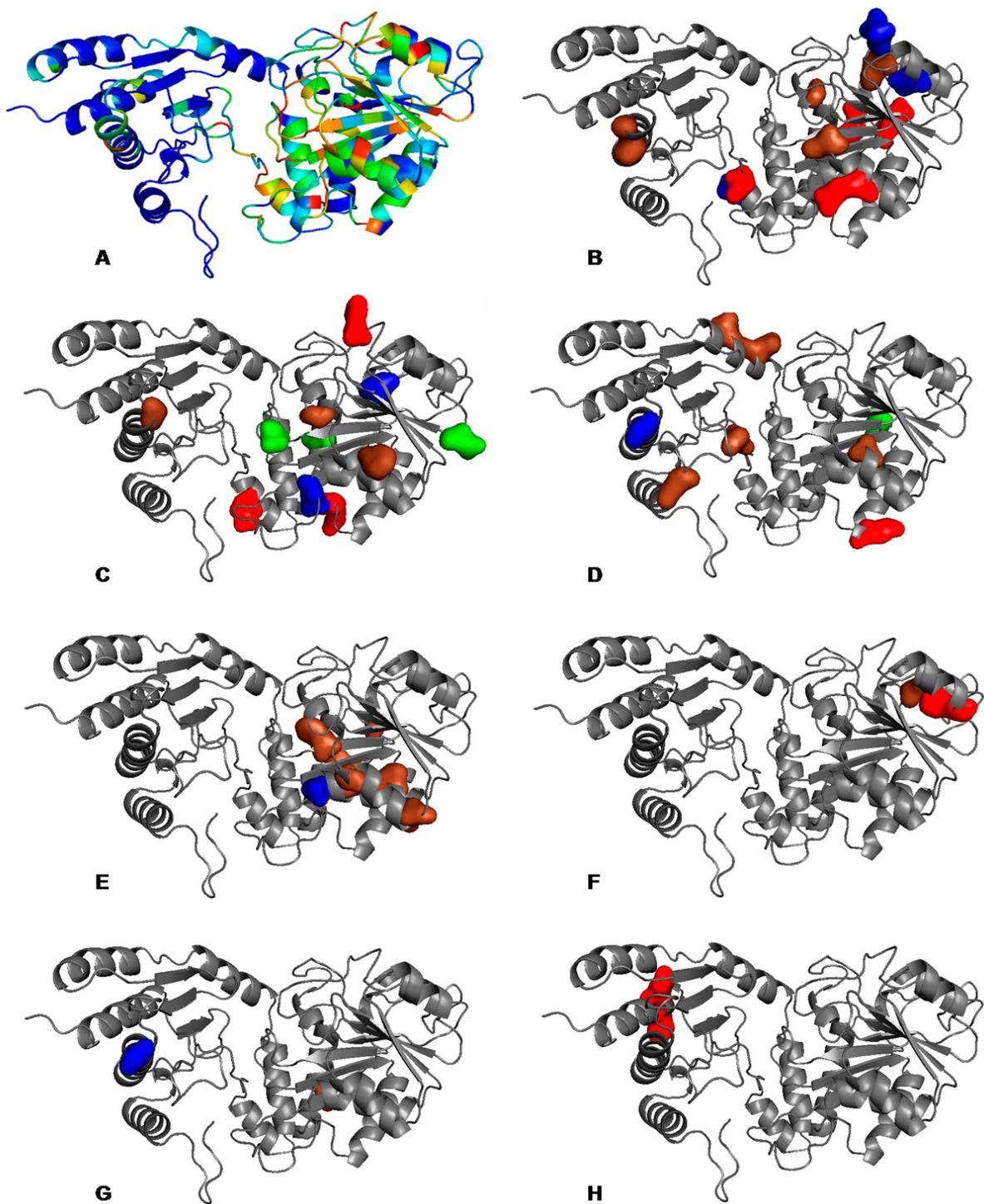
**Tabela 3. 7** - Grupo 6 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	P115	S396
P115	59,2	-	85,9
S396	62,6	90,8	-

**Tabela 3. 8** - Grupo 7 de correlação entre os resíduos da proteína SELA de *E. coli*. Em verde as correlações que apresentaram valores superiores a 15% comparados aos valores obtidos no alinhamento entre todas as sequências.

POS	ALL	L447	L449
L447	77,3	-	96,3
L449	79,0	98,4	-





**Figura 3. 8** - Análise de correlação entre os resíduos da proteína SELA de *E. coli*. **A**- modelo gerado pelo programa MODELLER com gradiente de conservação de resíduos; **B, C, D, E, F, G** e **H**- sobreposição à estrutura SELA modelada em cinza dos resíduos que possuem correlação resultante respectivamente aos grupos 1, 2, 3, 4, 5, 6, 7 agrupados na tabela 3.9.

Dos grupos de resíduos correlacionados acima, a maior parte se compõe de posições que já tem ocupação alta (acima de 80%) para um dado aminoácido. A exceção é para o grupo de número 3, em que os sítios de aminoácidos estão presentes em menos de 30% do total de sequências, mas apresentam correlação tão forte que a presença de apenas um deles leva a ocorrência de mais de 80% para qualquer outro. Esse tipo de comportamento sugere que em algum momento da evolução favoreceu-se a conservação simultânea desses destes resíduos, o que poderia estar relacionado a alguma função biológica. Curiosamente a SELA estudada nesse trabalho pertence a esse grupo, apresentando simultaneamente todos os oito aminoácidos encontrados. Esse grupo também é o único que apresenta anti-correlação com outro grupo dentre os encontrados. Pela tabela 3.10, podemos observar que a presença de resíduos no grupo 3 leva a diminuição de frequências dos resíduos na comunidade de número 6 e vice-versa, sendo a probabilidade dessa diminuição em média menor que  $10^{-10}$ .

**Tabela 3. 10** - Análise de correlação/anti-correlação para acoplamento entre procariotos e arqueobactérias. Os números correspondem à média dos logaritmos dos p-valor observado para correlações entre membros de uma comunidade em relação aos membros da outra. Sinais positivos são atribuídos quando há correlação, enquanto sinais negativos são usados para anti-correlação, isto é, a presença de um resíduo diminui a frequência de outro.

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
Grupo 1	10.059172	11.720280	0.951923	4.644231	2.615385	0.461538	6.115385
Grupo 2	8.692307	10.520661	0.579545	4.056818	2.227273	0.500000	4.545455
Grupo 3	3.769231	3.420455	22.531250	5.718750	6.000000	-16.875000	4.687500
Grupo 4	8.634615	9.363636	1.468750	8.781250	2.312500	0.125000	5.062500
Grupo 5	7.807693	7.181818	2.437500	3.437500	5.250000	-0.250000	8.000000
Grupo 6	1.230769	0.954545	-12.250000	0.187500	0.000000	6.000000	0.500000
Grupo 7	8.769231	7.954545	1.000000	3.812500	3.750000	0.500000	6.500000

### 3.5 Ensaios estruturais da proteína SELA e SELA-tRNA<sup>sec</sup> por SAXS

A proteína Selenocisteína Sintase (SELA) e a mesma incubada com tRNA<sup>sec</sup> foram estudados pela técnica de espalhamento de raios-X a baixos ângulos (SAXS) objetivando-se a determinação de sua estrutura tridimensional. Esses estudos foram realizados em colaboração com a Profa. Dra. Íris L. Torriane e Dr. Cristiano L. P. Oliveira no Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas – SP – Brasil.

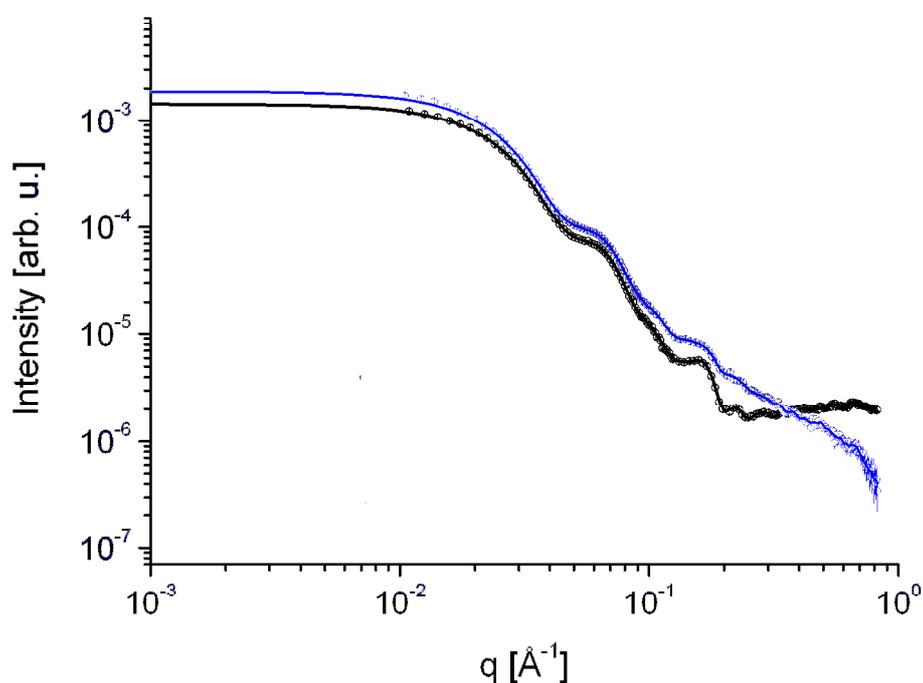
Todas as medidas de SAXS foram realizadas no LNLS, inicialmente utilizando a linha SAXS–1 para medidas da proteína SELA e posteriormente com a implantação da linha SAXS–2 para medidas tanto da proteína SELA como SELA-tRNA<sup>sec</sup>. As medidas utilizaram 15 µM da proteína SELA incubados com 75, 162 e 232 µM do tRNA<sup>sec</sup>, obtendo-se as respectivas proporções molares de 1 molécula de proteína para 5, 10 e 15 moléculas de tRNA. O comprimento de onda do raio incidente utilizado nas diversas experimentações foi  $\lambda=1.488\text{\AA}$ , as distâncias amostra – detector entre 407 e 1620mm. Para aquisição de dados utilizou-se um detector 1D sensível a posição (DSP), na linha SAXS–1 e um detector bidimensional (CCD), na linha SAXS–2.

Desse modo, realizou-se a subtração entre as curvas do espalhamento da solução protéica pela curva de espalhamento da solução tampão, resultando na curva final de espalhamento para a proteína em estudo. A partir da curva de espalhamento da proteína, o programa autoRg provido por EMBL (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/>) foi utilizado como guia para avaliação da curva de espalhamento. Esse programa informa os pontos utilizados e porcentagem de qualidade para o gráfico de Guinier<sup>66</sup> e cálculo do raio de giro – R<sub>g</sub> e I(0) para a curva de espalhamento I por q.

O método de transformada inversa de Fourier implementado no programa GNOM<sup>67, 68</sup>, foi utilizado para obtenção da função de distribuição de distâncias – p(r), diâmetro máximo da molécula – D<sub>max</sub> e o raio de giro – R<sub>g</sub>, também obtido anteriormente pela equação de Guinier. Para as medidas realizadas na linha SAXS-

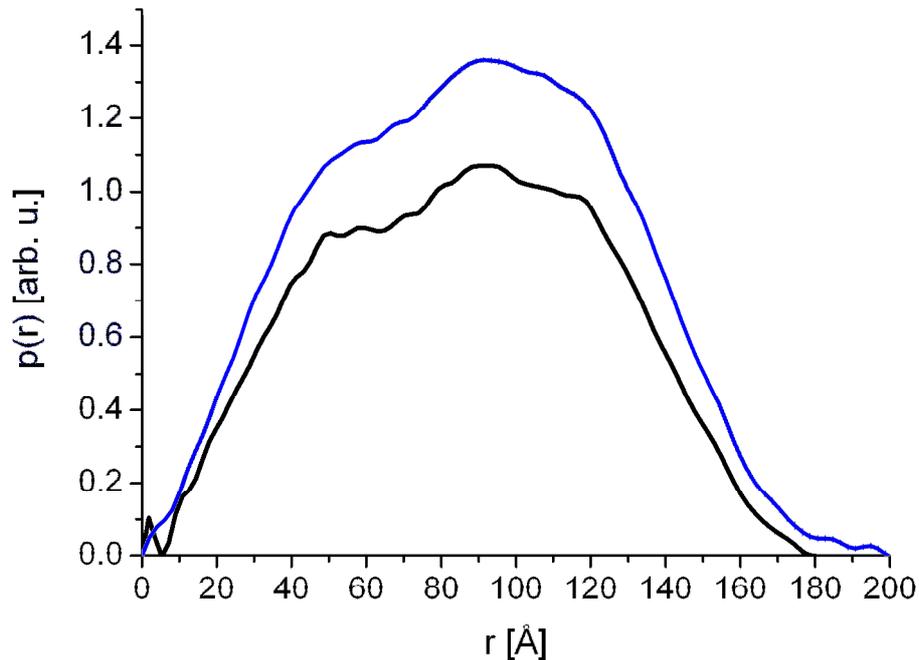
1, o efeito de deformação da curva (*smearing*), devido à altura e largura da janela de entrada do detector e aos parâmetros do feixe (feixe não pontual) foram adequadamente corrigidos no programa GNOM.

A curva de espalhamento medida experimentalmente e sua extrapolação de  $q$  igual a zero e normalizado para satisfazer a condição  $I(0)$  igual a 1, pelo programa GNOM, podem ser visualizados na figura 3.9 tanto para a proteína SELA como SELA-tRNA.



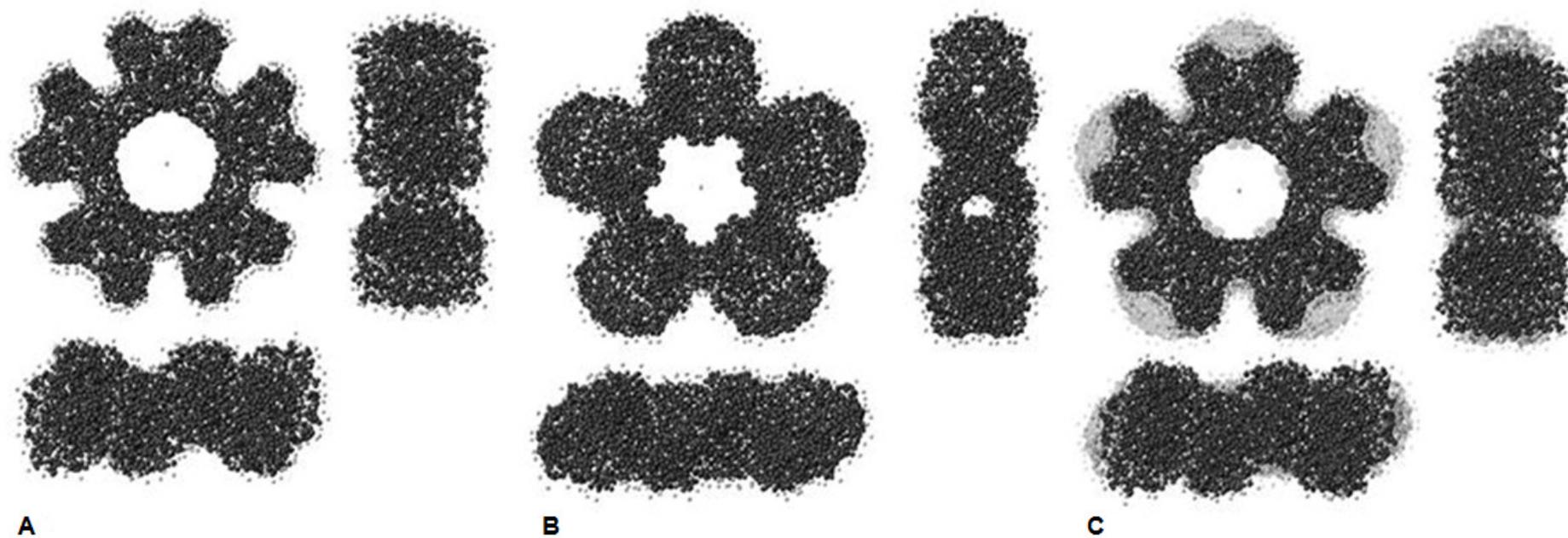
**Figura 3. 9** - Curvas de espalhamento obtidas experimentalmente (pontos) e respectivo ajuste pelo GNOM extrapolando a  $q=0$  (linha cheia). Em preto, proteína SELA e em azul, proteína SELA-tRNA.

Os parâmetros estruturais derivados da curva de espalhamento, como valores de dimensão máxima ( $D_{\max}$ ), raio de giro ( $R_g$ ) e massa molecular ( $MM$ ), bem como dados obtidos pela técnica de Microscopia Eletrônica de Transmissão (TEM)<sup>46</sup> e cálculos teóricos para massa molecular tanto para a proteína SELA como SELA-tRNA são visualizados na tabela 3.11. A figura 3.10 mostra a curva de distribuição de distância gerada pelo programa GNOM.

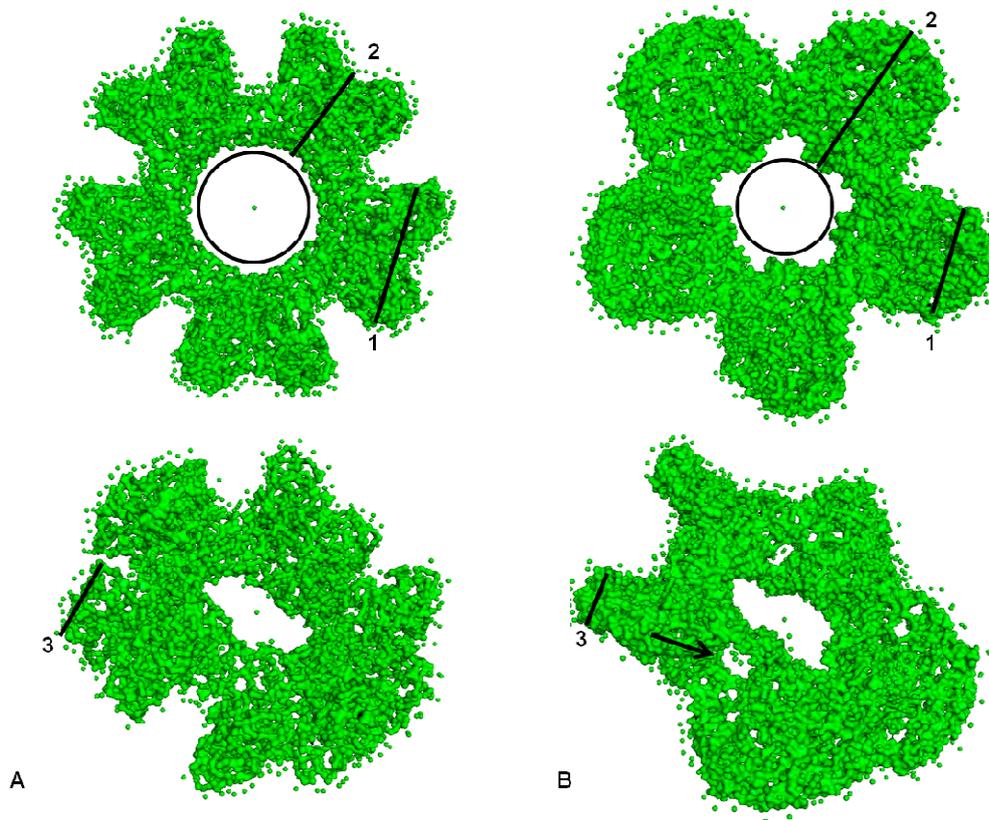


**Figura 3. 10** - Função de distribuição de distâncias a partir da curvas de espalhamento obtidas experimentalmente. Em preto proteína SELA e em azul proteína SELA-tRNA.

Os modelos de baixa resolução foram gerados utilizando procedimentos *ab initio*, implementados no programa GASBOR<sup>69</sup> e a partir de parâmetros iniciais tais como o arquivo de saída do programa GNOM, número de aminoácidos por monômero (463 aa) e a simetria da proteína (P52) obtido através de dados de Microscopia Eletrônica de Transmissão (TEM)<sup>46</sup>. Para verificar a validade do modelo, realizou-se 10 construções independentes, resultando em envelopes similares, validando o melhor modelo pelo programa DAMAVER ([http://www.embl-hamburg.de/ExternalInfo/ Research/Sax/](http://www.embl-hamburg.de/ExternalInfo/Research/Sax/)). Os modelos finais para a proteína SELA e SELA-tRNA podem ser visualizados na figura 3.11.



**Figura 3. 11** - Modelo *ab initio*, gerados pelo programa GASBOR. **A**- proteína Selenocisteína Sintase (SELA); **B**- SELA – tRNA; **C**- sobreposição das duas estruturas.



**Figura 3. 12** - Modelo *ab initio*, gerados pelo programa GASBOR e imagens tratadas pelo programa PYMOL. **A**- proteína Selenocisteína Sintase (SELA); **B**- SELA – tRNA. As linhas negras e respectivos números indicam as medidas da subunidade dimérica: 1- comprimento lateral; 2- comprimento em direção ao centro; 3- espessura e o círculo-diâmetro interno do orifício central. Valores encontrados na tabela 3.11. A seta negra indica um dos cinco orifícios encontrado apenas na estrutura SELA-tRNA.

**Tabela 3. 11** - Parâmetros estruturais para a proteína SELA e SELA-tRNA, obtidos por Microscopia Eletrônica de Transmissão (TEM) e Espalhamento de Raios-X a Baixos Ângulos (SAXS).

	$D_{max}$ (Å)	Orifício Central (Å)	Subunidade Dimérica (Å)			Rg (Å)	MM (kDa)
			Lateral	Em direção ao centro	Espessura		
<b>Proteína SELA</b>							
<b>TEM</b>	190	40	85	70	50 a 60	-	493 +/- 20
<b>SAXS</b>	180	60	88	40	61	66,5 +/- 0,1	552 +/- 28
<b>Teórico</b>	-	-	-	-	-	-	509,322
<b>Proteína SELA-tRNA (1:5)</b>							
<b>SAXS</b>	203	55	81	78	31	68,0 +/- 0,1	630 +/- 30
<b>Teórico</b>	-	-	-	-	-	-	664,322

Dados obtidos por TEM<sup>46</sup>; Dados teóricos obtidos a partir da literatura<sup>24, 40, 46</sup>; Dados de SAXS obtidos experimentalmente.

## 3.6 Ensaios estruturais da proteína SELA e SELA-tRNA<sup>sec</sup> por difração de raio-X

### 3.6.1 Proteína Selenocisteína Sintase

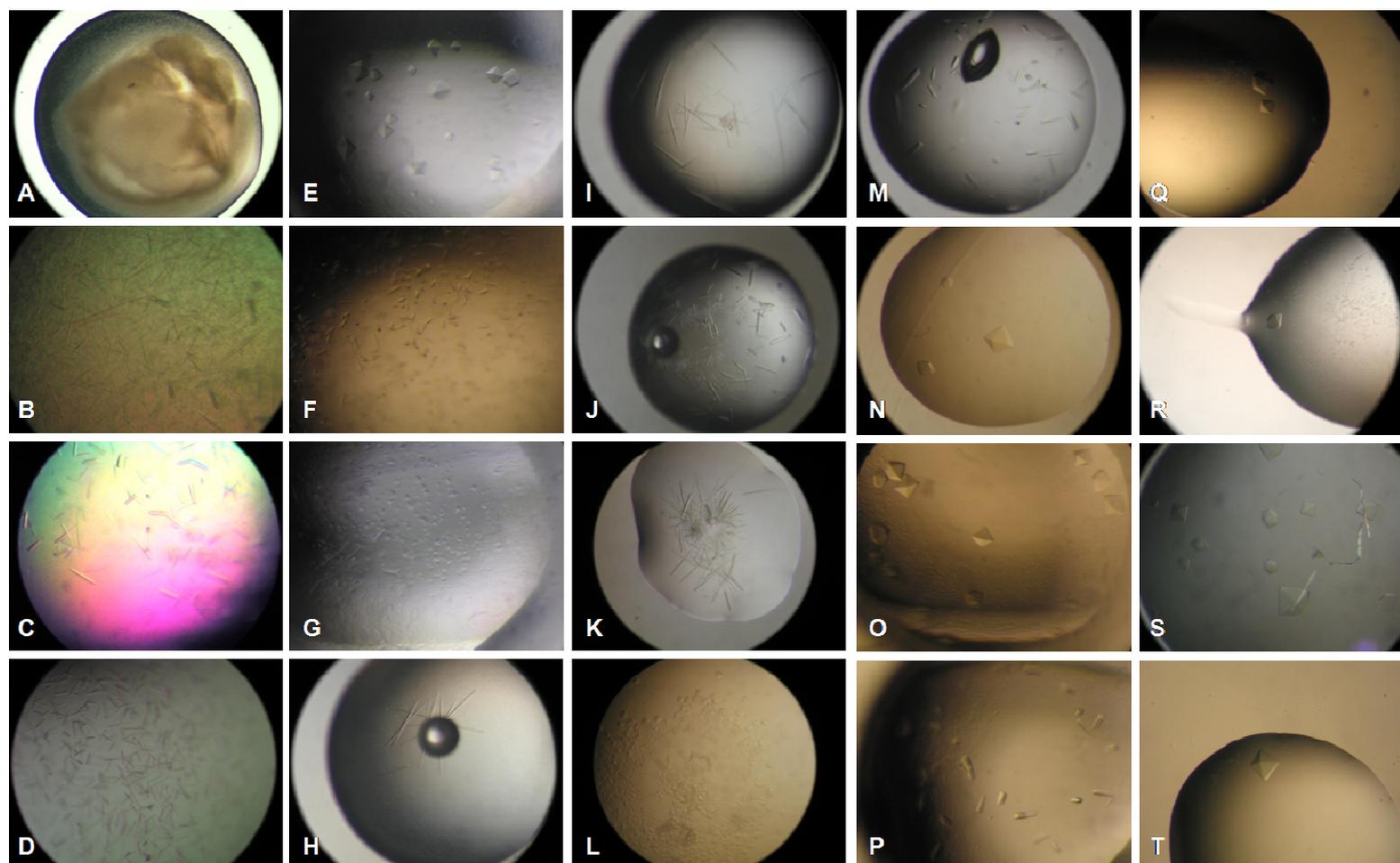
Os estudos de cristalografia foram realizados em colaboração com o Dr. Humberto d’Muniz Pereira e Prof. Dr. Eduardo Horjales Reboredo do grupo de cristalografia no Instituto de Física de São Carlos.

A abordagem utilizada para a maior parte dos ensaios de cristalização tanto da proteína SELA como da mesma ligada ao tRNA<sup>sec</sup> foi a técnica de difusão de vapor por gota suspensa (*hanging drop vapor diffusion*). Desse modo, 500 µL de solução de cristalização foram adicionados ao poço de uma placa de cristalização, 2 µL dessa solução foram adicionados a 2 µL de proteína em uma lamínula que rapidamente foi invertida sobre o poço previamente contornado com graxa de silicone, resultando em um sistema fechado. Posteriormente a aquisição do robô de cristalização “Honeybee 939” (Genomic Solution) pelo Instituto de Física de São Carlos (IFSC), passou-se a realizar triagens de cristalização automatizadas pela técnica de difusão de vapor por gota depositada (*sitting drop vapor diffusion*). As placas de cristalização utilizadas pelo robô necessitam de menores quantidades de solução de poço (80 µL) e proteína (1 µL). Após o preparo das placas de cristalização as mesmas foram deixadas em repouso em locais com temperaturas controladas a 4 e 18° C, livres de vibrações, para que as proteínas pudessem cristalizar.

Os fatoriais comerciais utilizadas nas triagens de cristalização são provenientes de Qiagen-Nextal e Hampton Research (tabela 3.12).

**Tabela 3. 12** - Fatoriais comerciais de cristalização utilizados para triagens de cristalização da proteína SELA e SELA-tRNA<sup>sec</sup>.

<b>Fabricante</b>	<b>Código</b>	<b>Fatorial de Cristalização</b>
Hampton Research	HR2 – 110	Crystal Screen
	HR2 – 112	Crystal Screen 2
	HR2 – 126	PEG/Ion Screen
	HR2 – 130	Crystal Screen HT
	HR2 – 136	SaltRx
	HR2 – 211	Grid Screen Ammonium Sulfate
	HR2 – 213	Grid Screen PEG 6000
	HR2 – 215	Grid Screen 2-Methyl-2-4-pentaneidiol
	HR2 – 219	Grid Screen Sodium Chloride
	HR2 – 221	Quik Screen Sodium/Potassium Phosphate
Quiagen-Nextal	130701	Classics Suite
	130704	PEGs Suite
	130716	PEGs II Suite
	130719	Nucleix Suite



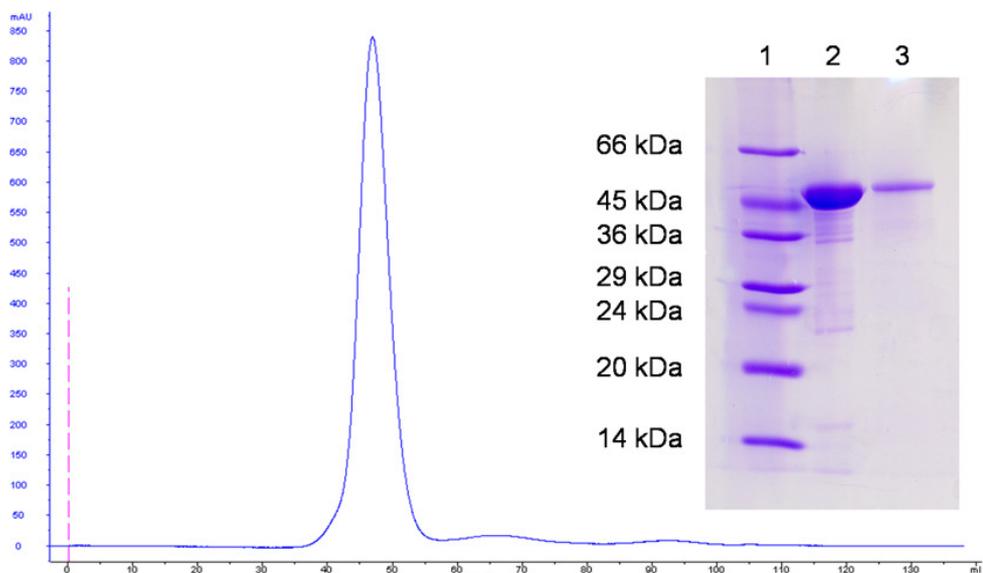
**Figura 3. 13** - Foto dos cristais obtidos ao longo das triagens e refinamentos das condições de cristalização para proteína Selenocisteína Sintase de *E. coli*. Em **A**- Cristal Screen 2 – 21: 2.0 M Sodium Chloride, 0.1 M MES pH 6.5, 0.2 M Na/K Phosphate; Em **B**- Sodium Chloride – B3: 0.1 M MES pH 6.0, 2.0 M Sodium Chloride; Em **C**- refinamento dos cristais A e B - 26% MPD, 0.1M Mes pH 6.2; Em **D**- refinamento dos cristais A e B - 29% MPD, 0.1M Mes pH 6.4; Em **E** a **L**- condições descritas na tabela 3.13 (página 73); Em **M** a **P**- condições descritas na tabela 3.14 (página 74); Em **Q**- cristal condição N formado em solução crio de 22% de glicerol; Em **R**- cristal condição N formado após “microseeding”; Em **S**- cristal condição N formado em solução crio 20% de etilenoglicol; Em **T**- refinamento cristal N crescido após digestão com Tripsina em 1.6 M do agente precipitante. Cristais S e T apresentaram melhores padrões de difração.

As triagens de cristalização para a proteína SELA resultaram em inúmeros cristais de sais e algumas condições em que micro-cristais, em forma de agulhas com grande quantidade de nucleações e precipitado, precisaram ser refinadas para testes posteriores (figura 3.13 – **A** e **B**). O refinamento deu-se com variações na concentração da proteína entre 10 a 17  $\mu\text{M}$  (5 a 9 mg/mL), concentração do agente precipitante, pH e adição de sais como cloreto de sódio, cloreto de magnésio, sulfato de amônio e acetato de sódio. As melhorias na forma e tamanho dos cristais foram obtidas em tampão 0,1 M de MES entre os pHs 6,0 e 6,5 com variações entre 20 e 29% de 2-Methyl-2,4-pentaneidiol (MPD) a 18° C (figura 3.13 – **C** e **D**).

Após a melhoria dos cristais, os mesmos apresentaram padrão de difração de sal, quando submetido ao feixe de raios-X na linha MX2 do Laboratório Nacional de Luz Síncrotron (LNLS).

Tendo em vista a não formação de cristais de proteína, apesar de ampla triagem em fatoriais de cristalização comerciais, supôs-se que problemas como contaminantes ou presença de mais de um estado oligomérico da proteína impediam sua cristalização. Desse modo, uma etapa de gel filtração (coluna superdex 200 – GE Healthcare - 1.6 x 60.0 cm), sugerida pela doutoranda Lívia Regina Manzine (IFSC – USP), foi adicionada como último passo ao protocolo de purificação anteriormente estabelecido<sup>51</sup>.

A coluna de gel filtração Superdex 200 é utilizada para separar proteínas com massas moleculares entre 10 e 600 kDa. Dessa forma, a proteína SELA homodecamérica (510 kDa) é esperada como primeira fração a ser eluída da coluna. Com maior pureza e um único estado oligomérico (figura 3.14.) repetiu-se os fatoriais de cristalização (tabela 3.12) com a proteína SELA e a mesma ligada ao tRNA<sup>sec</sup>.



**Figura 3. 14** - Cromatografia da proteína SELA após gel filtração (Superdex 200) e análise em SDS-PAGE: coluna **1**- padrão de massa molecular; coluna **2**- proteína SELA antes da coluna de gel filtração; coluna **3**- proteína SELA após coluna de gel filtração.

As triagens de cristalização após a passagem da proteína SELA pela coluna de gel filtração resultaram novamente em inúmeros cristais de sais, alguns cristais cuja forma precisava ser melhorada e cristais que não apresentaram padrões de difração. Com exceção das condições resultantes de cristais de sais, os demais cristais são listados na tabela 3.13 e ilustrados na figura 3.13 – **E - L**.

**Tabela 3. 13** - Micro-cristais e cristais da proteína SELA obtidos em triagens de cristalização após a adição da coluna de gel filtração Superdex 200 no protocolo de purificação.

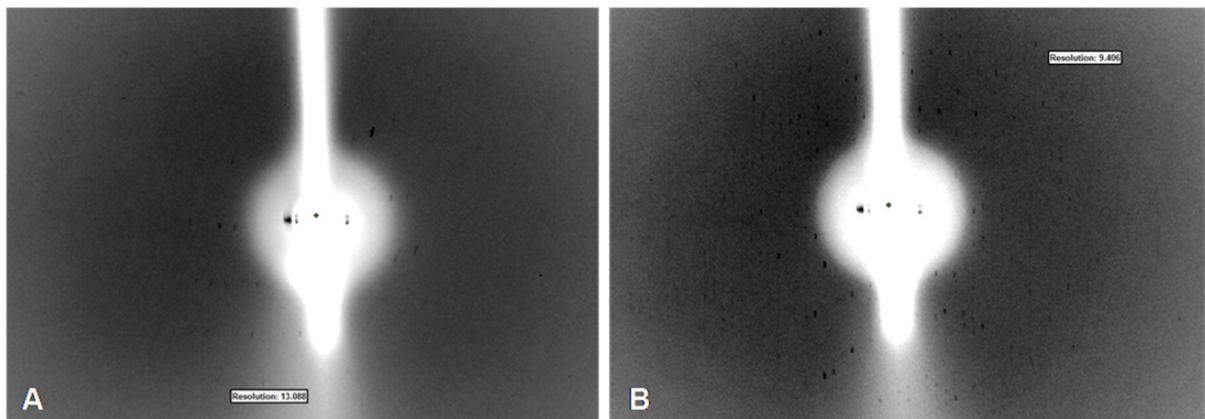
Proteína	Condição de Cristalização	Classificação do Cristal
Selenocisteína Sintase	<b>Hampton Research</b>	
	<b>Fatorial Salt Rx</b>	
	<b>E- Sal</b> 1.0 M Magnesium sulfate hydrate Tampão 0.1 M BIS-TRIS propane pH 7.0	Cristal não difratou
	<b>F- Sal</b> 3.2 M Sodium chloride Tampão 0.1 M BIS-TRIS propane pH 7.0	Cristal não difratou
	<b>G- Sal</b> 0.6 M Potassium sodium tartrate tetrahydrate Tampão 0.1 M BIS-TRIS propane pH 7.0	Cristal não difratou
	<b>Crystal Screen Lite</b>	
	<b>H- Sal</b> 0.2 M Magnesium Chloride hexahydrate Tampão 0.1 M HEPES - Na pH 7.5 Precipitante 15% v/v iso-Propanol	Melhorar cristais
	<b>I- Sal</b> 0.2 M tri-Sodium Citrate dihydrate Tampão 0.1 M HEPES - Na pH 7.5 Precipitante 10% v/v iso-Propanol	Melhorar cristais
	<b>J- Sal</b> None Tampão None Precipitante 2.0 M Sodium Formate	Cristal não difratou
	<b>Qiagen-Nestal</b>	
	<b>Classics Suite</b>	
	<b>κ- Sal</b> None Tampão 0.1 M HEPES sodium salt pH 7.5 Precipitante 0.8 M K/Na tartrate	Melhorar cristais
	<b>L- Sal</b> 0.1 M Sodium phosphate, 0.1 M Potassium phosphate Tampão 0.1 M MES pH 6.5 Precipitante 2.0 M Sodium chloride	Cristal não difratou

Quatro cristais de proteína com padrões de difração a baixa resolução cristalográfica foram obtidos, sendo um para a proteína SELA e três para proteína SELA-tRNA<sup>sec</sup> (tabela 3.14 e figura 3.13 – **M - P**).

**Tabela 3. 14** - Cristais da proteína SELA e SELA-tRNA<sup>sec</sup> obtidos em triagens de cristalização após a adição da coluna de gel filtração Superdex 200 no protocolo de purificação.

<b>Proteína</b>	<b>Condição de Cristalização</b>	<b>Classificação do Cristal</b>
Selenocisteína Sintase	<b>Qiagen-Nestal</b> <b>PEGsSuite</b>  M- Sal None Tampão 0.1 M Tris.HCl pH 8.5 Precipitante 30 %(v/v) PEG 400	Cristal de Proteína      (difração > 13Å)
Selenocisteína Sintase – tRNA <sup>sec</sup>	<b>Qiagen-Nestal</b> <b>Classic Suite</b>  N- Sal 0.1 M Sodium phosphate, 0.1 M Potassium phosphate Tampão 0.1 M MES pH 6.5 Precipitante 2.0 M Sodium chloride	Cristal de Proteína      (difração > 9.4Å)
	<b>Qiagen-Nestal</b> <b>Nucleix Suite</b>  o- Tampão 0.05 M Sodium cacodylate pH 6.5 Aditivo 0.08 M Magnesium acetate Precipitante 15 %(v/v) PEG 400	Cristal de Proteína      (difração > 20Å)
	P- Tampão 0.05 M HEPES Sodium salt pH 7.0 Aditivo 0.1 M Potassium chloride, 0.01 M Calcium chloride Precipitante 10 %(v/v) PEG 400	Cristal de Proteína      (difração > 20Å)

Apesar de inúmeras tentativas para conseguir um bom padrão de difração das condições descritas na tabela 3.13 acima, nenhuma condição de refinamento obteve sucesso. Entretanto, nas condições que apresentaram difração, tabela 3.14, foi observado que apesar dos padrões apresentarem baixas resoluções cristalográficas com poucos “spots” a proteína ligada ao tRNA atingia maiores resoluções e quantidades de “spots” quando comparada a proteína sem o tRNA ligante (figura 3.15). Também foram observados cristais com forma semelhantes na mesma condição de cristalização, entretanto, apresentando difração apenas na presença do tRNA (tabela 3.13 – **L** e tabela 3.14 – **N** e figura 3.13 – **L** e **N**).



**Figura 3.15** - Padrão de difração para a proteína SELA e SELA-tRNA<sup>sec</sup>. Em **A**- proteína SELA, em **B**- proteína SELA-tRNA<sup>sec</sup>. Ambas cristalizadas conforme descrito na tabela 3.14 – **M**, **N** e visualizados na figura 3.13 – **M**, **N**.

Os cristais tabela 3.14 – **O** e **P** (figura 3.13 – **O** e **P**) apresentaram problemas de difração pobre e de reprodutibilidade dos cristais, inviabilizando o refinamento para essas condições. Os cristais tabela 3.13 – **L** e tabela 3.14 – **M** e **N** (figura 3.13 – **L** – **N**) passaram por inúmeras condições de refinamento descritas a seguir.

Inicialmente foi observada uma piora do padrão de difração até ausência total de difração após 19 dias de vida do cristal. Devido à rápida formação dos cristais (aproximadamente 15 horas), estabeleceu seu congelamento após dois dias da preparação da gota de cristalização. Para o congelamento foram preparadas soluções crio - protetoras (preparo da solução mãe de cristalização com glicerol ou etilenoglicol 20% - concentração final). Desse modo, o cristal era transferido para

uma gota de solução crio - protetora por alguns minutos, recolocado no “loop” e congelado em nitrogênio líquido até o momento da exposição aos raios-X.

É sabido que quando a proteína cristaliza muito rápida resultando em cristais com padrões de difração ruins ou mesmo ausentes, pode significar problemas na organização interna do cristal, e uma alternativa para melhorá-los é diminuir a velocidade de sua formação<sup>70-72</sup>. Cristais tanto da proteína SELA como SELA-tRNA<sup>sec</sup> foram conseguidos quando as mesmas foram preparadas com utilização de óleos sobre a solução de poço, na tentativa de diminuir a velocidade de difusão na gota, e assim melhorar a organização interna durante a formação dos cristais. A adição de dois tipos de óleos: “Paraffin e Silicon Oil da Hampton Research” foram utilizados conforme especificação do fabricante. Os resultados obtidos foram retardo e diminuição da nucleação, entretanto nenhum ganho no padrão de difração foi obtido.

Variações na concentração de proteína 5,8 a 17  $\mu\text{M}$  (3 a 9 mg/mL), concentração do agente precipitante, pH e relação proteína : solução de poço foram realizados. Observou-se que concentrações abaixo de 7,8  $\mu\text{M}$  (4 mg/mL) não resultavam em cristais e gotas em que a proteína estava acima de 9,8  $\mu\text{M}$  (5 mg/mL) pareciam semelhantes. Desse modo nos ensaios seguintes utilizou-se a concentração de 9,8  $\mu\text{M}$  (5 mg/mL). As variações na concentração do agente precipitante, pH e relação proteína : solução de poço não resultaram em melhorias no padrão de difração.

Outro problema que pode ocorrer em proteínas que difratam mal são desarranjos estruturais do cristal quando o mesmo é banhado na solução de crio – protetora, principalmente em proteínas que resultem em cristais com grandes quantidades de solvente<sup>73</sup>. De modo a minimizar esses danos, passagens em gradientes crescentes de glicerol ou etilenoglicol (5 – 25%) foram preparadas, infelizmente sem sucesso no ganho de resolução dos padrões de difrações. Também foram preparadas soluções de cristalização acrescidas de glicerol ou etilenoglicol (5 – 25%), resultando em cristal que seria formado já em solução crio, minimizando os danos sofridos ao cristal durante seu congelamento. Nessa abordagem os cristais apresentaram pouca nucleação com cristais maiores nas concentrações de 16, 18 e 22% de glicerol (tabela 3.14 – **N** e figura 3.13 – **Q**),

porém, sem ganho de resolução em seus padrões de difração. Também foram testados outros agentes crio protetores como PEG 200, 400 e 500, também sem melhorias.

Uma nova triagem foi realizada na condição de cristalização do cristal **N** (tabela 3.14 – **N**) em que 96 aditivos diferentes (Additive Screen HR2-428 – Hampton Research) foram adicionados a gota de cristalização. Um total de 35 condições aditivadas resultaram em cristais, todos com a forma do cristal nativo **N**, dos quais alguns não apresentaram padrão de difração enquanto outros apresentaram padrões de difração com resolução superior a 10 Å e “spots” mal definidos (figuras dos cristais e imagens de difração não mostrados).

Devido às inúmeras abordagens para a formação de cristais que apresentassem bons “spots” a alta resolução falharem, iniciaram-se tratamentos pós-cristalização como “annealing”, desidratação e “seeding” para os cristais já formados da proteína SELA ligada ao tRNA (tabela 3.14 – **N**).

Na tentativa de minimizar os problemas ocorridos durante o congelamento, muitas vezes desigual, do cristal de proteína causando assim aumento da desordem interna, especialmente em grandes cristais ou cristais com grande quantidade de solvente. A técnica de “annealing”, de modo simplificado, trata-se de impedir o fluxo de nitrogênio que mantém o cristal congelado no suporte do difrator de raios-X, por alguns segundos até que o cristal atinja a temperatura ambiente seguido por um rápido congelamento, com o restabelecimento do fluxo de nitrogênio. Processo que pode ser realizado uma única vez, variadas vezes em intervalos determinados de tempo, ou mesmo pela remoção do cristal do “loop” para novo banho em solução crio – protetora e novamente rápido congelamento pela recolocação do cristal no “loop” e fluxo de nitrogênio<sup>73-75</sup>. Como resultados, foram obtidos uma piora significativa ou mesmo ausência de difração dos cristais para a proteína SELA. Também foram realizados testes que comprovassem se o congelamento dos cristais estava danificando sua organização interna, através da colocação do cristal em um capilar e posterior exposição ao feixe de raios-X, sem a necessidade do congelamento do cristal. Essa abordagem não foi utilizada para coletas de conjunto de dados, mas para verificar se o cristal apresentava um padrão de difração melhor comparado aos cristais congelados. O mesmo padrão a baixa resolução e qualidade

dos “spots” foram observados caracterizando problemas na organização do cristal e não no processo de congelamento.

A tentativa seguinte de melhorar o padrão de difração encontrado nos cristais foi sua desidratação. Segundo essa técnica o cristal formado em uma dada condição possui determinada quantidade de solvente internamente as suas moléculas. Essa quantidade de solvente é variável e pode ser parcialmente reduzida (desidratada), através de passagens sucessivas do cristal pela solução mãe de cristalização com gradientes crescentes do agente precipitante, promovendo um melhor empacotamento e organização (ordenamento) do cristal<sup>74, 76-78</sup>. Na condição de cristalização em questão, o agente precipitante NaCl (2,0 M) teve gradual aumento de 0.1 M até 3.0 M de NaCl. As passagens dos cristais foram realizadas primeiramente a cada 5 minutos e em uma segunda abordagem a cada 12 horas. Em ambas as metodologias os cristais resistiram às passagens ao longo do gradiente crescente de concentração do agente precipitante, sem apresentarem qualquer dano visível (rachaduras, quebras ou mesmo dissolução). Mesmo assim, nenhum ganho na resolução do padrão de difração foi obtido.

O terceiro tratamento pós-cristalização foram experimentos de “seeding”, mais especificamente “microseeding”, em que cristais são colocados em um tubo e triturados com bolinas de vidro (*glassbead*) sob forte agitação (*vortex*). Os fragmentos dos cristais triturados, denominados de “sementes”, são utilizados como núcleos a serem colocados em novas gotas de cristalização contendo a mesma concentração de proteína e pequena diminuição da concentração do agente precipitante. Diminuição necessária para evitar em grande parte a formação de novos núcleos, priorizando o crescimento dos núcleos ali colocados. Técnica aprendida durante o “IV CeBIME/LNLS Structural Molecular Workshop – techniques, tools and developments in protein crystallization 2009”, para o completo entendimento da técnica e suas variantes utilizar o livro editado por Bergfors<sup>79</sup>. Desse modo, com um pelo de rabo de cavalo previamente passado na gota de cristais triturados, carregando assim, “sementes” de micro-cristais, foram feitas estrias seriadas em gotas de proteína com solução de cristalização (possuindo seu agente precipitante em concentrações decrescentes) para crescerem. Houve crescimento de cristais (figura 3.13 – **R**), porém sem ganhos na resolução do padrão de difração.

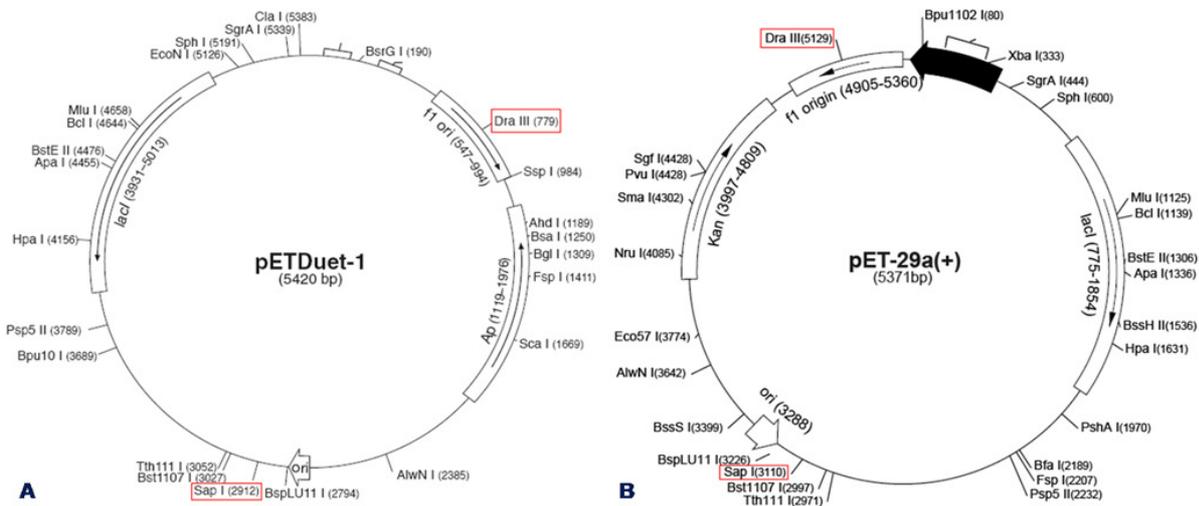
### 3.6.2 Proteína Selenocisteína Sintase-2

A doutoranda Lívia Regina Manzine (IFSC – USP), durante a preparação de seus experimentos de Fluorescência e determinação estequiométrica entre os tRNA<sup>sec</sup> ligantes à proteína homodecamérica Selenocisteína Sintase (SELA), observou através de amostras de proteína aplicadas em gel nativo corado inicialmente com brometo de etídeo (específico para DNA/RNA) e posteriormente com *Coomassie Blue* (específico para proteínas), que algum material permanecia ligado a proteína.

Como é sabido pela literatura<sup>26, 40, 46</sup> que a proteína liga-se a tRNAs, é possível que algumas moléculas de tRNA continuem ligadas a proteína mesmo após o processo de purificação e que esse material ligado resulte em heterogeneidade da amostra prejudicando os ensaios de cristalização, resultando nas baixas resoluções de difrações obtidas.

Na tentativa de conseguir a proteína SELA livre de moléculas de tRNA<sup>sec</sup> a mesma foi subclonada e expressa em cepas de *E. coli* WL81460 (FM433  $\Delta$ (selC)400), deficientes para o gene selC, gentilmente cedida pelo laboratório do Prof. Dr. August Böck do Instituto de Biociências Moleculares da Universidade de Frankfurt, Alemanha. Para isso, a doutoranda Lívia converteu a linhagem WL81460 para WL81460(DE3) competente, capaz de expressar a enzima T7 RNA Polimerase sobre controle do *lacUV5*, podendo assim, ser utilizada para expressão de qualquer gene clonado em vetores pET (Novagen). Devido à linhagem WL81460(DE3) ser resistente ao antibiótico kanamicina, o mesmo antibiótico utilizado na construção plasmideal pET29a-SELA, Lívia preparou a substituição do gene de kanamicina pelo de ampicilina encontrado no vetor pETDuet-1 (Novagen). Através das enzimas de restrição *Dra* III e *Sap* I (Fermentas), que apresentavam sítios de clivagens em regiões correspondentes entre os vetores e ausentes na sequência do gene selA (Figura 3.16). Como a substituição do gene de resistência para ampicilina foi realizada

a transformação do vetor agora denominado de pET29a-SELA-2 na cepa bacteriana de *E. coli* WL81460(DE3) competente.



**Figura 3. 16** - Mapas dos vetores pET (Novagen) destacando-se em vermelho os locais de clivagem para as enzimas *Dra* III e *Sap* I. Em **A**- vetor pETDuet -1, resistente a ampicilina. Em **B**- vetor pET-29a, resistente a kanamicina. Figura cedida pela doutoranda Lívia Regina Manzine.

A expressão e purificação da proteína denominada SELA-2, em bactérias que não produzem tRNA<sup>SEC</sup>, seguiram conforme protocolos estabelecidos anteriormente<sup>51</sup>. Os ensaios de cristalização foram repetidos com a utilização do robô de cristalização, novamente cristais de sais, cristais sem padrão de difração e a condição SELA-tRNA **N** (tabela 3.14 – **N**) com padrão de baixa resolução semelhantes aos para SELA-1 foram obtidos.

### 3.6.3 Ensaio de proteases, deleções e mutações sítio-dirigida para proteína Selenocisteína Sintase-2

Em casos onde a cristalização parece impossível ou os padrões de difração resultantes dos cristais obtidos são pobres, alternativas como uso de proteases, deleções de regiões específicas ou mutações sítio-dirigida, podem ser opções que resultem em cristais com bons padrões de difrações e conjuntos de dados suficientes para a resolução da estrutura cristalográfica<sup>74, 80</sup>.

A ausência de cristais ou mesmo cristais “ruins” podem não serem resultantes de problemas como heterogeneidade (presença de proteínas contaminantes, fragmentos resultantes de proteólises ou mesmo variados estados oligoméricos da proteína de interesse) ou instabilidade (devido a alterações na solubilidade por formações espontâneas, porém irreversíveis, de precipitados durante a fase de concentração protéica), mas pela impossibilidade da proteína estabelecer interações intermoleculares ou diminuição de regiões flexíveis, condições necessárias para nucleação e crescimento dos cristais. Quando problemas desses tipos ocorrem a partir da sequência completa da proteína em estudo, alternativas como a remoção dessas regiões interferentes ao processo de cristalização ou mesmo pela limitação dos ensaios cristalográficos a domínios funcionais da proteína (domínios de ligação de complexos, sinalizadores nucleares, dentre outros) são boas abordagens<sup>80</sup>.

Alguns exemplos de proteínas que apresentaram problemas que impediam suas cristalizações resultantes de: heterogeneidades devido à modificações pós-transcricionais como na proteína Tirosina Quinase<sup>81</sup>; heterogeneidade devido à ação de proteases na proteína Receptora de Interferon  $\gamma$ <sup>82</sup>; extremidades flexíveis de proteínas que inviabilizam a cristalização<sup>83</sup>; problemas de agregação solucionada pela mutação de aminoácidos hidrofóbicos por outros mais hidrofílicos<sup>84</sup> ou mesmo pela adição de resíduos indispensáveis à cristalização como no caso da presença do resíduo isoleucina no C-terminal da proteína Fibronectina<sup>85</sup>.

Determinar a sequência problema na proteína não é fácil, tampouco sua remoção por proteases. Mesmo com o conhecimento dos sítios de atuação das

proteases, muitas vezes não é sabido se tais locais estarão acessíveis a elas após o enovelamento ou oligomerização da proteína. Testes com diferentes proteases, em variadas concentrações e tempos de incubação são necessários. Desse modo ensaios de proteólises foram realizados na proteína SELA-2 na tentativa de remover regiões flexíveis localizadas externamente a sua estrutura homodecamérica. Para isso, duas metodologias foram utilizadas:

- metodologia **A**, em que a proteína de interesse é incubada com a protease, seguida pela inativação da reação de proteólise por inibidores de protease, purificação utilizando uma coluna de gel filtração (Superdex 200), concentração, incubação (ou não) com o tRNA ligante e preparação das gotas de cristalização<sup>80</sup>;

- metodologia **B**, em que a proteína de interesse é incubada (ou não) com o tRNA ligante, seguida pela incubação com a protease e preparação das gotas de cristalização<sup>86, 87</sup>.

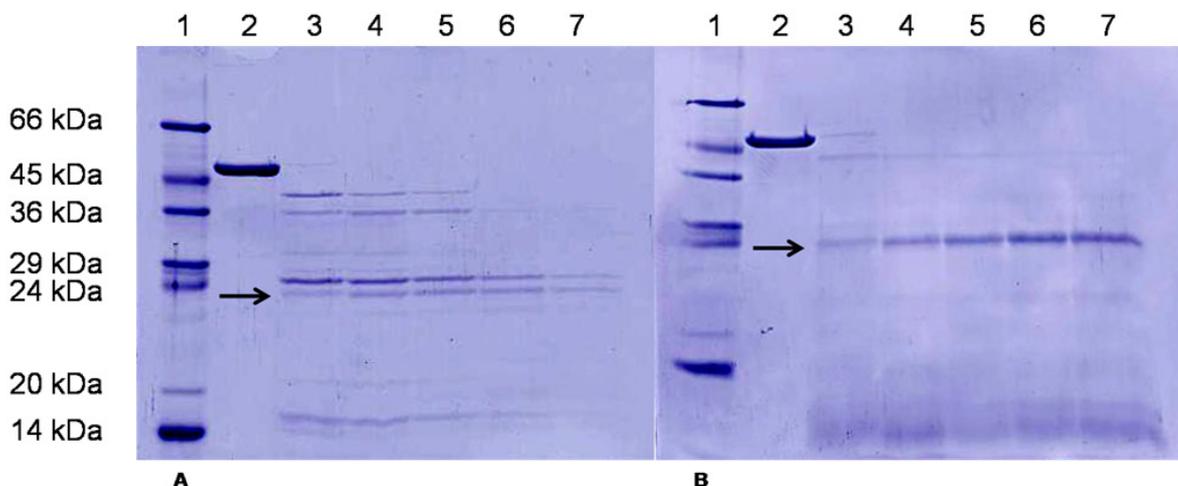
A principal diferença entre as duas metodologias é a ausência/presença do tRNA durante a incubação com a protease, que pode resultar na diferente atuação da protease decorrente de alguns sítios reconhecidos pela mesma estarem encobertos pelo tRNA ou ainda de organizações estruturais diferentes assumida pela proteína quando na presença da molécula ligante. Segundo os autores do trabalho, para a segunda metodologia a preparação das gotas de cristalização na presença da protease, não deve ser um problema, devido a diminuta quantidade de protease comparada a proteína em estudo, bem como a sua provável inativação quando misturada as soluções de cristalização<sup>87</sup>.

Dentre as proteases sugeridas e utilizadas, Tripsina e Papaína foram as enzimas escolhidas para os ensaios de proteólises. A primeira por ser a que demonstrou melhores resultados na literatura<sup>86, 87</sup>, a segunda por seu tampão de incubação ser muito parecido ao tampão de purificação da proteína de interesse<sup>80</sup>.

Testes iniciais com as duas proteases foram realizados para verificação de suas atividades na proteína de interesse. Desse modo, concentrações crescentes de proteases (20, 40, 60, 80 e 100 µg/mL) foram incubadas com 200 µg/mL de proteína SELA-2 por 60 minutos a 25° C.

A digestão com a protease Tripsina resultou de inúmeros a poucos fragmentos da proteína SELA-2 decorrentes do aumento da concentração de protease. É possível observar a presença da proteína SELA-2 sem digestão a 20 e 40 ug/mL de protease e uma grande digestão após 80 ug/mL. Desse modo escolheu-se a concentração 60 ug/mL de Tripsina para os ensaios pré-cristalização (figura 3.17 – **A**).

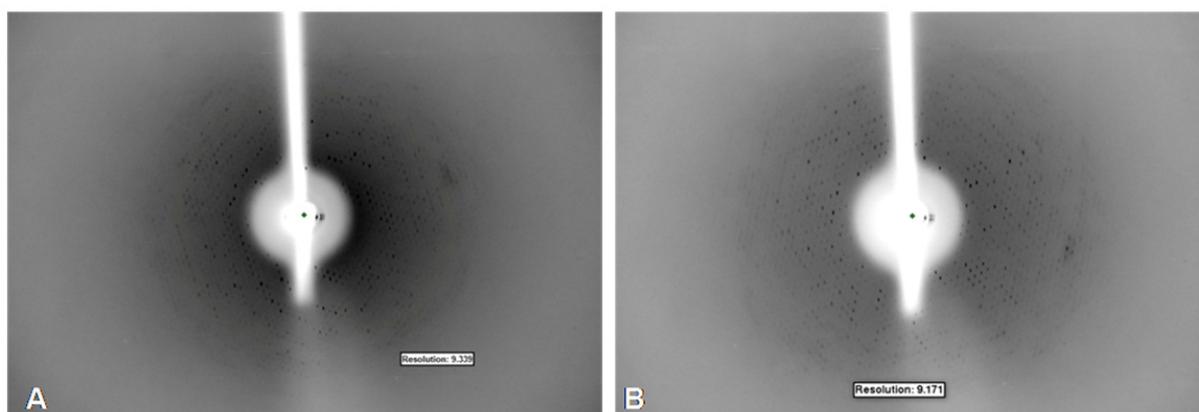
A digestão com a protease Papaína resultou na degradação total da proteína SELA-2, independente de sua concentração (figura não mostrada). Desse modo, o tempo de incubação foi reduzido para 20 minutos e apesar da presença da proteína SELA-2 sem digestão na concentração de 20 ug/mL a presença de um grande arrasto de fragmentos de proteínas ao longo das concentrações de protease, desqualificaram a utilização dessa protease nos ensaios pré-cristalização (figura 3.17 – **B**).



**Figura 3. 17** - Análise em SDS-PAGE dos ensaios de protease para proteína SELA-2. Em **A**- incubação com Tripsina – 60 min/25° C; Em **B**- incubação com Papaína – 20 min/25° C. Coluna **1**- padrão de massa molecular; coluna **2**- proteína SELA-2 controle; colunas **3** a **7**- proteína SELA-2 incubada em concentrações crescentes de protease 20, 40, 60, 80,100 ug/mL respectivamente. A seta negra indica respectivamente a protease Tripsina 23 kDa e Papaína 23,8 kDa.

Ensaio de cristalização após proteólise da proteína SELA-2 (na ausência e presença do tRNA) foram realizados segundo a metodologia **A**, para as condições **L**, **M** e **N** (tabela 3.13 – **L** e tabela 3.14 – **M** e **N**). Variações nas concentrações do agente precipitante (1,0 a 2,0 M de NaCl) e do pH (pH 6,5 e 7,0) para a condição **N**

resultou em um único cristal para as concentrações do agente precipitante de 1,6 e 2,0 M, ambos a pH 6,5, sem formação de cristais para as demais condições. As formas dos cristais foram semelhantes aos cristais obtidos anteriormente a proteólise, com padrão de difração ruim quando exposto aos raios-X no difrator do IFSC, para o cristal obtido na concentração de 2,0 M. O cristal com concentração de 1,6 M (figura 3.13 – **T**) apresentou padrão de difração com baixa resolução, mas com “spots” bons, semelhantes ao melhor padrão obtidos para proteína SELA1-tRNA ambos congelados em solução crio-protetora 20% etilenoglicol (tabela 3.14 – **N** e figura 3.13 – **S**). Enquanto para o cristal **S** apenas três imagens foram coletadas, para o cristal **T** um conjunto de dados com 130 imagens foi coletado na tentativa de indexar as imagens e determinar seu grupo espacial e parâmetros de cela (figura 3.18 e tabela 3.15).



**Figura 3. 18** - Padrão de difração para a proteína SELA-tRNA<sup>sec</sup>. Em **A**- melhor padrão de difração para proteína SELA1-tRNA<sup>sec</sup> congelada em solução crio-protetora 20% etilenoglicol, em **B**- proteína SELA2-tRNA<sup>sec</sup> digerida com protease Tripsina pela metodologia A, descrita no texto. Ambas cristalizadas conforme descrito na tabela 3.14 **N** (com variação do agente precipitante para 1.6M para SELA2-tRNA) e visualizados na figura 3.13 – **S** e **T**.

**Tabela 3. 15** - Informações sobre as condições de cristalização e análises dos padrões de difrações obtidos para a proteína Selenocisteína Sintase ligada ao seu tRNA.

	SELA : tRNA nativa	SELA : tRNA após proteólise
<b>Proteína: tRNA<sup>sec</sup> estequiometria (razão molar)</b>	1 : 12	1 : 12
<b>Concentrações Inicial (mg/mL)</b>		
<b>Proteína</b>		
Selenocisteína Sintase	5,0	5,0
<b>tRNA</b>		
Selenocisteína – tRNA <sup>sec</sup>	3,8	3,8
<b>Condição de cristalização</b>		
Tampão	0,1 M MES pH 6,5	0,1 M MES pH 6,5
Sal	0,1 M Fosfato de Sódio, 0,1 M Fosfato de Potássio	0,1 M Fosfato de Sódio, 0,1 M Fosfato de Potássio
Agente precipitante	2,0 M Cloreto de Sódio	1,6 M Cloreto de Sódio
Agente crio-protetor	20% Etilenoglicol	20% Etilenoglicol
<b>Ação de protease</b>	----	Tripsina
<b>Grupo espacial</b>	P3 (hP)	P3 (hP)
	$\alpha = 90$	$\alpha = 90$
	$\beta = 90$	$\beta = 90$
	$\gamma = 120$	$\gamma = 120$
<b>Dimensões da cela (Å)</b>	a = 274,9 b = 274,9 c = 498,7	a = 275,2 b = 275,2 c = 486,2
<b>Limites da difração (Å)</b>	9,33 – 28	9,17 – 28

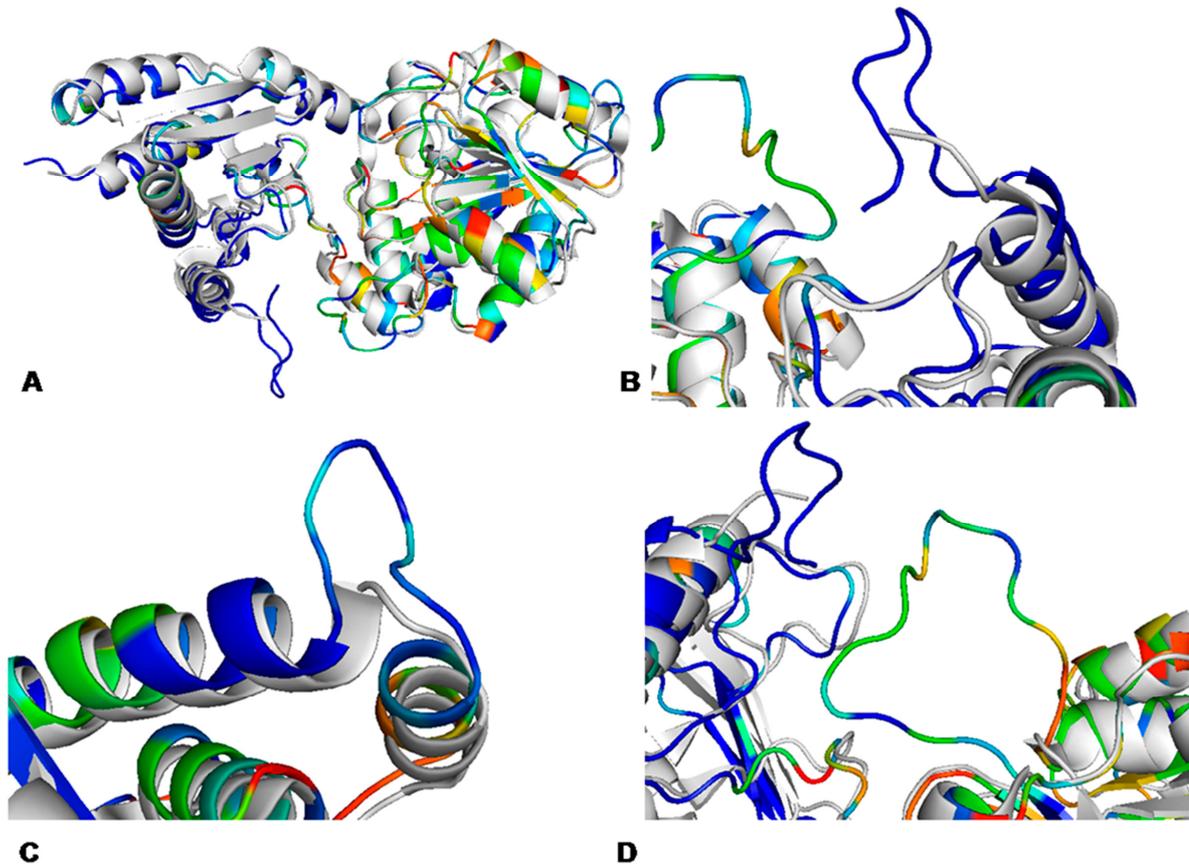
Massa molecular monomérica da proteína Selenocisteína Sintase (SELA) = 50667 Da;  
 Massa molecular do Selenocisteína – tRNA<sup>sec</sup> (tRNA<sup>sec</sup>) = 31000 Da

Devido aos baixíssimos rendimentos de proteínas resultantes da proteólise pela metodologia **A**, teve início a metodologia **B** em larga escala com a utilização do robô de cristalização, repetindo os fatoriais de cristalização (tabela 3.12), com a proteína SELA-2 e a mesma ligada ao tRNA<sup>sec</sup>. Resultante desses fatoriais, 35 cristais foram obtidos para proteína ligada ao tRNA, apenas alguns cristais na ausência do ligante, infelizmente todos cristais de sais.

Outra abordagem realizada foi à deleção de regiões na proteína SELA através de técnicas de biologia molecular. A partir do artigo de proteólise<sup>86</sup>, observou-se que na grande maioria das proteínas que não cristalizavam ou apresentavam padrões de

difração ruim ou mesmo ausente, após a proteólise de sequências flexíveis, principalmente localizadas em seu N-terminal, resultavam em cristalizações com padrões de difrações entre 1,7 – 2,8 Å, onde antes não havia cristais ou seus padrões de difração eram de baixa resolução.

Análises na proteína SELA através do modelo cristalográfico obtido da SELA de arqueobactéria *M. jannaschii* (figura 3.6), mostraram uma grande região flexível em sua extremidade N-terminal e outras duas regiões flexíveis ao longo da estrutura (figura 3.19). Oligonucleotídeos foram sintetizados para amplificação e clonagem dos genes *SelA* com a remoção dessas regiões flexíveis (tabela 3.16), denominadas de: *SelA*-mut\_nTerm (deleção do N-terminal – resíduo 1 a 61 figura 3.19 – **B**); *SelA*-mut\_alc2 (deleção da segunda região flexível – resíduo 102 a 108 figura 3.19 – **C**) e *SelA*-mut\_alc3 (deleção da terceira região flexível – resíduo 139 a 150 figura 3.19 – **D**). Essas deleções são colocadas como uma possibilidade a serem realizadas com cautela, pois decorrem de análises de sobreposição entre a estrutura modelada de bactéria e o modelo cristalográfico de arqueobactéria (figura 3.19 – **A**) e o alinhamento entre a sequência de *E. coli* e *M. jannaschii* demonstra falta de alinhamento para essas regiões (figura 3.5). Porém, quando analisadas pela coloração resultante da conservação sequencial de resíduos (figura 3.7) as mesmas resultam de baixa ou nenhuma conservação, indicando a possibilidade para mutação ou mesmo deleção dessas regiões.



**Figura 3. 19** - Análises das regiões flexíveis do modelo para proteína SELA de *E. coli* (colorido) sobreposto ao modelo cristalográfico de SELA de *M. jannaschii* (prateado). Em **A**-sobreposição dos dois modelos; **B**- evidenciando a região flexível para o N-Terminal (nTerm); **C**- evidenciando a região flexível para a segunda alça (alc2); **D**- evidenciando a região flexível para a terceira alça (alc3).

Nas sequências dos oligonucleotídeos é possível observar em verde uma pequena sequência amplificada que será removida, porém importante para a ancoragem da enzima de restrição, em vermelho o sítio de restrição para cada uma das construções e o grifo negro em baixo da trinca de nucleotídeos indicando o início da fase de leitura para a deleção do N-terminal (SelA-mut\_nTerm), bem como a junção para as deleções internas (SelA-mut\_alc2 e SelA-mut\_alc 3) (tabela 3.16).

**Tabela 3. 16** - Oligonucleotídeos sintetizado para as deleções das regiões flexíveis encontrados na sequência da proteína SELA. Em verde é mostrado uma pequena sequência utilizada para ancoragem pelas enzimas de restrição e que será eliminada após digestão, em vermelho os sítios para enzimas *Nde* I e *Hind* III (SelA-mut\_nTerm), *Bgl* II (SelA-mut\_alc2) e *Hind* III (SelA-mut\_alc3) e o grifo negro abaixo da trinca de nucleotídeos indica a fase de leitura para as respectivas deleções.

---

### SelA-mut\_nTerm

Oligonucleotídeo sense

5'- **ACT GTA TCA TAT** GGA AAA CTG GGC GCA AGA AGT CGA TGC C -3'

Oligonucleotídeo antisense

5' **TAG CTA AGC** TTT CAT TTC AAC AAC ATC TCC AAA AAC CG -3'

### SelA-mut\_alc2

Oligonucleotídeo sense

5'- **ACT GTA TAG ATC** TGC GCA GGC TAT GCG TTC GCC AGT GAC CCT CGA GTA TGA TCT GGA CG -3'

Oligonucleotídeo antisense

5'- **ATA CAG TAG ATC** TGC CCT GTA AAG CTC GCC CAA GGT TGG TAT GCA GCA CGG TTC CCG -3'

### SelA-mut\_alc3

Oligonucleotídeo sense

5'- **ACT GTA TAA GCT** TAA TAA CAA TGC GGC GGC GGT GTT ATT GAT GTT GGC GGC CAC TGC CAG CGG -3'

Oligonucleotídeo antisense

5'- **TGA CAT AAA GCT** TCA GCA GCT GCG CCA GCG CCC GAT CGC GAT GTC CGC G -3'

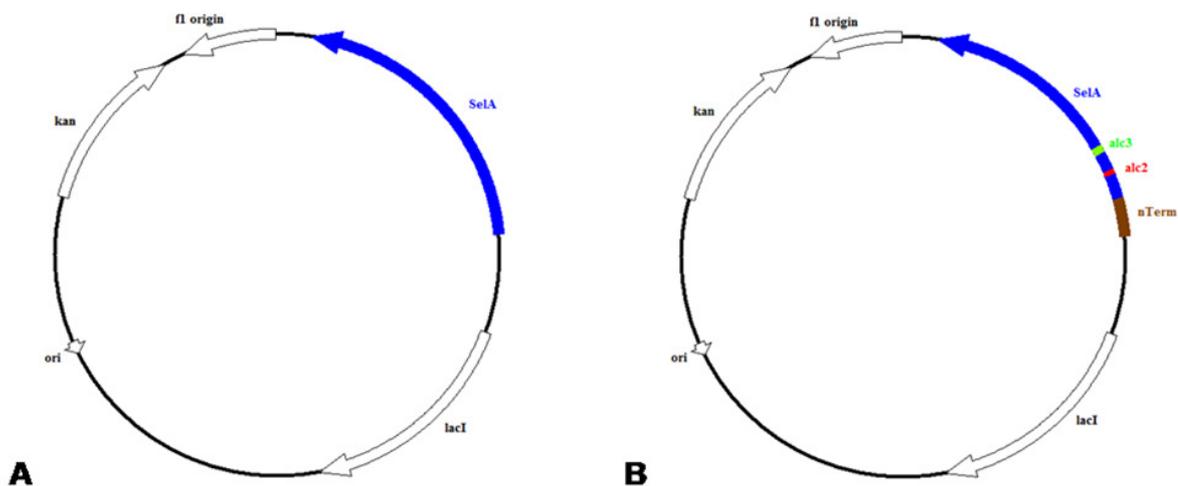
---

Para a preparação dos genes mutantes duas estratégias foram escolhidas:

- a deleção do N-terminal, realizou-se com a amplificação da sequência do gene SelA a partir da posição 184 de sua sequência nucleotídica, eliminando os 61 primeiros aminoácidos quando expressa. Após sua amplificação, o gene mutante foi clonado no vetor de propagação PTZ (Fermentas) e subclonado no vetor pET29a resistente a ampicilina (o mesmo usado para pET29a-SELA-2) através das enzimas de restrição *Nde* I e *Hind* III;

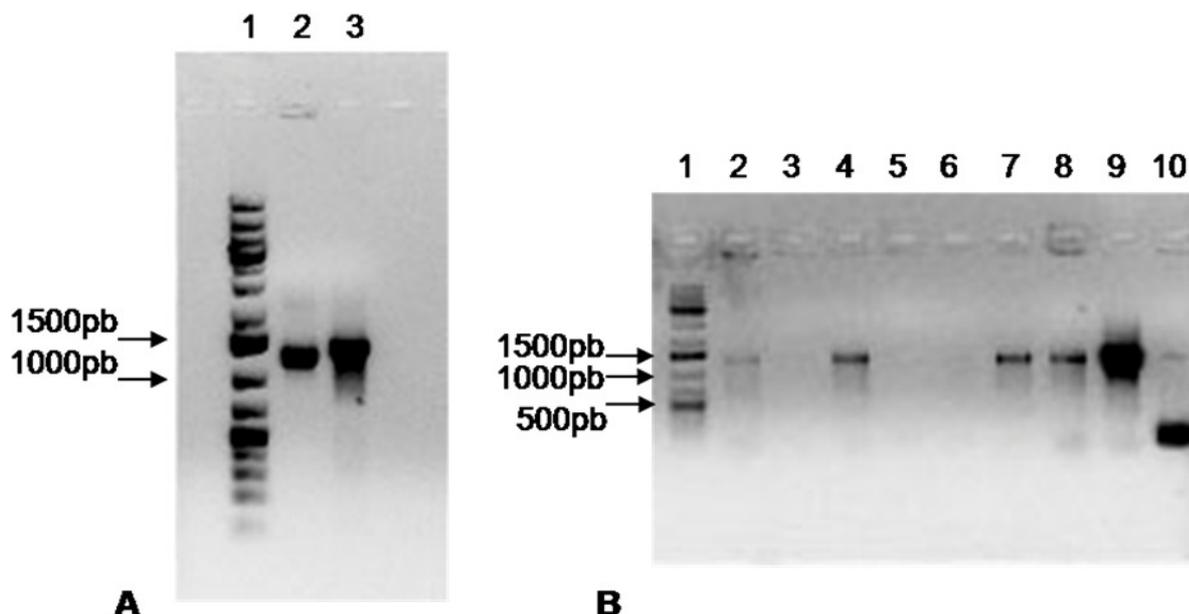
- as deleções das regiões alc2 e alc3, por se localizarem internamente ao gene SelA, será utilizada uma estratégia modificada da empregada na técnica de mutação sítio-dirigida "QuikChange Site-Directed Mutagenesis Kit" (Stratagene). Desse modo, oligonucleotídeos localizados nas extremidades da região que se deseja remover, portando sítios para uma mesma enzima de restrição em suas

extremidades 5', serão utilizados não apenas para amplificar a região que se deseja do gene como todo o vetor plasmideal. Em nosso experimento foram escolhidas as enzimas *Bgl* II e *Hind* III respectivamente para as deleções *alc2* e *alc3*. Após a reação de amplificação, suas extremidades são digeridas pela enzima correspondente e posteriormente ligadas obedecendo à complementaridade de bases resultante da digestão e refazendo a fase de leitura do gene (figura 3.20)



**Figura 3. 20** - Figura ilustrativa dos plasmídeos pET29a contendo o gene *SelA* de *E. coli* e o mesmo plasmídeo mostrando as três regiões deletadas. Em **A**- plasmídeo pET29a-SelA inteiro; **B**- plasmídeo pET29a-SelA com a região N-Terminal, *alc2* e *alc3* deletadas respectivamente em marrom, vermelho e verde.

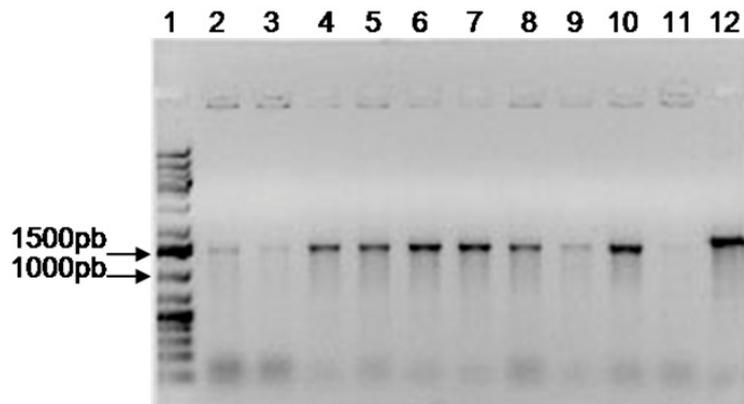
O gene *SelA*-mut\_nTerm, contendo a deleção dos primeiros 183 pb, foi amplificado por PCR a partir do DNA plasmideal pET29a-SELA-2. O fragmento amplificado de 1209 pb (figura 3.21 - **A**) foi purificado de gel de agarose 1% (Wizard SV Gel and PCR Clean Up System - Promega), clonado no vetor de propagação PTZ (Fermentas) e transformado em cepa bacteriana *E. coli* DH5- $\alpha$  competente. As buscas por colônias recombinantes, através de PCR de colônia utilizando oligonucleotídeos comerciais para o vetor PTZ resultaram em cinco colônias positivas (figura 3.21 - **B**). As colônias positivas foram sequenciadas (MegaBACE-1000 - GE Healthcare) no IFSC para verificação da deleção e fase de leitura do gene.



**Figura 3. 21** - Amplificação do gene *SelA*-mut\_nTerm e busca por plasmídeos recombinantes (PTZ-*SelA*-mut\_nTerm) pela técnica de PCR de colônia visualizado em gel de agarose 1% - TAE [1X] corado com brometo de etídeo. Em **A**- coluna 1- padrão molecular "1Kb plus DNA ladder"; coluna 2- amplificação do gene *SelA*-mut\_nTerm (1209 pb); coluna 3- amplificação controle do gene *SelA* (1392 pb). Em **B**- coluna 1- padrão molecular "1Kb plus DNA ladder"; coluna 2, 4, 7, 8, 9- colônias 12, 14, 17, 18, 19 PTZ-*SelA*-mut\_nTerm positivas respectivamente (amplificação do gene de interesse e parte da região de múltipla-clonagem do vetor PTZ totalizando aproximadamente 1500 pb); coluna 10- controle negativo (amplificação da região de múltipla-clonagem do vetor PTZ de aproximadamente 300 pb); coluna 3, 5, 6- nenhuma amplificação ocorreu.

O gene *SelA*-mut\_nTerm clonado no vetor pTZ foi digerido em suas extremidades pelas enzimas de restrição *Nde* I e *Hind* III, o mesmo foi realizado para o vetor pET29a-SELA-2. Ambos, gene *SelA*-mut\_nTerm e vetor pET29a, foram purificados de gel de agarose 1% (Wizard SV Gel and PCR Clean Up System - Promega) ligados e transformado em cepa bacteriana *E. coli* WL81460(DE3) competente.

Novamente a busca por colônias positivas foram realizadas por PCR de colônia utilizando oligonucleotídeos comerciais para o vetor pET29a, resultando em dez colônias positivas (figura 3.22). As colônias 3, 4, 5, 6, 7 e 9 foram escolhidas para testes de expressão segundo protocolo estabelecido anteriormente para a proteína SELA sem deleção.



**Figura 3. 22** - Busca por plasmídeos recombinates (pET29a-SelA-mut\_nTerm pela técnica de PCR de colônia visualizado em gel de agarose 1% - TAE [1X] corado com brometo de etídeo e testes de expressão das colônias positivas em SDS-PAGE. Em **A**- coluna 1- padrão molecular "1Kb plus DNA ladder"; coluna 2 a 11- colônias 1 a 10 pET29a-SelA-mut\_nTerm positiva (amplificação do gene de interesse e parte da região de múltipla-clonagem do vetor pET29a, totalizando 1407 pb); coluna 12- controle positivo (amplificação do gene SelA inteiro e parte da região de múltipla-clonagem, totalizando 1585 pb).

Ensaio de expressão, purificação e cristalização utilizando a nova construção pET29a-SelA-mut\_nTerm ainda não foram realizados, assim como as construções para as deleções internas pET29a-SelA-mut\_alc2 e pET29a-SelA-mut\_alc3, ficando assim disponíveis para estudos futuros.

Uma abordagem menos agressiva comparada às deleções de fragmentos ao longo do gene de interesse, são as mutações sítio-dirigidas, em que resíduos de aminoácidos são substituídos por outros através de amplificações por PCR utilizando oligonucleotídeos específicos para a região que se deseja mutar no gene<sup>80</sup>.

Novamente esbarramos no problema da identificação dos aminoácidos a serem mutados decorrente ao problema que desejamos estudar, seja ele, a identificação do sítio ativo, problemas de solubilidade, resíduos responsáveis pela oligomerização da estrutura, obtenção de cristais, melhoria em seus padrões de difração dentre outros. O foco das mutações sítio-dirigidas em nosso estudo é a obtenção de cristais e melhoria em seus padrões de difrações.

Para análises dos resíduos a serem mutados foi utilizado o servidor "Surface Entropy Reduction Prediction Server" (SERp)<sup>88</sup> (<http://nihserver.mbi.ucla.edu/SER/intro.php>), que procura identificar os melhores sítios para serem mutados visando otimizar a cristalização da proteína em estudo. Tendo em vista que as proteínas

possuem em sua superfície uma somatória alta de entropia devido as suas cadeias laterais longas, flexíveis e polares capazes de impedirem interações moleculares que resultem na organização cristalina, que deve possuir baixa entropia.

O SERp analisa a superfície exposta da proteína em estudo, substituindo os resíduos com alta entropia pelo resíduo Alanina, que possui baixo valor de entropia. Dessa forma, a partir da sequência de aminoácidos da proteína alvo o programa analisa:

- a predição secundária da estrutura, buscando por regiões em “coil” que preferencialmente são tidas como boas regiões para mutação por tenderem a estar expostas ao solvente;

- o perfil de entropia, em que os valores de entropia da cadeia lateral para todos os resíduos da proteína alvo são normalizados ao valor do resíduo com maior entropia (Glutamina). Assim, regiões com alta entropia são consideradas boas para mutação;

- alinhamento em bancos de sequências, buscando regiões conservadas (desfavoráveis, mas não excludentes para mutações) e regiões não conservadas possíveis de serem mutadas. Também realiza o alinhamento com o resíduo mutado (Alanina) das regiões propostas, promovendo maior confiabilidade a indicação de mutação.

Cada análise gera uma pontuação (positiva ou negativa) para cada resíduo na sequência que combinados identificam o resíduo mais favorável para a mutação. Regiões com pontuações mais altas são mais favoráveis a serem mutadas.

Após a identificação o servidor informa uma lista dos resíduos a serem mutados com suas respectivas posições e pontuações. Candidatos ideais para mutações são resíduos não-conservados, com alta entropia expostos a superfície da proteína (bons candidatos possuem pontuação superior a 3).

As análises para a proteína SELA encontraram três grupos candidatos a mutação (tabela 3.17), em que é sugerida a mutação dos resíduos destacados em verde para o aminoácido Alanina. É possível observar pelos dados fornecidos na tabela que os aminoácidos sugeridos para serem mutados possuíam altos valores

de entropia e baixos valores de conservação quando submetidos a alinhamentos múltiplos para essa região e que esse valor são mantido após a substituição para Alanina.

**Tabela 3. 17** - Grupos candidatos a mutação sítio-dirigida analisados pelo servidor “Surface Entropy Reduction Prediction Server” (SERp)<sup>88</sup>. Resíduos propostos a mutação são destacados em verde; resíduos possíveis de mutação e com alta entropia são destacados em azul; resíduos com baixa entropia são destacados em amarelo; demais resíduos não são destacados (<http://nihserver.mbi.ucla.edu/SER/intro.php>).

### Grupo 1

Resíduos: **KEAQSA**

- SERp pontuação: **4,96**
- Extensão do grupo: **74 - 79**
- Extensão proposta: **74 - 77**
- Mutações: **3**
- AA mutável: **K, E, Q**
- AA alvo: **A**

Resíduo	Entropia	Alinhamento pré-mutação	Alinhamento pós-mutação
<b>74: K</b>	<b>0,818</b>	<b>0,087</b>	<b>0,128</b>
<b>75: E</b>	<b>0,624</b>	<b>0,024</b>	<b>0,046</b>
76: A	0,612	0,123	0,123
<b>77: Q</b>	<b>0,532</b>	<b>0,144</b>	<b>0,081</b>
78: S	0,532	0,105	0,032
79: A	0,263	0,115	0,115

### Grupo 2

Resíduos: **KE**

- SERp pontuação: **4.96**
- Extensão do grupo: **274 - 275**
- Extensão proposta: **274 - 275**
- Mutações: **2**
- AA mutável: **K, E**
- AA alvo: **A**

Resíduo	Entropia	Alinhamento pré-mutação	Alinhamento pós-mutação
<b>274: K</b>	<b>0,624</b>	<b>0,119</b>	<b>0,170</b>
<b>275: E</b>	<b>0,624</b>	<b>0,665</b>	<b>0,006</b>

### Grupo 3

Resíduos: **EALSEK**

- SERp pontuação: **4.35**
- Extensão do grupo: **345 - 350**
- Extensão proposta: **349 - 350**
- Mutações: **2**
- AA mutável: **K, E**
- AA alvo: **A**

Resíduo	Entropia	Alinhamento pré-mutação	Alinhamento pós-mutação
345: E	0,291	0,316	0,017
346: A	0,342	0,057	0,057
347: L	0,263	0,224	0,166
348: S	0,554	0,093	0,080
<b>349: E</b>	<b>0,836</b>	<b>0,333</b>	<b>0,026</b>
<b>350: K</b>	<b>0,676</b>	<b>0,121</b>	<b>0,012</b>

SERp pontuação- pontuação total calculada pelo servidor para a extensão a ser mutada; AA- aminoácido; Entropia- entropia média da cadeia lateral para três resíduos; Alinhamento pré/pós mutação- valores de conservação para alinhamento do resíduo pré/pós mutação.

Para a realização das mutações sítio-dirigidas serão utilizados longos oligonucleotídeos complementares, contendo internamente a região a ser mutada e em suas extremidades sequências complementares ao gene de interesse (tabela 3.18). Desse modo, ocorrerá a hibridação junto ao DNA molde (região do gene que se deseja a mutação) e amplificação de todo o gene e vetor plamideal, conforme especificações de “QuikChange Site-Directed Mutagenesis Kit” (Stratagene).

Infelizmente não houve tempo hábil para realização das mutações sítio-dirigidas de modo que essa parte do trabalho fica disponível para estudos futuros.

**Tabela 3. 18** - Oligonucleotídeos sintetizado para as mutações sítio-dirigidas na sequencia gênica da proteína SELA. Em vermelho localizam-se os nucleotídeos mutados, o grifo negro indica o códon responsável pela tradução em Alanina e os demais nucleotídeos seguem a sequência nativa do gene para hibridação e extensão da reação de amplificação.

---

#### **SelA2\_mut\_74-79**

Oligonucleotídeo sense

5'- GCG CAA GAA GTC GAT GCC CGG TTG ACG GCA GCA GCG GCG AGC GCG CTG CGT CCG -3'

Oligonucleotídeo antisense

5'- CGG ACG CAG CGC GCT CGC CGC TGC TGC CGT CAA CCG GGC ATC GAC TTC TTG CGC -3'

#### **SelA2\_mut\_274-275**

Oligonucleotídeo sense

5'- CGA TCT TAG CCA GTA CGG TTT GCC GGC AGC GCC AAT GCC GCA GGA GTT GAT TGC G -3'

Oligonucleotídeo antisense

5'- CGC AAT CAA CTC CTG CGG CAT TGG CGC TGC CGG CAA ACC GTA CTG GCT AAG ATC G -3'

#### **SelA2\_mut\_349-350**

Oligonucleotídeo sense

5'- C CCT GCA GCT CTG AGT GCA GCA TTA CCG ACC CTG CGC CTG CTT ACC CGC -3'

Oligonucleotídeo antisense

5'- GCG GGT AAG CAG GCG CAG GGT CGG TAA TGC TGC ACT CAG AGC TGC AGG G -3'

---

## CAPÍTULO 4

### 4 DISCUSSÃO

Desde sua caracterização em 1991<sup>24, 40</sup>, apenas um estudo estrutural para a proteína Selenocisteína Sintase (SELA) de *Escherichia coli*, foi publicado utilizando-se da técnica de Microscopia Eletrônica de Transmissão (TEM)<sup>46</sup>. Mais recentemente, em 2005, uma proteína putativa homóloga de arqueobactéria (*Methanococcus jannaschii*) com 30% de identidade foi identificada e teve sua estrutura cristalográfica resolvida<sup>26</sup>.

Com o desenvolvimento de um novo protocolo de purificação<sup>51</sup>, aumentando em 15 vezes o rendimento da proteína SELA purificada e a utilização do “kit Megashortscript – Ambion” desenvolvido para transcrição de pequenas sequências de RNA, foi possível a experimentação estrutural da proteína SELA bem como da mesma ligada ao tRNA<sup>sec</sup> tanto pelas técnicas de Espalhamento de Raios-X a Baixos Ângulos (SAXS), como de ensaios de cristalização para Difração de Raios-X de monocristais (Cristalografia).

Os estudos de SAXS, realizados em colaboração com a Profa. Dra. Íris L. Torriane (LNLS) e Prof. Dr. Cristiano L. P. Oliveira (IF-USP), a partir de informações disponíveis no trabalho de Microscopia Eletrônica de Transmissão (TEM)<sup>46</sup> guiaram a imposição da simetria em P52, possibilitando a determinação dos envelopes da proteína SELA e SELA-tRNA (figura 3.11 – **A** e **B**). É nítida a diferença encontrada quando se compara os dois envelopes obtidos, visualizados pela sobreposição dos mesmos (figura 3.11 – **C**). É possível observar uma protuberância no centro de cada uma das subunidades diméricas, indicando possivelmente o local de ligação da molécula do tRNA<sup>sec</sup>. Quando visualizado a lateral da molécula SELA-tRNA (figura 3.11 e 3.12 – **B**) é possível observar um dos cinco orifícios que se repetem ao longo

da molécula e que não são visualizados na proteína sem o tRNA (figura 3.11 e 3.12 – **A**). Também é possível observar uma acentuada torção das cinco extremidades e cinco pequenas protuberâncias no interior do orifício central, decorrentes de modificações conformacionais resultantes da presença do tRNA ou mesmo do próprio tRNA (figura 3.12). A partir das curvas de espalhamentos parâmetros estruturais como dimensão máxima ( $D_{max}$ ), raio de giro ( $R_g$ ) e massa molecular (MM) foram determinados e seus valores podem ser comparados aos obtidos previamente por TEM (tabela 3.11). Os valores de massa molecular obtido a partir dos dados experimentais estão próximos aos calculados teoricamente e as dimensões do orifício central e da subunidade dimérica foram obtidos a partir dos PDBs gerados por SAXS e analisados no programa Pymol<sup>60</sup>. Assim como os dados obtidos por TEM<sup>46</sup> e por experimentos de Fluorescência, realizados pela doutoranda Lívia Regina Manzine também envolvida nos estudos da via de Selenociteína do grupo IFSC – USP, os dados de SAXS indicaram que a proteína homodecamérica SELA é capaz de ligar-se a cinco moléculas de tRNA, um tRNA por dímero da proteína.

Os ensaios de cristalização para a proteína SELA na forma apo e ligada ao tRNA foram realizados a partir da proteína expressa tanto em cepas de *E. coli* BL21(DE3) como WL81460(DE3) deficiente para o gene selC (clonagem realizada pela doutoranda Lívia Regina Manzine). Os ensaios de cristalização utilizaram a técnica de difusão de vapor por gota suspensa em sua grande maioria, sendo posteriormente a aquisição do robô de cristalização “Honeybee 939” (Genomic Solution) substituído pela técnica de difusão de vapor por gota depositada. Entretanto, todas as condições de refinamento foram realizadas manualmente por gota suspensa. Os cristais obtidos da proteína ligada ao tRNA resultaram em padrões de difrações melhores comparados a proteína sem ligante (tabela 3.13 e 3.14) e (figura 3.15), desse modo foi dada prioridade a condição de cristalização da proteína SELA-tRNA obtida na condição **N** (tabela 3.14) e inúmeros testes para tentar melhorar seu padrão de difração foram realizados conforme descrito a seguir:

- rápida formação dos cristais (aproximadamente 15 horas). Desse modo a adição de óleos sobre a solução de poço, na tentativa de diminuir a velocidade de difusão na gota e assim melhorar a organização interna durante a formação dos cristais;

- ausência total de difração após 19 dias de vida do cristal. Estabelecendo seu congelamento após dois dias da preparação da gota de cristalização;

- congelamento dos cristais em soluções crio – protetoras (preparo da solução mãe de cristalização com glicerol ou etilenoglicol 20% - concentração final). A solução crio – protetora com etilenoglicol foi mais eficaz;

- congelamento dos cristais através de passagens sucessivas por soluções crio – protetoras com concentrações crescente do agente crio – protetor (5 a 25% de glicerol ou etilenoglicol). Os resultados foram similares ao banho do cristal diretamente em solução crio – protetora 20%;

- preparação da cristalização já em solução crio – protetora (5 a 25% de glicerol, etilenoglicol, PEG 200, PEG 400 ou PEG 500), na tentativa de evitar desarranjos estruturais do cristal quando o mesmo é banhado na solução crio – protetora. Nessa abordagem os cristais apresentaram poucas nucleações com cristais maiores nas concentrações de 16, 18 e 22% de glicerol;

- variações na concentração da proteína (3 a 9 mg/mL) e estequiometria do tRNA ligante (1 proteína : 5; 7,5; 12; 16; 20). Observando que concentrações inferiores a 4 mg/mL não resultavam em cristais, concentrações acima de 5 mg/mL os cristais obtidos possuíam padrões de difrações semelhantes estabelecendo a concentração de 5mg/mL como padrão para os ensaios de cristalização. Com relação à estequiometria obteve-se cristais até a relação 1 proteína : 12 tRNA, relações maiores não resultavam em cristais. Como cristais na ausência do tRNA e que não apresentavam difração, semelhantes a condição **N**, foram obtidos (tabela 3.13 – **L**) estabeleceu-se a relação 1 proteína : 12 tRNA de modo que os possíveis dez sítios de ligação ao tRNA fossem supridos pela molécula ligante;

- adição de 96 aditivos (Additive Screen HR2-428 – Hampton Research) na gota de cristalização previamente a formação do cristal. Um total de 35 condições aditivadas resultaram em cristais com forma semelhante a do cristal nativo. Dentre esses cristais alguns não apresentaram padrão de difração e os que apresentaram resultaram em resolução superior a 10 Å e “spots” mal definidos.

Infelizmente as abordagens acima realizadas para obtenção de cristais que apresentassem melhores resoluções de difração falharam. Desse modo três novas

abordagens pós - cristalização como “annealing”, desidratação e “seeding” foram realizadas. Na técnica de “annealing” ocorreu uma piora até a ausência total de difração. Os cristais que foram desidratados resistiram bem às sucessivas passagens pelo gradiente crescente de concentração do agente precipitante, não apresentando danos visíveis como rachaduras, quebras ou mesmo dissolução do cristal, porém nenhum ganho na resolução do padrão de difração foi obtido. Por fim, a metodologia de “microseeding” foi aplicada com a obtenção de cristais crescidos das “sementes” “semeadas” na gota de cristalização (figura 3.13 – **R**), mas sem ganhos na resolução do padrão de difração.

Finalizando os estudos de cristalização da proteína SELA, abordagens como o uso de proteases, deleções de regiões específicas e mutações sitio-dirigida foram iniciadas. Após a escolha da protease, Tripsina, duas metodologias foram abordadas (digestão prévia a interação com o tRNA e digestão após interação com tRNA), em que a primeira apresentou cristais semelhantes ao cristal nativo, com padrões de difrações comparáveis ao melhor padrão de difração obtido pela proteína SELA-tRNA (figura 3.18). O que sugere que a protease agiu na proteína foi o retardo na eluição da proteína após a cromatografia por exclusão de tamanho e a cristalização de um único cristal em concentrações de 1,6 e 2,0 de agente precipitante, comportamentos completamente atípicos às condições previamente estabelecidas com a proteína nativa. Os padrões obtidos até o momento permitiram a determinação do grupo espacial e parâmetros de cela para a proteína SELA ligada ao tRNA (tabela 3.15). As deleções de regiões flexíveis específicas foram identificadas após análise comparativas entre o modelo gerado para SELA de *E. coli* e a estrutura cristalográfica de *M. jannaschii*. Nesse modelo foram observadas três regiões flexíveis: SelA-mut\_nTerm (deleção do N-terminal – resíduo 1 a 61); SelA-mut\_alc2 (deleção da segunda região flexível – resíduo 102 a 108) e SelA-mut\_alc3 (deleção da terceira região flexível – resíduo 139 a 150) (figura 3.19). Novamente é ressaltado, que essa abordagem é colocada como uma possibilidade para cristalizar a proteína SELA, porém, deve ser vista com cautela, pois decorrem de regiões com ausência de alinhamento entre as sequências. Até o momento foram realizadas a amplificação e clonagem do gene SELA com a deleção para a região codificante ao N-terminal (SelA-mut\_nTerm). A confirmação da deleção e fase aberta de leitura foram obtidas por seqüenciamentos realizados em nosso laboratório sendo o

próximo passo a verificação da expressão de forma solúvel da proteína mutante. Complementando as deleções específicas, uma abordagem menos agressiva é a mutação sítio-dirigida, obtida a partir de análises *in silico* da sequência de aminoácidos da proteína alvo pelo programa “Surface Entropy Reduction Prediction Server” (SERp)<sup>88</sup>. Essa análise encontrou três grupos candidatos a mutação (tabela 3.17), em que os resíduos alvos são substituídos pelo resíduo Alanina, que possui baixos valor de entropia. Até o momento apenas as análises de quais resíduos são sugeridos serem mutados e a síntese dos oligonucleotídeos foram realizados.

Paralelamente aos estudos realizados no laboratório, análises computacionais de modelagem, conservação e acoplamento estatísticos para a proteína SELA foram realizados com auxílio do Dr. Humberto d’Muniz Pereira, Dr. Lucas Bleicher e Mestre José Fernando Ruggiero Bachega. O modelo realizado para a proteína SELA de *E. coli* foi obtido a partir do alinhamento com a estrutura cristalográfica putativa da proteína SELA de *M. jannaschii*. Apesar da baixa identidade (30%) entre as duas sequências, o programa MODELLER<sup>57</sup> foi capaz de gerar um bom modelo (figura 3.6), que foi avaliado pelo servidor PROCHECK quanto a distribuição dos resíduos no diagrama de Ramachandran (figura 3.4) e (tabela 3.1). A construção desse modelo foi importante para os estudos de conservação e acoplamento estatístico, bem como para o planejamento das deleções de regiões flexíveis específicas. A identificação dos resíduos mais conservados e dos resíduos que possuíssem correlação entre si, foi possível a partir do alinhamento múltiplo de 237 sequências. Como resultados, um gradiente de conservação dos resíduos pode ser observado na figura 3.7, com o interior da estrutura apresentando resíduos mais conservados e tanto o N como o C terminal quase completamente sem conservação. Também é possível observar a presença do resíduo Lisina (K-295) ligante ao cofator PLP, o resíduo Asparagina (N-85) com alto grau de conservação e os resíduos Histidina (H-92) e Glutamina (Q-301) com grau menor de conservação, porém, todos localizados próximos ao sítio ativo. Com o intuito de identificar resíduos correlacionados entre si e que possam desempenhar funções na oligomerização da proteína SELA ou ligação da mesma à molécula de tRNA, estudos de acoplamento estatísticos foram realizados. Sete grupos destacando o respectivo resíduo, sua posição e a classificação quanto à característica de sua cadeia lateral foram reunidos na tabela 3.9 e projetados no modelo da proteína SELA (figura 3.8), para melhor visualização

da localização dos dados resíduos. Dentre os resultados obtidos, o grupo 3 chamou atenção devido aos resíduos presentes nesse grupo apresentarem-se menos de 30% do total de sequências, entretanto com correlações superiores a 80%. Um comportamento que sugere um favorecimento a conservação simultânea desses resíduos ao longo da evolução. Também é observado nesse grupo um forte comportamento de anti-correlação ao grupo 6. Infelizmente, ainda não foi encontrado nenhuma relação que pudesse identificar resíduos relacionados a oligomerização ou a região ligante ao tRNA. Ainda são necessárias mais informações estruturais para a proteína SELA.

#### **4.1 PERSPECTIVAS FUTURAS**

Dando continuidade aos experimentos iniciados durante esse projeto de doutorado, serão abordados:

**Deleções de regiões específicas** – realização dos testes de solubilidade, purificação e novas triagens de cristalização para a deleção N-terminal (SelA-mut\_nTerm), bem como, a amplificação, clonagem, testes de solubilidade, purificação e triagens de cristalização para as deleções internas SelA-mut\_alc2 e SelA-mut\_alc3 da proteína Selenocisteína Sintase (SELA);

**Mutação sítio-dirigida** – amplificação do vetor contendo o gene selA, a partir de oligonucleotídeos portando respectivamente as três mutações sítio-dirigidas: selA2\_mut\_74-79; selA2\_mut\_274-275 e selA2\_mut\_349-350. Posteriormente a transformação em cepas competentes, testes de solubilidade, purificação e triagens por novas condições de cristalização serão realizados para as três formas mutantes;

**Coleta de dados no ESRF** – coleta de dados de difração de raios – X dos cristais da proteína SELA-tRNA (tabela 3.14 – **N** e figura 3.13 – **N**) pelo Dr. Humberto d’Muniz Pereira no “European Synchrotron Radiation Facility” (ESRF) em Grenoble – França (setembro de 2010).



## REFERÊNCIAS

- 1 WHANGER, P. D. Selenocompounds in plants and animals and their biological significance. **Journal of the American College of Nutrition**, v. 21, n. 3, p. 223-232, 2002.
- 2 WESSJOHANN, L. A.;SCHNEIDER, A.;ABBAS, M.BRANDT, W. Selenium in chemistry and biochemistry in comparison to sulfur. **Biological Chemistry**, v. 388, n. 10, p. 997-1006, 2007.
- 3 CATON-WILLIAMS, J.HUANG, Z. Biochemistry of selenium-derivatized naturally occurring and unnatural nucleic acids. **Chemistry & Biodiversity**, v. 5, n. 3, p. 396-407, 2008.
- 4 PAPP, L. V.;LU, J.;HOLMGREN, A.KHANNA, K. K. From selenium to selenoproteins: synthesis, identity, and their role in human health. **Antioxidants Redox Signaling**, v. 9, n. 7, p. 775-806, 2007.
- 5 MOGHADASZADEH, B.BEGGS, A. H. Selenoproteins and their impact on human health through diverse physiological pathways. **Physiology (Bethesda)**, v. 21, n., p. 307-315, 2006.
- 6 STADTMAN, T. C. Selenium-dependent enzymes. **Annual Review Biochemistry**, v. 49, n., p. 93-110, 1980.
- 7 JOHANSSON, L.;GAFVELIN, G.ARNER, E. S. Selenocysteine in proteins-properties and biotechnological use. **Biochimica et Biophysica Acta**, v. 1726, n. 1, p. 1-13, 2005.
- 8 GROMER, S.;EUBEL, J. K.;LEE, B. L.JACOB, J. Human selenoproteins at a glance. **Cellular and Molecular Life Sciences**, v. 62, n. 21, p. 2414-2437, 2005.
- 9 STADTMAN, T. C. Biosynthesis and function of selenocysteine-containing enzymes. **The Journal of Biological Chemistry**, v. 266, n. 25, p. 16257-16260, 1991.
- 10 HATFIELD, D. L.GLADYSHEV, V. N. How selenium has altered our understanding of the genetic code. **Molecular and Cellular Biology**, v. 22, n. 11, p. 3565-3576, 2002.
- 11 GLADYSHEV, V. N.HATFIELD, D. L. Selenocysteine-containing proteins in mammals. **Journal of Biomedical Science**, v. 6, n. 3, p. 151-160, 1999.

12 DUFFIELD, A. J.; THOMSON, C. D.; HILL, K. E.; WILLIAMS, S. An estimation of selenium requirements for New Zealanders. **American Journal of Clinical Nutrition**, v. 70, n. 5, p. 896-903, 1999.

13 KOHRLE, J.; BRIGELIUS-FLOHE, R.; BOCK, A.; GARTNER, R.; MEYER, O.; FLOHE, L. Selenium in biology: facts and medical perspectives. **The Journal of Biological Chemistry**, v. 381, n. 9-10, p. 849-864, 2000.

14 NIRENBERG, M. Protein synthesis and the RNA code. **The Harvey Lectures**, v. 59, p. 155-185, 1965.

15 BOCK, A.; STADTMAN, T. C. Selenocysteine, a highly specific component of certain enzymes, is incorporated by a UGA-directed co-translational mechanism. **Biofactors**, v. 1, n. 3, p. 245-250, 1988.

16 ALLMANG, C.; WURTH, L.; KROL, A. The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. **Biochimica et Biophysica Acta**, v. 1790, n. 11, p. 1415-1423, 2009.

17 ALLMANG, C.; KROL, A. Selenoprotein synthesis: UGA does not end the story. **Biochimie**, v. 88, n. 11, p. 1561-1571, 2006.

18 LOBANOV, A. V.; KRYUKOV, G. V.; HATFIELD, D. L.; GLADYSHEV, V. N. Is there a twenty third amino acid in the genetic code? **Trends in Genetics**, v. 22, n. 7, p. 357-360, 2006.

19 BARON, C.; BOCK, A. The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. **The Journal of Biological Chemistry**, v. 266, n. 30, p. 20375-20379, 1991.

20 IBBA, M.; SOLL, D. Aminoacyl-tRNA synthesis. **Annual Review Biochemistry**, v. 69, p. 617-650, 2000.

21 IBBA, M.; SOLL, D. Aminoacyl-tRNAs: setting the limits of the genetic code. **Genes & Development**, v. 18, n. 7, p. 731-738, 2004.

22 YAREMCHUK, A. D.; TUKALO, M. A.; KRIKLIVYI, I. A.; MEL'NIK, V. N.; BERTHET-COLOMINAS, C.; CUSACK, S.; LEBERMAN, R. Crystallization of the seryl-tRNA synthetase-tRNA(Ser) complex from *Thermus thermophilus*. **Journal of Molecular Biology**, v. 224, n. 2, p. 519-522, 1992.

23 YAREMCHUK, A. D.; TUKALO, M. A.; KRIKLIVYI, I.; MALCHENKO, N.; BIOU, V.; BERTHET-COLOMINAS, C.; CUSACK, S. A new crystal form of the complex between seryl-tRNA synthetase and tRNA(Ser) from

*Thermus thermophilus* that diffracts to 2.8 Å resolution. **Federation of European Biochemical Societies Letters**, v. 310, n. 2, p. 157-161, 1992.

24 FORCHHAMMER, K.;LEINFELDER, W.;BOESMILLER, K.;VEPREK, B. BOCK, A. Selenocysteine synthase from *Escherichia coli*. Nucleotide sequence of the gene (*selA*) and purification of the protein. **The Journal of Biological Chemistry**, v. 266, n. 10, p. 6318-6323, 1991.

25 GANICHKIN, O. M.;XU, X. M.;CARLSON, B. A.;MIX, H.;HATFIELD, D. L.;GLADYSHEV, V. N. WAHL, M. C. Structure and catalytic mechanism of eukaryotic selenocysteine synthase. **The Journal of Biological Chemistry**, v. 283, n. 9, p. 5849-5865, 2008.

26 KAISER, J. T.;GROMADSKI, K.;ROTHER, M.;ENGELHARDT, H.;RODNINA, M. V. WAHL, M. C. Structural and functional investigation of a putative archaeal selenocysteine synthase. **Biochemistry**, v. 44, n. 40, p. 13315-13327, 2005.

27 FORCHHAMMER, K.;LEINFELDER, W. BOCK, A. Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. **Nature**, v. 342, n. 6248, p. 453-456, 1989.

28 ROTHER, M.;WILTING, R.;COMMANS, S. BOCK, A. Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. **Journal of Molecular Biology**, v. 299, n. 2, p. 351-358, 2000.

29 FAGEGALTIER, D.;HUBERT, N.;YAMADA, K.;MIZUTANI, T.;CARBON, P. KROL, A. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. **European Molecular Biology Organization Journal**, v. 19, n. 17, p. 4796-4805, 2000.

30 LEINFELDER, W.;FORCHHAMMER, K.;VEPREK, B.;ZEHELEIN, E. BOCK, A. In vitro synthesis of selenocysteinyl-tRNA(UCA) from seryl-tRNA(UCA): involvement and characterization of the *selD* gene product. **Proceedings of National Academy of Science of United States of America**, v. 87, n. 2, p. 543-547, 1990.

31 LOW, S. C.;HARNEY, J. W. BERRY, M. J. Cloning and functional characterization of human selenophosphate synthetase, an essential component of selenoprotein synthesis. **The Journal of Biological Chemistry**, v. 270, n. 37, p. 21659-21664, 1995.

32 GUIMARAES, M. J.;PETERSON, D.;VICARI, A.;COCKS, B. G.;COPELAND, N. G.;GILBERT, D. J.;JENKINS, N. A.;FERRICK, D. A.;KASTELEIN, R. A.;BAZAN, J. F. ZLOTNIK, A. Identification of a novel *selD* homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? **Proceedings of National Academy of Science of United States of America**, v. 93, n. 26, p. 15086-15091, 1996.

33 CHAPPLE, C. E.GUIGO, R. Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. **Public Library of Science**, v. 3, n. 8, p. e2968, 2008.

34 LIU, Z.;RECHES, M.;GROISMAN, I.ENGELBERG-KULKA, H. The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in Escherichia coli. **Nucleic Acids Research**, v. 26, n. 4, p. 896-902, 1998.

35 KOROTKOV, K. V.;NOVOSELOV, S. V.;HATFIELD, D. L.GLADYSHEV, V. N. Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. **Molecular and Cellular Biology**, v. 22, n. 5, p. 1402-1411, 2002.

36 TORMAY, P.;WILTING, R.;HEIDER, J. BOCK, A. Genes coding for the selenocysteine-inserting tRNA species from *Desulfomicrobium baculatum* and *Clostridium thermoaceticum*: structural and evolutionary implications. **The Journal of Bacteriology**, v. 176, n. 5, p. 1268-1274, 1994.

37 BARON, C.;WESTHOF, E.;BOCK, A.GIEGE, R. Solution structure of selenocysteine-inserting tRNA(Sec) from *Escherichia coli*. Comparison with canonical tRNA(Ser). **Journal of Molecular Biology**, v. 231, n. 2, p. 274-292, 1993.

38 LI, C.;RECHES, M.ENGELBERG-KULKA, H. The bulged nucleotide in the *Escherichia coli* minimal selenocysteine insertion sequence participates in interaction with SelB: a genetic approach. **The Journal of Bacteriology**, v. 182, n. 22, p. 6302-6307, 2000.

39 ROTHER, M.;RESCH, A.;GARDNER, W. L.;WHITMAN, W. B. BOCK, A. Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3' non-translated region. **Molecular Microbiology**, v. 40, n. 4, p. 900-908, 2001.

40 FORCHHAMMER, K. BOCK, A. Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. **The Journal of Biological Chemistry**, v. 266, n. 10, p. 6324-6328, 1991.

41 GLASS, R. S.;SINGH, W. P.;JUNG, W.;VERES, Z.;SCHOLZ, T. D. STADTMAN, T. C. Monoselenophosphate: synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. **Biochemistry**, v. 32, n. 47, p. 12555-12559, 1993.

42 ARAISO, Y.;PALIOURA, S.;ISHITANI, R.;SHERRER, R. L.;O'DONOGHUE, P.;YUAN, J.;OSHIKANE, H.;DOMAE, N.;DEFRANCO, J.;SOLL, D. NUREKI, O. Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. **Nucleic Acids Research**, v. 36, n. 4, p. 1187-1199, 2008.

43 MIHARA, H.;KURIHARA, T.;WATANABE, T.;YOSHIMURA, T. ESAKI, N. cDNA cloning, purification, and characterization of mouse liver selenocysteine lyase. Candidate for selenium delivery protein in selenoprotein synthesis. **The Journal of Biological Chemistry**, v. 275, n. 9, p. 6195-6200, 2000.

44 BURK, R. F. Molecular biology of selenium with implications for its metabolism. **The Journal of the Federation of American Societies for Experimental Biology**, v. 5, n. 9, p. 2274-2279, 1991.

45 XU, X. M.;CARLSON, B. A.;MIX, H.;ZHANG, Y.;SAIRA, K.;GLASS, R. S.;BERRY, M. J.;GLADYSHEV, V. N.HATFIELD, D. L. Biosynthesis of selenocysteine on its tRNA in eukaryotes. **Public Library of Science**, v. 5, n. 1, p. e4, 2007.

46 ENGELHARDT, H.;FORCHHAMMER, K.;MULLER, S.;GOLDIE, K. N. BOCK, A. Structure of selenocysteine synthase from *Escherichia coli* and location of tRNA in the seryl-tRNA(sec)-enzyme complex. **Molecular Microbiology**, v. 6, n. 23, p. 3461-3467, 1992.

47 TORMAY, P.;WILTING, R.;LOTTSPREICH, F.;MEHTA, P. K.;CHRISTEN, P. BOCK, A. Bacterial selenocysteine synthase--structural and functional properties. **European Journal of Biochemistry**, v. 254, n. 3, p. 655-661, 1998.

48 KRYUKOV, G. V.;CASTELLANO, S.;NOVOSELOV, S. V.;LOBANOV, A. V.;ZEHTAB, O.;GUIGO, R. GLADYSHEV, V. N. Characterization of mammalian selenoproteomes. **Science**, v. 300, n. 5624, p. 1439-1443, 2003.

49 ZHANG, Y.;FOMENKO, D. E. GLADYSHEV, V. N. The microbial selenoproteome of the Sargasso Sea. **Genome Biology**, v. 6, n. 4, p. R37, 2005.

50 LOBANOV, A. V.;FOMENKO, D. E.;ZHANG, Y.;SENGUPTA, A.;HATFIELD, D. L. GLADYSHEV, V. N. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. **Genome Biology**, v. 8, n. 9, p. R198, 2007.

51 CASSAGO, A. **Estudos moleculares da Selenocisteína Sintase de *Escherichia coli***. 2005. 87 f. Dissertação (Mestrado em Genética e Evolução) - Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, 2005.

52 BOCK, A.;FORCHHAMMER, K.;HEIDER, J.;LEINFELDER, W.;SAWERS, G.;VEPREK, B. ZINONI, F. Selenocysteine: the 21st amino acid. **Molecular Microbiology**, v. 5, n. 3, p. 515-520, 1991.

53 SCHON, A.;BOCK, A.;OTT, G.;SPRINZL, M. SOLL, D. The selenocysteine-inserting opal suppressor serine tRNA from *E. coli* is highly unusual in structure and modification. **Nucleic Acids Research**, v. 17, n. 18, p. 7159-7165, 1989.

54 FRUGIER, M.;FLORENTZ, C.;HOSSEINI, M. W.;LEHN, J. M. GIEGE, R. Synthetic polyamines stimulate in vitro transcription by T7 RNA polymerase. **Nucleic Acids Research**, v. 22, n. 14, p. 2784-2790, 1994.

55 STURCHLER, C.;WESTHOF, E.;CARBON, P.KROL, A. Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). **Nucleic Acids Research**, v. 21, n. 5, p. 1073-1079, 1993.

56 HARTMANN, R. K.;BINDEREIF, A.;SCHÖN, A.WESTHOF, E., **Handbook of RNA Biochemistry**. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA, 2005. V. 1.

57 SALI, A.BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of Molecular Biology**, v. 234, n. 3, p. 779-815, 1993.

58 CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Research**, v. 16, n. 22, p. 10881-10890, 1988.

59 LASKOWSKI, R. A.;MACARTHUR, M. W.;MOSS, D. S.THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, p. 283-291, 1993.

60 DELANO, W. L. The PyMOL molecular graphics system. *San Carlos*:. **DeLano Scientific**, 2002.

61 BACHEGA, J. F.;NAVARRO, M. V.;BLEICHER, L.;BORTOLETO-BUGS, R. K.;DIVE, D.;HOFFMANN, P.;VISCOGLIOSI, E.GARRATT, R. C. Systematic structural studies of iron superoxide dismutases from human parasites and a statistical coupling analysis of metal binding specificity. **Proteins**, v. 77, n. 1, p. 26-37, 2009.

62 LOCKLESS, S. W.RANGANATHAN, R. Evolutionarily conserved pathways of energetic connectivity in protein families. **Science**, v. 286, n. 5438, p. 295-299, 1999.

63 BLEICHER, L. **Implementação da análise de acoplamento estatístico e sua aplicação à família de proteínas Tirosina Fosfatase**. 2009. 167 f. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos da Universidade de São Paulo, Universidade de São Paulo, São Carlos, 2009.

64 WATERHOUSE, A. M.;PROCTER, J. B.;MARTIN, D. M.;CLAMP, M.BARTON, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189-1191, 2009.

65 KATOH, K.;MISAWA, K.;KUMA, K.MIYATA, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059-3066, 2002.

66 SVERGUN, D. I.KOCH, M. H. Small-angle scattering studies of biological macromolecules in solution **Reports on Progress in Physics**, v. 66, p. 1735-1782, 2003.

67 SEMENYUK, A. V.;SVERGUN, D. I. GNOM - A program package for small angle scattering data processing. **Journal of Applied Crystallography**, v. 24, n. 5, p. 537-540, 1991.

68 SVERGUN, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal of Applied Crystallography**, v. 25, p. 495-503, 1992.

69 SVERGUN, D. I.;PETOUKHOV, M. V.;KOCH, M. H. Determination of domain structure of proteins from X-ray solution scattering. **Biophysical Journal**, v. 80, n. 6, p. 2946-2953, 2001.

70 CHAYEN, N. E.;SHAW STEWART, P. D.;MAEDER, D. L.;BLOW, D. M. An automated system for microbatch protein crystallization and screening. **Journal of Applied Crystallography**, v. 23, p. 297-302, 1990.

71 CHAYEN, N. E.;STEWART, P. D.;BLOW, D. M. Microbatch crystallization under oil — a new technique allowing many small-volume crystallization trials **Journal of Crystal Growth**, v. 122, n. 1-4, p. 176-180, 1992.

72 D'ARCY, A.;ELMORE, C.;STIHLE, M.;JOHNSTON, J. E. A novel approach to crystallising proteins under oil **Journal of Crystal Growth**, v. 168, n. 1-4, p. 175-180, 1996.

73 PFLUGRATH, J. W. Macromolecular cryocrystallography--methods for cooling and mounting protein crystals at cryogenic temperatures. **Methods**, v. 34, n. 3, p. 415-423, 2004.

74 HERAS, B.;MARTIN, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. **Acta Crystallographica D - Biological Crystallography**, v. 61, n. pt 9, p. 1173-1180, 2005.

75 WARKENTIN, M.;THORNE, R. E. Slow cooling of protein crystals. **Journal of Applied Crystallography**, v. 42, n. pt 5, p. 944-952, 2009.

76 MUELLER, M.;JENNI, S.;BAN, N. Strategies for crystallization and structure determination of very large macromolecular assemblies. **Current Opinion in Structural Biology**, v. 17, n. 5, p. 572-579, 2007.

77 KUO, A.;BOWLER, M. W.;ZIMMER, J.;ANTCLIFF, J. F.;DOYLE, D. A. Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. **Journal of Structural Biology**, v. 141, n. 2, p. 97-102, 2003.

78 HERAS, B.;EDELING, M. A.;BYRIEL, K. A.;JONES, A.;RAINA, S.;MARTIN, J. L. Dehydration converts DsbG crystal diffraction from low to high resolution. **Structure**, v. 11, n. 2, p. 139-145, 2003.

79 BERGFORS, T. M., **Protein crystallization: techniques, strategies, and tips**. La Jolla: International University Line, 1999. 306p.

80 GIEGE, R.DUCRUIX, A., **Crystallization of nucleic acids and proteins: a practical approach**. 2nd ed. Oxford: Oxford University Press, 1999. 435p.

81 WEI, L.;HUBBARD, S. R.;HENDRICKSON, W. A.ELLIS, L. Expression, characterization, and crystallization of the catalytic core of the human insulin receptor protein-tyrosine kinase domain. **The Journal of Biological Chemistry**, v. 270, n. 14, p. 8122-8130, 1995.

82 WINDSOR, W. T.;WALTER, L. J.;SYTO, R.;FOSSETTA, J.;COOK, W. J.;NAGABHUSHAN, T. L.WALTER, M. R. Purification and crystallization of a complex between human interferon gamma receptor (extracellular domain) and human interferon gamma. **Proteins**, v. 26, n. 1, p. 108-114, 1996.

83 ZHANG, G.;LIU, Y.;QIN, J.;VO, B.;TANG, W. J.;RUOHO, A. E.HURLEY, J. H. Characterization and crystallization of a minimal catalytic core domain from mammalian type II adenylyl cyclase. **Protein Science**, v. 6, n. 4, p. 903-908, 1997.

84 JENKINS, T. M.;HICKMAN, A. B.;DYDA, F.;GHIRLANDO, R.;DAVIES, D. R.CRAIGIE, R. Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. **Proceeding of National Academy of Science of United States of America**, v. 92, n. 13, p. 6057-6061, 1995.

85 DICKINSON, C. D.;GAY, D. A.;PARELLO, J.;RUOSLAHTI, E.ELY, K. R. Crystals of the cell-binding module of fibronectin obtained from a series of recombinant fragments differing in length. **Journal of Molecular Biology**, v. 238, n. 1, p. 123-127, 1994.

86 DONG, A. et al. In situ proteolysis for protein crystallization and structure determination. **Nature Methods**, v. 4, n. 12, p. 1019-1021, 2007.

87 WERNIMONT, A.EDWARDS, A. In situ proteolysis to generate crystals for structure determination: an update. **Public Library of Science**, v. 4, n. 4, p. e5094, 2009.

88 GOLDSCHMIDT, L.;COOPER, D. R.;DEREWENDA, Z. S.EISENBERG, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. **Protein Science**, v. 16, n. 8, p. 1569-1576, 2007.