## **Relatório – Análise de Sequência no BLAST e STRING**

**Aluno(a):** Rafael da SIlva Barros  
**Disciplina:** Bioinformática  
**Organismo estudado:** *Escherichia coli* O157:H7 cepa Sakai

Durante a realização da análise das sequências proteicas derivadas dos cinco quadros de leitura (frames) do genoma de *Escherichia coli* O157:H7 cepa Sakai, foi utilizado o algoritmo BLASTp para comparar as sequências traduzidas com aquelas depositadas no banco de dados GenBank. A avaliação dos alinhamentos gerados baseou-se nos seguintes parâmetros: *Max Score*, *Total Score*, *Query Cover*, *E-value* e *Percent Identity*. Estes indicadores são essenciais para determinar a significância e a confiabilidade biológica dos alinhamentos obtidos.

O **Max Score** (pontuação máxima) representa a maior pontuação de alinhamento obtida para um segmento da sequência consultada. Essa pontuação reflete a qualidade do alinhamento local entre a sequência da consulta e a sequência de referência.

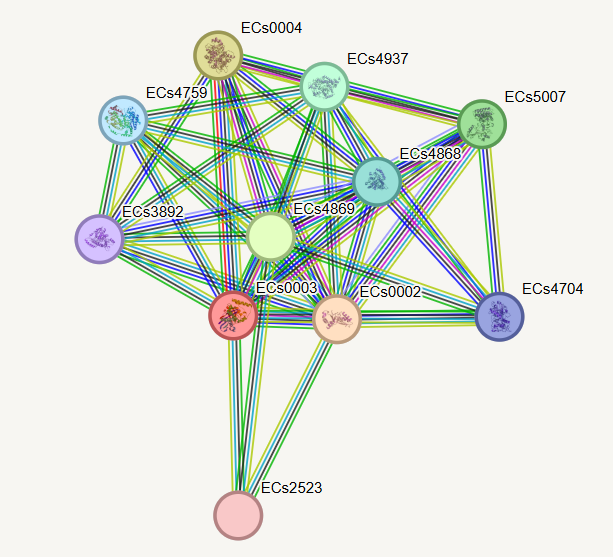
O **Total Score** corresponde à soma das pontuações de todos os segmentos alinhados entre a sequência consultada e a sequência do banco de dados. Quando o *Total Score* é igual ao *Max Score*, como observado nos resultados desta análise, significa que o alinhamento foi realizado em um único segmento contínuo, sugerindo uma correspondência altamente confiável.

A **Query Cover** (cobertura da consulta) representa a porcentagem da sequência analisada que foi coberta pelo alinhamento. Um valor de 100% implica que a totalidade da sequência da proteína foi alinhada com a sequência de referência, o que confere robustez à predição funcional da sequência.

O **E-value** (valor esperado) estima a probabilidade de o alinhamento ocorrer por acaso. Valores próximos de zero, como os obtidos nesta análise (0,0), indicam alta significância estatística e baixa probabilidade de ocorrência aleatória. Portanto, alinhamentos com E-value igual a zero são considerados biologicamente relevantes.

Por fim, a **Percent Identity** (identidade percentual) refere-se à proporção de resíduos de aminoácidos idênticos entre as sequências comparadas. Os valores obtidos nesta análise foram superiores a 99%, o que evidencia que as sequências possuem alta similaridade e, possivelmente, a mesma função biológica.

**homoserine kinase**



#### **1. Rede de Interação Proteica**

A imagem a cima representa a rede de interação da proteína **ECs0003**, obtida a partir do banco de dados STRING.

A rede demonstra um alto grau de conectividade entre ECs0003 e diversas outras proteínas funcionalmente associadas. As conexões coloridas indicam evidências distintas de interação, como coexpressão, associação em bases de dados, experimentação direta e correlação genômica. Esse padrão evidencia que a proteína ECs0003 participa de uma rede metabólica multifuncional, desempenhando papel relevante na homeostase celular e na integração de processos bioquímicos.

#### **2. Função da Proteína ECs0003 e Seu Papel na Rede Metabólica**

A proteína **ECs0003** está anotada como **glyceraldehyde-3-phosphate dehydrogenase (GAPDH)**, uma enzima chave da via glicolítica. Essa enzima catalisa a sexta etapa da glicólise, convertendo o gliceraldeído-3-fosfato (G3P) em 1,3-bisfosfoglicerato (1,3-BPG), com a redução de NAD⁺ a NADH.

No contexto da rede metabólica, a ECs0003 exerce função central por:

* Atuar na **glicólise**, principal via catabólica para obtenção de energia em forma de ATP;
* Produzir **NADH**, essencial para o metabolismo oxidativo e fermentativo;
* Servir como ponte metabólica entre a glicólise e outras vias biossintéticas.

Na rede apresentada, ECs0003 interage com enzimas das vias glicolítica, da via das pentoses fosfato e do metabolismo de aminoácidos, indicando sua participação em rotas anapleróticas e de resposta ao estresse oxidativo.

#### **3. Função da Rede Metabólica na *E. coli* Sakai**

A glicólise é uma via metabólica conservada e vital em organismos procarióticos e eucarióticos. Na *E. coli* O157:H7 Sakai, a importância dessa via é intensificada pelo seu papel em:

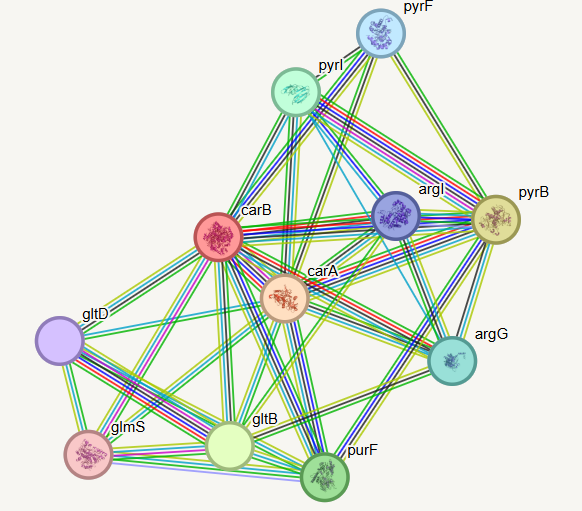
* **Fornecimento rápido de ATP**, necessário para proliferação e colonização no hospedeiro;
* **Produção de precursores metabólicos**, usados na síntese de aminoácidos, ácidos nucleicos e lipídios;
* **Adaptação metabólica**, especialmente sob condições anaeróbias ou com fontes alternativas de carbono.

A expressão e funcionalidade adequadas da proteína ECs0003 são, portanto, críticas para a sobrevivência, crescimento e virulência da *E. coli* Sakai em ambientes como o trato gastrointestinal humano.

#### **4. Referências**

* STRING Database. Protein-protein interaction network of *Escherichia coli* O157:H7 Sakai. Disponível em:<https://string-db.org>. Acesso em: maio de 2025.
* EcoCyc Database. *Escherichia coli* K-12 substr. MG1655: Glyceraldehyde-3-phosphate dehydrogenase. Disponível em:<https://ecocyc.org>.
* Stryer, L. (2019). *Bioquímica*. 8. ed. Rio de Janeiro: Guanabara Koogan.
* Nelson, D. L.; Cox, M. M. (2021). *Lehninger Princípios de Bioquímica*. 8. ed. Porto Alegre: Artmed.

**carbamoyl-phosphate synthase large subunit**



#### **1. Rede de Interação Proteica**

A figura a seguir ilustra a rede de interação da proteína **carB**, gerada com base na ferramenta STRING:

*Fonte: STRING Database*

A rede destaca a **carB** como um nó central, com diversas conexões fortes (linhas verdes, azuis e rosas) com outras proteínas, incluindo **carA**, **argI**, **gltB**, **pyrF**, entre outras. A densidade e diversidade dessas conexões indicam que a carB participa de uma complexa rede metabólica integrada, relacionada a múltiplas vias biossintéticas.

#### **2. Função da Proteína carB e Sua Atuação na Rede Metabólica**

A proteína **carB** codifica a subunidade grande da enzima **carbamoil-fosfato sintetase**, que atua na síntese de **carbamoil-fosfato**, um intermediário essencial para a biossíntese das bases pirimídicas (como uracila e citosina) e da arginina.

Sua função específica é:

* Catalisar, junto à subunidade **carA**, a conversão de glutamina, dióxido de carbono (CO₂) e ATP em carbamoil-fosfato;
* Servir como ponto de ramificação na síntese de **nucleotídeos** e **aminoácidos**.

Na rede apresentada, as conexões com **pyrB**, **pyrF** e **pyrI** indicam sua participação na **via das pirimidinas**, enquanto as ligações com **argG** e **argI** revelam seu envolvimento na **síntese de arginina**. As interações com **gltB**, **gltD** e **glmS** apontam sua integração com a regulação do metabolismo do nitrogênio e glutamina.

#### **3. Papel da Rede Metabólica na *E. coli* Sakai**

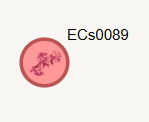
A rede centrada em **carB** regula dois eixos metabólicos críticos:

* **Biossíntese de pirimidinas**: fundamental para a produção de RNA e DNA. A regulação precisa dessa via garante o crescimento e divisão celular.
* **Biossíntese de arginina**: aminoácido essencial para a síntese proteica, ciclo da ureia (em organismos superiores) e regulação do pH intracelular.

Na *E. coli* O157:H7 Sakai, um patógeno entérico, essas rotas são cruciais para a rápida proliferação no hospedeiro e para a produção de fatores de virulência dependentes de síntese proteica e regulação genética. A disfunção da carB pode comprometer essas rotas, impactando a viabilidade celular.

#### **4. Referências**

* STRING Database. Protein-protein interaction network of *Escherichia coli* O157:H7 Sakai. Disponível em:<https://string-db.org>. Acesso em: maio de 2025.
* EcoCyc Database. *Escherichia coli* K-12 substr. MG1655: Carbamoyl-phosphate synthase large subunit (carB). Disponível em:<https://ecocyc.org>.
* Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L. (2019). *Bioquímica*. 8. ed. Rio de Janeiro: Guanabara Koogan.
* Nelson, D. L.; Cox, M. M. (2021). *Lehninger Princípios de Bioquímica*. 8. ed. Porto Alegre: Artmed.

**plasmid mobilization**  
  
**1. Contextualização e Visualização**

A figura a cima representa a proteína **ECs0089**, isolada na rede de interação gerada pela ferramenta STRING:

*Fonte: STRING Database*

A ausência de conexões visíveis com outras proteínas indica que **ECs0089** é uma proteína **hipotética ou pouco caracterizada**, sem interações funcionais ou físicas conhecidas registradas até o momento.

#### **2. Interpretação Biológica**

A anotação "ECs0089" sugere que essa proteína está presente no genoma da cepa O157:H7 da *E. coli*, mas sua função ainda **não foi experimentalmente confirmada**. As proteínas hipotéticas são comuns em genomas bacterianos, representando alvos promissores para futuras investigações funcionais, especialmente em cepas patogênicas como a Sakai.

A ausência de interações na STRING pode significar:

* Falta de dados experimentais ou computacionais disponíveis;
* Função altamente específica e ainda não explorada;
* Possível pseudogene ou produto de anotação incorreta.

#### **3. Importância da Investigação**

Apesar da ausência de interações conhecidas, o estudo de proteínas hipotéticas como a ECs0089 é essencial para:

* **Revelar novas vias metabólicas ou regulatórias** em bactérias;
* Identificar **potenciais alvos terapêuticos** ou de diagnóstico;
* Compreender mecanismos de **virulência e resistência** em cepas patogênicas.

A caracterização funcional pode ser realizada por abordagens como mutagênese gênica, análise de expressão proteica e estudos de bioinformática estrutural comparativa.

#### **4. Referências**

* STRING Database. Protein-protein interaction network of *Escherichia coli* O157:H7 Sakai. Disponível em:<https://string-db.org>.
* Galperin, M. Y.; Koonin, E. V. (2004). ‘Conserved hypothetical’ proteins: prioritization of targets for experimental study. *Nucleic Acids Research*, 32(18), 5452–5463.
* EcoCyc Database. Hypothetical proteins in *Escherichia coli*. Disponível em:<https://ecocyc.org>.
* Nelson, D. L.; Cox, M. M. (2021). *Lehninger Princípios de Bioquímica*. 8. ed. Porto Alegre: Artmed.